Comparative genomics of rumen methanogens

A thesis presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biochemistry

At Massey University, Palmerston North, New Zealand

Yang Li

2016

Abstract

Methane (CH_4) emissions from agriculture represent around 9% of global anthropogenic greenhouse gas emissions. The single largest source of this CH_4 is animal enteric fermentation, predominantly from ruminant livestock, where it is produced mainly in their fermentative forestomach (or reticulo-rumen) by a group of archaea known as methanogens.

In order to reduce CH_4 emissions from ruminants, it is necessary to understand the role of methanogenic archaea in the rumen, and to identify their distinguishing characteristics that can be used to develop CH_4 mitigation technologies. To gain insights into the role of methanogens in the rumen environment, two methanogens have been isolated from ovine rumen and their genomes were sequenced: methanogenic archaeon ISO4-H5 of the order Methanomassiliicoccales and *Methanobrevibacter* sp. D5 of *Methanobrevibacter gottschalkii* clade.

Genomic analysis suggests ISO4-H5 is an obligate hydrogen-dependent methylotrophic methanogen, able to use methanol and methylamines as substrates for methanogenesis. Like other organisms within this order, ISO4-H5 does not possess genes required for the first six steps of hydrogenotrophic methanogenesis. Comparison between the genomes of different members of the order Methanomassiliicoccales revealed strong conservation in energy metabolism, particularly in genes of the methylotrophic methanogenesis pathway, as well as in the biosynthesis and use of pyrrolysine. Unlike members of Methanomassiliicoccales from human sources, ISO4-H5 does not contain the genes required for production of coenzyme M (CoM), and requires external supply of CoM to survive.

Methanobrevibacter sp. D5 is a hydrogenotrophic methanogen predicted to utilise $CO_2 + H_2$ and formate as substrates. Comparisons between the available *Methanobrevibacter* genomes has revealed a high conservation in energy metabolism and characteristics specific to each clade. The coexistence of different *Methanobrevibacter* species in the rumen may be partly due to the physical association *Methanobrevibacter* species with different microorganisms and host surface, which allow unique niches to be established.

Acknowledgements

It was my privilege to work under the guidance of Dr. Graeme Attwood, Dr. Sinead Leahy and Associate Professor Jasna Rakonjac, I am grateful for their contant guidance and unflinching support, they are a source of inspiration and motivation. This thesis would not be possible without them.

I am grateful to the New Zealand Agricultural Greenhouse Gas Research Centre and Massey University for the financial support and the opportunity they gave me to conduct a truly meaningful research project contributing towards CH_4 mitigation. I would also like to thank AgResearch Ltd. for providing the excellent facility for me to conduct my research at the Grasslands Research Centre. My gratitude also to the Pastoral Greenhouse Gas Research Consortium (PGgRc) for kindly providing the methanogen cultures.

The experience and skills I have learned in my time here are invaluable, but the best thing that happened during my PhD is finding my better, wiser half, Jana Muller, who is also a PhD student here in AgResearch. She and my family provided me the strength to finish this long journey, which often seems never ending.

My project would not be possible without the kind assistance and continued support offered by all the nice people of our rumen microbiology team, my fellow PhD students and the experts of Massey University. I would like to thank Dr. Bill Kelly for teaching me Pulsed Field Gel Electrophoresis and the scientific insights he offered, Dr. Eric Altermann for his expertise in bioinformatics and insights into prophage, Dr Peter Janssen for his knowledge in methanogen biology and insights into bioenergetics, Dr Gemma Henderson for kindly providing the source culture H6, and her scientific insights. I would like to extend my thanks to my predecessor, Dr Jeyamalar Jeyanathan who I never met, for her excellent work on Rumen Cluster C enrichment and kindly providing the Rumen Cluster C cultures, Dr. Suzanne Lambie for teaching me how to extract genomic DNA from methanogens, Dong Li for his scientific insights and assistance in trouble shooting molecular biology problems, Dr. Janine Kamke and Priya Soni for working together on the transcriptome project, Caroline Kim for her experience in *Methanobrevibacter* sp. D5 isolation, Dr Siva Ganesh for his expertise on statistical analysis and Rstudio, Debjit Dey for teaching me techniques on anaerobic cultivation, Dr Linda Samuelson and Dr. Patrick Edwards for their assistance and expertise on ^{1}H- nuclear magnetic resonance (NMR) and metabolomics, Douglas Hopcroft and Jordan Taylor of Manawatu Microsocopy and Imaging

Centre (MMIC) for assistance in transmission electron microscopy, Dr. Linley Schofield, Dr. Dragana Gagic, Dr Sandra Kittleman, Dr. Henning Seedorf and Dr. Ron Ronimus for their scientific insights, Kerri Reilly and Savannah Devente for the singing and humming in the lab, Carrie Sang for keeping lab consumables supplied, Dr. Kenneth Teh for the his insights on the taxonomic relationship of PhD students and zombies, Dr. Sonal Shewaramani for sharing her experience in RNA extraction and genome assembly, and special thanks to Faith Cox for her friendship, teaching me techniques on anaerobic cultivation, purification of *Succinivibrio dextrinisolvens* H5 and working together on Rumen Cluster C. Special thanks to Dr. Nick Palevich and Dr. Preeti Raju for their friendship, being great office mates and supporting each other throughout our PhD. I would also like to thank my fellow PhD students for their friendship and support, and Dr. Sunny Sun and his family for their friendship and support.

Table of contents

List of Tables

Page

List of Figures

Page

Abbreviations

Non-standard abbreviations:

aa	Amino acids (length of peptide chain or sequence identity)
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BES	2-bromoethanesulfonic acid
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUbstitution Matrix
BRIG	BLAST Ring Image Generator
CAI	Codon adaptation index
CDS	Coding DNA sequence
CH_4	Methane
CO	Carbon monoxide
CO_2	Carbon dioxide
CoA	Coenzyme A
CoB	Coenzyme B
COG	Clusters of Orthologous Groups
CoM	Coenzyme M
CRISPR	Clustered regularly interspaced short palindromic repeat
D_2O	Deuterated water
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
$F_{390/420/430}$	Cofactor $F_{390/420/430}$
FDR	False discovery rate
Fdx	Ferredoxin
FGD	Functional genome distribution
GHGs	Greenhouse gas
GIT	Gastrointestinal
H_4MPT	Tetrahydromethanopterin
HMM	Hidden Markov model
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
IS	Insertion sequence

IVOM	Interpolated variable ordered motif
KEGG	Kyoto Encyclopedia of Genes and Genomes
KW	Kruskal-Wallis rank sum test
M3MSP	Methyl-3-methylthiopropionate
M3SP	Methylmercaptopropionate
MCL	Maximum Composite Likelihood
MDS	Multidimensional scaling
MF	Methanofuran
MMIC	Manawatu Microscopy and Imaging Centre
mRNA	Messenger RNA
N_2	Nitrogen (gas or liquid)
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthase
NZ	New Zealand
O_2	Oxygen
ORB	Origin recognition box
ORF	Open reading frame
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
RCC	Rumen Cluster C
RNA	Ribonucleic acid
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
SSPGMS	*Succinivibrio* spent pectin growth media supernatant
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TE	Tris EDTA
TEM	Transmission electron micrograph
TMH	Transmembrane helix
TOCSY	Total Correlation Spectroscopy

tRNA	Transfer RNA
UPGMA	Unweighted pair group method with arithmetic mean
UV	Ultra violet
VFA	Volatile fatty acid

Measurement Units:

^{o}C	Degree Celsius
µg	Microgram
µL	Microlitre
µm	Micrometer
µM	Micromolar
bp	Base pair
h	Hour
kcal	kilocalorie
kb	kilobase
kDa	kilodaltons
kpa	kilopascal
kV	kilovolts
L	Litre
M	Molar
Mb	Million base pairs
mg	Milligram
MHz	Mega hertz
min	Minutes
mL	Millilitre
mm	Millimeter
mM	Millimolar
mV	Millivolts
ng	Nanogram
nm	Nanometer
ppm	Parts per million
s	Seconds
v/v	Volume/volume
w/v	Weight/volume

Chapter 1
Introduction

This chapter reviews the literature before my thesis start date (11/2011), information published during the course of my thesis which was directly related to my thesis topic will be discussed as part of the summary, conclusions and future directions in Chapter 7.

1.1. Greenhouse gases and global warming

Greenhouse gases (GHGs) are atmospheric gases that absorb and emit radiation, warming the planet Earth and making it habitable to life (Karl and Trenberth 2003). GHGs include water vapour, carbon dioxide (CO_2), ozone, CH_4 and other trace gases, which contribute 60%, 26%, 8% and 6% radiative forcing warming effect respectively (Kiehl and Trenberth 1997). Since the start of the industrial revolution in the 1760s, atmospheric concentrations of CO_2, CH_4 and nitrous oxide have increased by 40%, 150% and 20% respectively and continue to rise today. Human activities involving fossil fuel usage, land use change and agriculture, are the main driver of the increasing GHGs (Cubasch *et al.* 2013). The averaged global combined land and ocean surface temperature showed an increase of 0.85 degree Celsius (^oC) between 1880 and 2012, consistent with the elevated GHG emissions due to anthropogenic activity (Hartmann *et al.* 2013). The elevated temperature directly contributes to the shrinking of glaciers, reduction of Arctic sea ice and ice sheets and the rise of global average sea level at an average rate of 1.8 millimeter (mm) per year (Church and White 2011). Global warming has also been linked to the increased frequency of hurricanes (Saunders and Lea 2008), as well as other extreme weather conditions (Rahmstorf and Coumou 2011; Schiermeier 2011). The elevated level of CO_2 has also caused ocean acidification (Fabry *et al.* 2008), as well as extinction of species and loss of biodiversity (Balint *et al.* 2011; Selbmann *et al.* 2012). Therefore, to limit the effects of global warming, it is necessary to reduce anthropogenic GHG emissions.

1.2. New Zealand agriculture and associated CH_4 emissions

CH_4 is the second largest source of anthropogenic GHGs, with an atmospheric lifetime of 12.4 years and 28 times the global warming potential of CO_2 over 100 years (Cubasch *et al.* 2013). CH_4 emissions accounted for 44% of New Zealand (NZ)'s total GHG emissions in 2013, of which 90.4% came from agriculture (Ministry for the Environment, 2015). NZ is unusual

because it is one of the few developed country where agricultural GHG emissions play a major role in the national emissions profile. Agriculture was responsible for 32.5% of NZ's total exports in 2014 (Statistics NZ 2015), which is sustained by a high number of livestock, including 29.8 million sheep, 6.7 million dairy cattle, 3.7 million beef cattle and 0.96 million deer (MacPherson 2014). The high numbers of ruminant livestock results in high production of CH_4 via enteric fermentation, accounting for 79.9% of NZ's total CH_4 emissions (Ministry for the Environment, 2015). Because enteric CH_4 emissions account for such a large proportion of NZ's total GHG emissions, mitigation of enteric CH_4 emissions could have a significant effect in lowering the national GHG emissions. In order to formulate strategies for enteric CH_4 mitigation, it is necessary to understand the source of CH_4 and its formation in the rumen environment.

1.3. The rumen environment and CH_4 formation

The stomach of ruminant animals, such as cattle, sheep, deer and goats, is divided into four compartments that perform different digestive functions. The largest compartment, called the rumen, is where symbiotic microorganisms breakdown the digesta; the reticulum controls the movement of digesta as well as containing a small amount of digesta and microorganisms, the omasum absorbs nutrients and re-absorbs water, and the abomasum is the true acid stomach that secretes acidic digestive juices (Boomker 2000). The digestion of feed is carried out mainly in the rumen and reticulum by a combination of rumination and enzymatic degradation by rumen microorganisms. Ruminants do not possess the enzymes necessary to digest the structural carbohydrate components of plant cell walls, such as cellulose, hemicelluloses and pectin. Ruminants rely on the complex microbial community within their rumen to hydrolyze and ferment plant polysaccharides (Hungate 1966). The microbial community in the rumen and reticulum includes bacteria, archaea fungi, viruses, and protozoa, the number and diversity of which is as rich as any other natural microbial habitat (Hungate 1966). Ruminant animals and their rumen microbes have achieved a mutually beneficial symbiosis, in which the animal benefits from the volatile fatty acids (VFAs), acetate, propionate and butyrate which are the products of forage digestion by microbes, while the microbes benefit from a stable, buffered environment in which they are able to grow (Hungate 1966).

In addition to the VFAs mentioned above, there are also intermediate fermentation products that are produced and consumed continuously, such as succinate, lactate, formate, ethanol,

methanol, methylamines, and hydrogen (H_2). Several of these intermediary compounds are used by rumen methanogenic archaea to produce CH_4 (Hungate 1967).

The abundance of rumen archaea within the ovine, bovine and caprine rumen ranges from 0.3% to 3.3% of total rumen microorganisms, with an average of 1.69%, as estimated by archaea-specific deoxyribonucleic acid (DNA) probes targeting a conserved region of the small subunit ribosomal ribonucleic acid (RNA) genes (Lin *et al.* 1997). It has also been observed that 2.8% to 4.0% of rumen microorganisms showed cofactor F_{420} (F_{420}) autofluorescence, a unique property of methanogens (Yanagita *et al.* 2000), indicating that the majority of rumen archaea appear to be methanogens.

Methanogenic archaea depend solely on methanogenesis for energy generation, and can be classified into three categories based on the mechanism used. i) Hydrogenotrophic methanogens are predominant in the rumen where they use H_2 (or formate) as an energy source and reduce CO_2 to CH_4. This use of H_2 prevents its accumulation and an increase of H_2 partial pressure that can inhibit microbial fermentation (Hungate 1967). ii) Methylotrophic methanogens use methylated compounds such as methanol, methylamine and methylthiol as substrates for CH_4 formation. iii) The acetoclastic methanogens are the least common in the rumen, and they use acetate for methanogenesis which is energetically inefficient for the animal (Thauer 1998). It is believed that methanogenic archaea that depend solely on acetoclastic methanogenesis likely grow too slow to persist under normal conditions in the rumen and disappear from the rumen due to the wash out effect (Rowe *et al.* 1979; Hobson and Stewart 1997).

A model of fermentation has been proposed in which the H_2 concentration has a strong influence on the reactions that are thermodynamically favoured within the rumen (Janssen 2010). Figure 1.1 shows the main pathways for VFA and H_2 production and demonstrates how these reactions are influenced by changing ruminal H_2 concentrations.

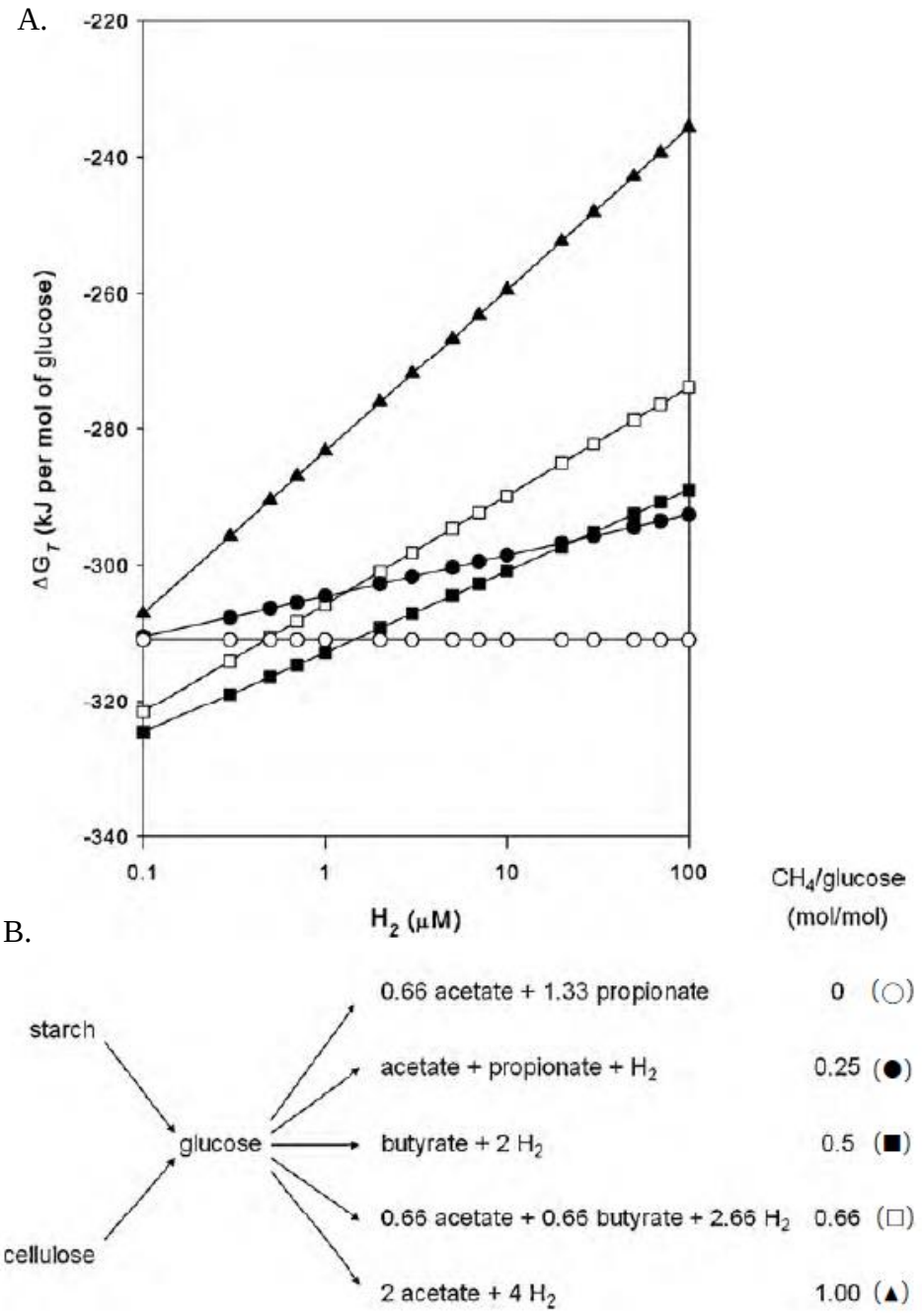


Figure 1.1. The thermodynamic relationship between H_2 concentration and VFA production from cellulose and starch. **A.** Gibbs free energy changes (ΔG_T) of glucose fermentation at different H_2 concentrations. **B.** Possible pathway of cellulose and starch fermentation via glucose to acetate, propionate and butyrate, for simplicity, CO_2, H^+ and H_2O are not shown. This figure is reproduced from Janssen 2010 with permission.

Because most rumen methanogens use H_2, methanogen activity and growth are intricately linked to H_2 producing microbes, such as rumen protozoa and bacteria, whose activity is influenced by the feed eaten by the animal. In NZ, ruminants are fed mainly pasture forages, while conserved forms of pasture forages (silage or hay) or forage crops (maize, brassicas) are also important at various times of the year. Less commonly, grains or mixtures of the previous feeds, are fed. Pasture forages typically contains more cellulose and hemicellulose, compared to grains, and its slower rate of fermentation in the rumen results in more CH_4 formation. Grains are rich in starch and are broken down more readily by rumen microorganisms, producing more propionate when fermented in the rumen (Huntington 1997), acting as an alternative sink for electrons rather than going to CH_4 (Moss *et al.* 2000). The general observation is that the slower the rate of digestion in the rumen, the more CH_4 is formed (Moe and Tyrrell 1979), and that forage has a slower passage rate through the rumen, and the high cellulose content favours the formation of acetate (Janssen 2010).

Understanding the characteristics of rumen methanogens and their complex relationship with other rumen microbes, are therefore essential to formulate strategies to mitigate enteric CH_4 emissions, either directly against methanogens or indirectly against rumen microbes supporting methanogen activities.

1.4. CH_4 mitigation strategies

CH_4 mitigation strategies must consider the entire rumen microbial community and its contribution to the nutrition of the host. Strategies must be viable without reducing animal production or have a negative impact on the farming operation (Clark *et al.* 2011). The products of animals (most commonly milk, meat and wool) must not contain residues of any compound used to reduce CH_4 emissions. It is likely that methanogen species in the rumen occupy different niches and are adapted to grow at different levels of H_2 and/or on different substrates, therefore mitigation strategies must be able to tackle methanogens living under different circumstances. This applies to their physical location as some methanogens are free-living while others are attached to, or live within H_2 producing protozoa (Fenchel and Finlay 1991). Finally, if methanogens are eliminated from the rumen, then an alternative H_2 disposal method must take the place of the methanogenesis, otherwise the ruminant host can suffer from accumulation of H_2, reduced fibre degradation and lower fermentation rates.

There are many CH_4 mitigation strategies that indirectly reduce methanogens by reducing H_2 production, such as redirecting VFA production from acetate to propionate (McAllister and Newbold 2008), or removal of H_2 producing protozoa and cellulolytic bacteria (Morgavi *et al.* 2012). Low fibre, high starch grain diets such as barley, result in faster digestion time and reduced energy loss via CH_4 from 6.5% to 3% (Beauchemin and McGinn 2005). While this feeding strategy is cost effective in countries with abundant supply of cheap grains, it is not applicable in NZ, which depends heavily on forage grazing. Feed additive such as canola oil, coconut oil, sunflower oil have also been shown to reduce CH_4 production (McGinn *et al.* 2004), as fatty acids may act as alternative H_2 sinks (Johnson and Johnson 1995), however this method is not ideal as it lowers the animal performance by decreasing fibre digestibility (Beauchemin and McGinn 2006). Application of essential oils as feed additives have reported similar findings (Patra and Yu 2012).

Many feed additives have been tested for CH_4 mitigation. Antibiotic ionophores such as monensin, inhibit protozoa and cellulolytic bacteria (Hino *et al.* 1993), therefore the introduction of monensin as a feed additive is thought to be able to reduce H_2 production and as a result mitigate CH_4 production by up to 10% (Beauchemin *et al.* 2008). However, usage of antibiotics has been banned by the European Union, because of the risk that antibiotics could get absorbed from the rumen and contaminate milk and meat (Pugh 2002). Furthermore, the reductions in CH_4 have been observed to be transient (Guan *et al.* 2006), and prolonged use of antibiotics could potentially contribute to antibiotic resistance. Organic acids, such as fumarate and malate, can be used as feed additives to redirect bacterial VFA production from acetate to propionate, as many rumen bacteria are known to produce propionate via a reverse citric acid cycle, of which malate and fumarate are key intermediates in the pathway (Martin 1998). Despite the reduction of CH_4 production by fumarate and malate observed *in vitro* (Newbold *et al.* 2005; Lin *et al.* 2013), the *in vivo* trials thus far remain inconclusive (Beauchemin and McGinn 2006; Foley *et al.* 2009). Cashew nut shell liquid has also been tested as a feed additive as it contains high amounts of anacardic acid, which inhibits Gram-positive bacteria (Kubo *et al.* 2003), redirecting rumen fermentation to propionate production and reducing CH_4 production, although reduced digestibility has been observed (Shinkai *et al.* 2012).

Protozoa are one of the primary H_2 producers in the rumen, and methanogens have been shown to be endo- and ectosymbiont of protozoa (Finlay *et al.* 1994). Endosymbiotic methanogens within protozoa could account for 10% to 20% of rumen methanogens (Chagan *et al.* 1999). Therefore removal of protozoa could be a viable strategy for CH_4 mitigation. The removal of

protozoa from the rumen is known as defaunation and several different methods exist, including chemical drenching and diet alteration to lower the pH in the rumen. The *in vivo* results for reducing CH_4 via defaunation thus far are inconsistent and it is debatable whether defaunation is capable of reducing CH_4 emissions (Bird *et al.* 2008). Defaunation is likely unsuitable in NZ as a CH_4 mitigation option, because pasture-fed animals tend to have lower number of protozoa (Towne *et al.* 1990; Silva *et al.* 2014), and it would be very difficult to prevent re-faunation of the rumen from natural sources.

Another method to reduce protozoa is the incorporation of plant secondary metabolites, such as saponins and tannins, as feed additives. Saponins are natural detergents with emulsification properties and are abundant in tea leaves. Saponins significantly reduce protozoa populations (Guo *et al.* 2008) and possibly impair protozoal membranes by forming complexes with sterols (Wallace *et al.* 2002). Overall, the effect of saponin addition on CH_4 reduction has been inconsistent (Jayanegara *et al.* 2014). Tannins are a class of polyphenols of high molecular weight, and are abundant in tropical legumes. The effect of tannins on protozoa and methanogens is inconsistent. Sheep feed on plants containing high levels of tannins have up to 30% reduction in CH_4 production without adverse effects (Carulla *et al.* 2005; Tavendale *et al.* 2005), and both methanogens and protozoa significantly declined as the concentration of tannin was increased in the feed (Tan *et al.* 2011). However, some studies have found tannins to have no effect on CH_4 reduction (Beauchemin *et al.* 2007). Therefore despite increasing the proportion of tannin-rich plants in forage being applicable in NZ, it would require more research with statistical support to ascertain its viability as a CH_4 mitigation option.

The traditional method of selective animal breeding could also reduce CH_4 emission, if CH_4 emissions are a heritable trait. Selecting for feed conversion efficiency and productivity may reduce CH_4 emissions per unit of product formed. Changes in farming practices, such as extending lactation in the dairy industry, could reduce the energy demand from herd and reduce overall emissions (Eckard *et al.* 2010; Buddle *et al.* 2011).

The anaerobic cultivation technique developed by Robert Hungate in 1947, allowed the cultivation of rumen methanogens *in vitro,* yet it remains difficult to cultivate methanogens *in vitro* even today. This has stunted the development of CH_4 mitigation strategies directly targeting rumen methanogens. The advancement in sequencing technologies has enabled culture-independent methods to evolve, and has made CH_4 mitigation strategies directly targeting rumen methanogens possible. One example is the use of methanogen genomic

information to develop vaccines that specifically target rumen methanogens (Attwood *et al.* 2008; Buddle *et al.* 2011). This approach relies on the use of specific methanogen proteins to elicit a secretory antibody response in the saliva of ruminants. Cattle secrete approximately 1.5 to 2.5 rumen volumes of saliva per day (Bailey and Balch 1961), thus a vaccinated animal would secrete antibodies against methanogens via its saliva. This would achieve a sustained inhibition and prevent re-colonization of rumen methanogens (Subharat *et al.* 2015), providing a practical approach for grazing ruminants in NZ. Another example is the development of small molecule inhibitors, via a chemogenomics approach, targeting pathways and enzymes unique to rumen methanogens. However this approach needs to overcome the dilution effect of rumen as well as requirement to continuously deliver the inhibitor into the animal to prevent re-colonization of rumen methanogens (Leahy *et al.* 2010; Aung *et al.* 2015).

The ideal targets for vaccination and inhibitors are genes conserved across all rumen methanogens which are absent from the animal host and other members of the rumen microbiome (Attwood *et al.* 2008). All rumen methanogens need to be targeted because if any are not inhibited, they would be able to quickly adapt and populate the vacant niche. One possible target is the methyl CoM reductase gene, which is inhibited by the CoM analogue inhibitor, 2-bromoethanesulfonic acid (BES) (Immig *et al.* 1996). The genome sequences of rumen methanogens can improve our understanding of their biology and identify conserved genes that may serve as targets for developing effective CH_4 mitigation technologies.

Methanogens are not the only H_2 utilisers in the rumen, and in the absence of methanogens the rumen microbiome would require an alternative H_2 sink so the host digestion would not be impaired. Homoacetogens are organisms that consume H_2 and produce acetate, acting as alternative H_2 sink and increasing the available energy to the host by 4% to 12% (Joblin 1999). However, the H_2 level in the rumen strongly favours methanogenesis over acetogenesis due to the low H_2 threshold and the energy generated under these conditions. The energy yield from methanogenesis is four times higher than acetogenesis (Thauer 1998; Muller 2003), therefore in the natural rumen environment, homoacetogens cannot compete with methanogens for H_2. Despite the abundance of strategies in CH_4 mitigation, our lack of understanding of the rumen microbiome limits the capability to develop effective long-term strategies.

1.5. Rumen methanogens

In order to formulate effective strategies of CH_4 mitigation, it is necessary to understand the broadness of diversity of rumen methanogens as well as their variety of metabolic processes. Methanogenic archaea belong to the Euryarchaeota phylum, which consists of seven orders; the Methanobacteriales, Methanomicrobiales, Methanosarcinales and Methanomassiliicoccales are found in the rumen as well as other environments, while the Methanocellales, Methanococcales and Methanopyrales are found outside of the rumen (Liu and Whitman 2008; Sakai *et al.* 2008; Iino *et al.* 2013). There are at least 28 genera and 113 species of methanogens (Garrity *et al.* 2007).

A survey of published 16*S* rRNA gene sequences has shown that rumen methanogenic archaea form nine separate clades (Janssen and Kirs 2008). Globally, the majority (92.3%) of rumen archaeal sequences fall within three main methanogen groups: *Methanobrevibacter*, *Methanomicrobium* and the newly established Methanomassiliicoccales order. *Methanobrevibacter* is the predominant genus of rumen methanogens, accounting for 61.6% of rumen archaea. Around 33.6% of archaeal 16*S* rRNA gene sequences fall within the *Mbb. gottschalkii* clade and 27.3% in the *Mbb. ruminantium* clade (Janssen and Kirs 2008). Members of Methanomassiliicoccales order account for 15.8% of global rumen archaeal population, and can make up to 85% of rumen methanogens in particular ruminants (Huang *et al.* 2012), while *Methanomicrobium* account for 14.9% of the rumen archaeal population. The remaining low abundance groups in the rumen includes *Methanomicrococcus* spp., *Methanobacterium* spp., and *Methanosarcinales* spp. (Yanagita *et al.* 2000; Shin *et al.* 2004; Wright *et al.* 2007; Janssen and Kirs 2008). In NZ ruminants, the Methanobacteriales account for 89.6% of rumen methanogens, of which the *Mbb. gottschalkii* clade accounts for 42.4% and *Mbb. ruminantium* clade accounts for 32.9% (Seedorf *et al.* 2015). Methanomassiliicoccales account for 10.4% of methanogens in NZ ruminants and is the second most abundant order (Seedorf *et al.* 2015).

Only a handful of rumen methanogens have been isolated and grown in pure cultures due to difficulties in culturing these strict anaerobes. The majority of cultures belong to the predominant genus, *Methanobrevibacter* (*Mbb.*), and display morphologies of short rods with a tendency to form pairs, chains or irregular clumps (Fournier *et al.* 2011). The first rumen methanogen described was *Mbb. ruminantium,* and was isolated in 1958 using serial dilution in a rumen simulating medium (Smith and Hungate 1958). The original strain was lost, but was re-isolated in 1965 (Bryant 1965), and subsequent verification and phylogenomic study

renamed this species as *Mbb. ruminantium* $M1^T$ ($M1^T$) (Balch *et al.* 1979). Subsequently, several species and strains belonging to the *Methanobrevibacter* genus have been isolated from a variety of ruminant sources. A member of the *Mbb. wolinii* clade, *Methanobrevibacter* sp. AbM4 (AbM4), has been isolated from ovine abomasum (Simcock *et al.* 1999). Four formate utilizing methanogens were isolated in 2007, all belonging to the genus, *Methanobrevibacter* (Rea *et al.* 2007). Three isolates were described as strains of *Mbb. olleyae*, two of which were from a bovine rumen source and one from the ovine rumen. The fourth isolate was described as *Mbb. millerae* from the bovine rumen. A member of the *Mbb. ruminantium* clade, *Mbb. olleyae* YLM1 (YLM1), and representatives of the *Mbb. gottschalkii* clade, *Methanobrevibacter* sp. D5 (D5), and *Mbb. millerae* SM9 (SM9) have been isolated from ovine rumen (Kim 2012). Three strains of *Mbb. boviskoreani* have been isolated from bovine rumen in South Korea (Lee *et al.* 2013).

Other groups of methanogens, such as *Methanomicrococcus*, *Methanosphaera* and *Methanobacterium* were either less abundant or were not detected in all studies (Yanagita *et al.* 2000; Shin *et al.* 2004; Wright *et al.* 2007; Janssen and Kirs 2008). The family *Methanobacteriaceae*, which includes the genera *Methanobrevibacter, Methanobacterium,* and *Methanosphaera*, composed from 30% to 99% of rumen archaea, while the family Methanomicrobiales account for 0 to 54% of rumen archaea. *Methanobacterium formicium,* and *Methanomicrobium mobile* have been isolated as pure cultures from the bovine rumen (Jarvis *et al.* 2000). The family Methanosarcinales were a small minority that ranged from 2% to 3% of rumen archaea (Janssen and Kirs 2008), and similar conclusions have been drawn from fluorescence *in situ* hybridization studies (Soliva *et al.* 2003). *Methanosarcina barkeri* have been isolated from the bovine rumen as pure cultures (Jarvis *et al.* 2000). *Methanosphaera* is present in the rumen of deer, sheep and cattle, and studies have found there are two major clusters of *Methanosphaera* spp., a particular strain ISO3-F5 has been isolated from ovine rumen and cultivated *in vitro* (Jeyanathan *et al.* 2011).

CH_4 mitigation strategies often involves *in vitro* experimental validation on methanogen cultures, such as testing potential methanogen-inhibitory compounds, therefore the scarcity of pure cultures of rumen methanogens is a limitation, making it difficult to validate whether the strategies could effectively inhibit all methanogens within the rumen. This also highlights the importance of DNA sequencing efforts in the formulation of CH_4 mitigation strategies.

1.6. Methanogenesis

Methanogenesis is the sole energy generating mechanism of rumen methanogens, therefore conserved enzymes in the methanogenesis pathways would make suitable targets for methanogen inhibitors. The pathways of hydrogenotrophic, acetoclastic, methylotrophic methanogenesis as well as the methanogenesis pathway of Methanomassiliicoccales are displayed in Figure 1.2 (Thauer 1998; Borrel *et al.* 2014).

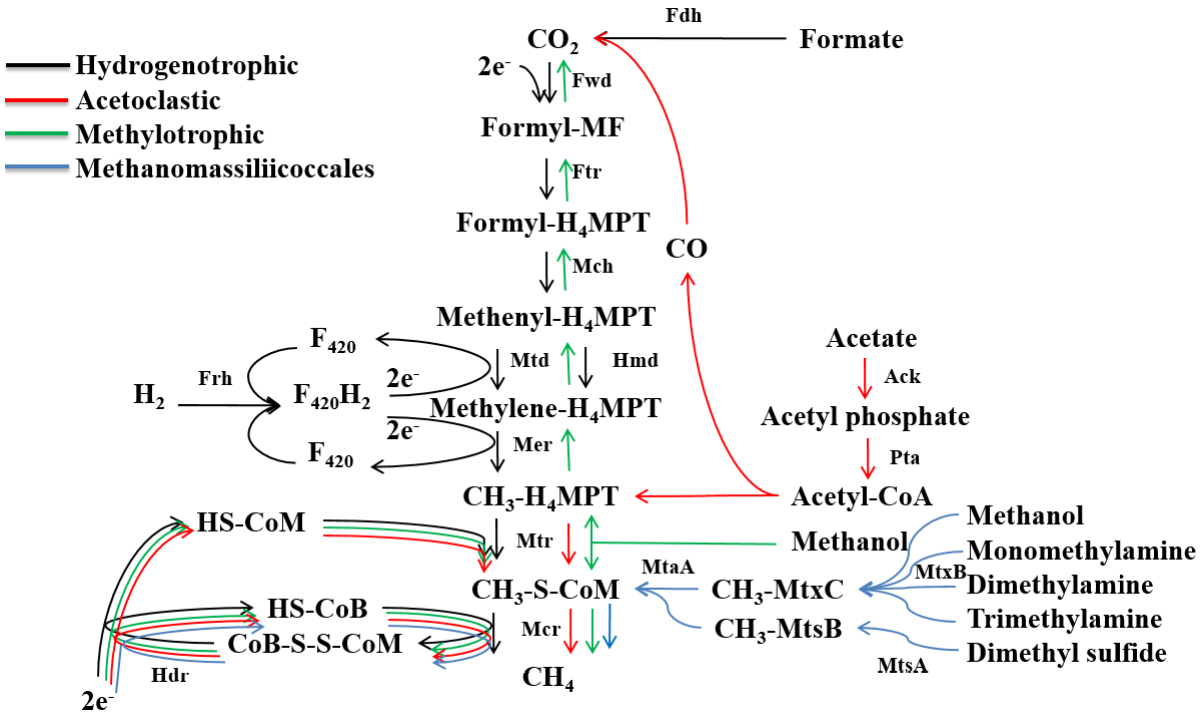


Figure 1.2. Methanogenesis pathways. nicotinamide adenine dinucleotide (NAD)-dependent formate dehydrogenase (Fdh), formylmethanofuran dehydrogenase (Fwd), formylmethanofuran (MF), formylmethanofuran:H_4MPT formyltransferase (Ftr), tetrahydromethanopterin (H_4MPT), methenyl- H_4MPT cyclohydrolase (Mch), 8-hydroxy-f-deazaflavin (F_{420}), F_{420}-reducing hydrogenase (Frh), F_{420}-dependent methylene-H_4MPT dehydrogenase (Mtd), H_2-forming methylene-H_4MPT dehydrogenase (Hmd), F_{420}-dependent methylene-H_4MPT reductase (Mer), coenzyme M (CoM), 2-sulfanylethanesulfonate (CoM), methyl-H4MPT: CoM methyltransferase (Mtr), methyl: CoM reductase (Mcr or Mrt), acetate kinase (Ack), phosphoacetyl transferase (Pta), methanol/mono- , di-, tri-methylamine corrinoid methyltransferase (MtxB), methanol/mono- , di-, tri-methylamine corrinoid protein (MtxC), methylthiol:corrinoid methyltransferase (MtsA), methylthiol:corrinoid protein (MtsB), methyl:coenzymeM methyltransferase (MtaA), heterodisulfide reductase (Hdr), coenzyme B (CoB), 7-mercaptoheptanoylthreoninephosphate (CoB) (Thauer 1998).

Hydrogenotrophic methanogenesis is the most common methanogenic pathway observed in rumen methanogens, and is utilised by the predominant *Methanobrevibacter* spp. (Leahy *et al.* 2010; Leahy *et al.* 2013). In the hydrogenotrophic pathway, CO_2 is reduced to CH_4 through a series of enzymatic steps as illustrated in Figure 1.2. The initial reduction of CO_2 with the C_1 carrier methanofuran (MF) to formylmethanofuran by formylmethanofuran dehydrogenase (Fwd) is driven by a sodium gradient in *Methanothermobacter thermoautotrophicus* (de

Poorter *et al.* 2003). The tetrahydromethanopterin (H_4MPT) is the second C_1 carrier, and the transfer of the formyl group is catalysed by formylmethanofuran:H_4MPT formyltransferase (Ftr) (Shima *et al.* 2002). The formyl-H_4MPT is then reduced to methyl-H_4MPT via methenyl-H_4MPT cyclohydrolase (Mch), F_{420}-dependent methylene-H_4MPT dehydrogenase (Mtd), and F_{420}-dependent methylene-H_4MPT reductase (Mer) (Shima *et al.* 2002). The Mtd and Mer depend on reducing potential supplied from H_2 via F_{420}-reducing hydrogenase (Frh). The Frh depends on the F_{420} (Cheeseman *et al.* 1972), which is so named due to its autofluorescence at 420 nanometer (nm). This property is particularly useful in enabling visualisation of rumen methanogens under ultra violet (UV) illumination (Doddema and Vogels 1978). The methyl group of methyl-H_4MPT is transferred to CoM via the transmembrane methyl-H_4MPT: CoM methyltransferase (Mtr), which drives a sodium gradient (Gottschalk and Thauer 2001). The membrane potential generated from Fwd and Mtr contribute to the membrane potential involved in energy conservation. The formation of methyl-CoM, the recycling of CoM and the formation of CH_4 is conserved with the other methanogenesis pathways. In the acetoclastic pathway, acetate is first converted to acetyl-Coenzyme A (CoA) by acetyl-CoA synthetase or by acetate kinase (Ack) and phosphoacetyl transferase (Pta). Carbon monoxide (CO) dehydrogenase then cleaves acetyl-CoA, releases CO and transfers the methyl group to H_4MPT, entering the methanogenesis pathway at Mtr. The CO is oxidized by ferredoxin (Fdx) and H_2, producing CO_2, which feeds into the methanogenesis pathway (Fournier and Gogarten 2008).In methylotrophic methanogenesis, the methyl group from different C_1 substrates is transferred to CoM and enters the methanogenesis pathway at Mcr. Members of the orders Methanosarcinales and Methanomassiliicoccales and the genus *Methanosphaera* are capable of carrying out methylotrophic methanogenesis. The Methanosarcinales is the only order capable of disproportionating methanol to CH_4 and CO_2 as depicted in Figure 1.2 (Keltjens and Vogels 1993). Members of *Methanosphaera* are dependent on methylotrophic methanogenesis and are incapable of disproportionating methanol to CO_2 or utilising CO_2 for methanogenesis. The genome sequence of *Methanosphaera stadmanae* confirmed this is due to absence of genes encoding the molybdenum cofactor biosynthesis complex, which is important in producing the essential cofactor to the formylmethanofuran dehydrogenase in the hydrogenotrophic methanogenesis pathway (Fricke *et al.* 2006). Members of the order Methanomassiliicoccales also depend solely on methylotrophic methanogenesis for survival, but differ from the *Methanosphaera* in that they do not possess any of the genes encoding the hydrogenotrophic part of the methanogenesis pathway (Borrel *et al.* 2014) (Figure 1.2.). Therefore the Mcr/Mrt is the only enzyme shared by all the known methanogenesis pathways in rumen methanogens.

Inhibitors affecting this enzyme have been tested, including the CoM analogue BES and the CH_4 analogue, chloroform. BES inhibition of methanogens in the rumen is temporary, followed by quick adaptation of the methanogen community (Immig *et al.* 1996), while chloroform is toxic to the host animal (Lanigan *et al.* 1978).

In the methyltrophic methanogenesis pathway, the utilisation of each substrate involves three proteins carrying out transfer of methyl groups from substrates to the CoM. These include the substrate:corrinoid methyltransferases MtaB, MtmB, MtbB, MttB, MtsA for methanol, mono-, di-, tri- methylamine and methylthiol respectively, and the corrinoid proteins MtaC, MtmC, MtbC, MttC, MtsB for methanol, mono-, di-, tri- methylamine, and methylthiol respectively (Burke and Krzycki 1997; Ferguson and Krzycki 1997; Sauer and Thauer 1998; Ferguson *et al.* 2000; Bose *et al.* 2008). The methanol: CoM methyltransferase, MtaA, is shared by different substrates, except for methylthiol which uses the MtsA, a bifunctional enzyme which carries out both the transfer of methyl group from substrate onto corrinoid protein as well as the transfer from the corrinoid protein onto CoM (Tallant and Krzycki 1997; Tallant *et al.* 2001). The corrinoid proteins depend on the cobalt ion in their active sites to be in a reduced state, and a methyltransferase-activating protein, RamA, is responsible for activation of the corrinoid protein (Wassenaar *et al.* 1996). The remainder of the methylotrophic pathway is conserved with the hydrogenotrophic and acetoclastic methanogenesis pathways.

The formation of methyl-CoM is the point of convergence of methanogenesis pathways. The methyl- CoM reductase is responsible for the final reduction of the methyl group to CH_4. The functional components were first elucidated from *Methanobacterium thermoautotrophicum* (Gunsalus and Wolfe 1980). Two isozymes of this enzyme have been observed in *M. thermoautotrophicum* ΔH (Rospert *et al.* 1990), Mcr (MCR I) and Mrt (MCR II). Both isozymes are composed of α, β and γ subunits encoded by *mcrA*, *mcrB* and *mcrG* respectively, yet the isozymes differ in their operon arrangements, *mcrBDCGA* and *mrtBDGA*. The isozymes are differentially regulated by growth condition, with Mcr up-regulated during growth on limiting H_2, and Mrt up-regulated in abundant H_2 (Pihl *et al.* 1994; Reeve *et al.* 1997).

Not all members of Methanobacteriales possess both isozymes; *Mbb. ruminantium* $M1^T$ possess only the Mcr, which suggested it has adapted to a lifestyle at low H_2 availability (Leahy *et al.* 2010). The functional methyl-CoM reductase enzyme is a hexamer consisting of subunits Mcr/$MrtA_2B_2G_2$ with nickel containing cofactor F_{430} (F_{430}) in the active site. The function of

the McrCD and MrtD subunits remain elusive, but it has been proposed that their role may be in post-translational modification of Mcr/MrtA (Ellefson *et al.* 1982; Ermler *et al.* 1997).

The reduction of methyl-CoM to CH_4 is coupled to the formation of the heterodisulfide, CoM-S-S-CoB, which is then reduced by the heterodisulfide reductase complex (Hdr) to produce CoM-SH and allow it to be reused in methanogenesis (Noll *et al.* 1986; Hedderich *et al.* 1990; Hedderich *et al.* 2005). Two components of Hdr exists, the cytoplasmic HdrABC and the transmembrane HdrDE, which was found in *Methanosarcina barkeri*, with the HdrE being identified as a *b*-type cytochrome (Heiden *et al.* 1994).

The reducing equivalents required for heterodisulfide reduction come from H_2. HdrABC gains reducing potential either from Frh or from methylviologen hydrogenase (Mvh). As indicated above, Frh supplies reducing potential from H_2 by reducing F_{420}. The Mvh complex, MvhADG, does not require F_{420} to function. HdrABC has been observed to form a complex with MvhADG in *Methanobacterium thermoautotrophicum* (strain Marburg); the complex was found to be loosely membrane bound and proposed to be associated with an integral membrane proton pump (Fiebig and Friedrich 1989; Setzke *et al.* 1994). In *Methanosarcina barkeri*, the transmembrane HdrDE has been found as a complex with membrane-associated energy-converting hydrogenase (EchABCDEF), which reduces Fdx with H_2 (Kunkel *et al.* 1998; Hedderich 2004). The periplasmic methanophenazine-dependent hydrogenase, Vht and Vho was also found in *Methanosarcina mazei* in addition to the Ech complex; and differentially expressed dependent on utilisation of acetate (Guss *et al.* 2009).

The hydrogenases not only supply reducing potential to methanogenesis, they are also an important feature linking methanogenesis to energy conservation. The $F_{420}H_2$, reduced Fdxs or cytochromes produced by hydrogenases act in conjunction with proton pumps to link methanogenesis with membrane potential generation, which enables adenosine triphosphate (ATP) synthase to generate energy (Deppenmeier *et al.* 1999; Deppenmeier 2002). The $F_{420}H_2$ dehydrogenase complex (Fpo) is a large transmembrane proton pump; in *Methanosarcina mazei* it is a key enzyme that generates membrane potential by electron transport from $F_{420}H_2$ (Baumer *et al.* 2000; Welte and Deppenmeier 2011), as well as being linked to Hdr by electron transporting methanophenazine (Tietze *et al.* 2003). The membrane potential required for energy generation can be Na^+ or H^+. In the hydrogenotrophic methanogens, the Mtr complex generates Na^+ gradient (Lienard *et al.* 1996). The Fwd is driven by the Na^+ gradient in hydrogenotrophic methanogenesis (Kaesler and Schonheit 1989; de Poorter *et al.* 2003), and

may act in reverse to generate a Na^+ gradient when the methyl substrate is disproportionated in methylotrophic methanogenesis (Deppenmeier *et al.* 1999). The H^+ gradient is generated by the Fpo complex powered by the Hdr and Mvh (Deppenmeier *et al.* 1999; Schäfer *et al.* 1999; Deppenmeier 2002). In *Mbb. ruminantium* $M1^T$, Na^+ is utilized by the A_1A_0-ATP synthase for ATP generation, but under conditions of low pH and low Na^+, a H^+ gradient could also be used to generate ATP (McMillan *et al.* 2011).

1.7 Methanogen Genomes

Advances in next generation DNA sequencing technologies has made genome sequencing a rapid and economical research method (Grada and Weinbrecht 2013). Methanogen genomic information is useful in investigating ruminant CH_4 mitigation strategies, and one of the strengths of this approach is the ability to compare between genomes to identify which genes are conserved among particular types of methanogens and also which genes define the unique features of each microorganism. Difficulties in the cultivation and purification of rumen methanogens has hindered their widespread genome sequencing. However, the genomes of several cultivated rumen methanogens have been sequenced (Table 1.1.). The methanogen genome sequences currently completed range in size from 1.5 Mb to 4.5 Mb, with most between 1.5 Mb and 3 Mb. The genome of *Methanosarcina barkeri* CM1 is larger than other methanogen genomes. The larger genome size of members of the Methanosarcinales is correlated with their ability to form CH_4 via hydrogenotrophic, acetoclastic and methylotrophic pathways; and are thus capable of more complex metabolic capacities (Maeder *et al.* 2006).

Methanobrevibacter ruminantium $M1^T$ was the first genome sequence of a rumen methanogen (Leahy *et al.* 2010). Like other hydrogenotrophic methanogens, it has a small genome of 2.9 Mb with a %G+C of 33%, and it encodes each of the seven steps in the methanogenesis pathway using H_2 + CO_2. In hydrogentrophic methanogens, the last two steps of the methanogenesis pathway are strongly conserved, which includes the CoM methyltransferase and the methyl CoM reductase genes. Several methanogenesis marker genes, currently without any ascribed function, are also highly conserved (Gao and Gupta 2007; Liu and Whitman 2008). The $M1^T$ genome sequence explains why it needs acetate, 2-methylbutyrate and CoM to survive, as it lacks several genes involved in the use or synthesis of these compounds. Insights from the analysis of the $M1^T$ genome have led to the optimization of *Mbb. ruminantium* growth. The presence of two copies of nicotinamide adenine dinucleotide phosphate (NADP)-dependent alcohol dehydrogenase within the genome suggested possible

alcohol utilisation and it was found that methanol and ethanol could stimulate, but not support growth. Subsequently methanol has been added to media for growth of $M1^T$. The $M1^T$ genome sequence also lead to the discovery of a prophage sequence, designated φmru, and two non-ribosomal peptide synthase (NRPS) genes, which have never been found previously in archaea (Leahy *et al.* 2010). Several more rumen Methanobacteriales genomes have been sequenced, including AbM4, *Mbb. boviskoreani* $JH1^T$ ($JH1^T$), SM9, YLM1, *Mbb. wolinii* SH^T (SH^T), *Mbb. millerae* $ZA\text{-}10^T$ ($ZA\text{-}10^T$), *Methanosphaera* sp. ISO3-F5, *Methanobacterium formicicum* BRM9. Several other rumen methanogen genome sequencing projects are underway as indicated in Table 1.1, including *Mbb. millerae* HW02 and *Methanobrevibacter* sp. YE315 (YE315).

Table 1.1. Rumen methanogen genomes from Genomes OnLine Database (GOLD).

Strain Name	GOLD ID	Order	Genome Size (Mb)	GC (%)	ORFs	Completion	Reference
Methanosphaera sp. ISO3-F5		Methanobacteriales	2.8	31	2354	In Progress	
Methanobrevibacter ruminantium M1^T	Gc01203	Methanobacteriales	2.9	33	2283	Complete	Leahy *et al.*, 2010
Methanobrevibacter millerae SM9	Gi06992	Methanobacteriales	2.5	32	2321	Complete	Kelly *et al.*, 2016
Methanobrevibacter sp. YE286		Methanobacteriales				Draft	
Methanobacterium formicicum BRM9	Gi06999	Methanobacteriales	2.4	41	2352	Complete	Kelly *et al.*, 2014
Methanobacterium sp YE299		Methanobacteriales				Draft	
Methanobrevibacter olleyae YLM1	Gi07000	Methanobacteriales	2.2	27	1862	Draft	Kelly *et al.*, 2016
Methanobrevibacter sp. AbM4	Gi17672	Methanobacteriales	2	29	1730	Complete	Leahy *et al.*, 2013
Methanobrevibacter millerae HW02	Gi0074693	Methanobacteriales				In Progress	
Methanobrevibacter sp. YE315	N/A	Methanobacteriales				In Progress	
Methanobrevibacter wolinii SH^T	Gi0054485	Methanobacteriales	2	24	1736	Draft	
Methanobrevibacter millerae DSM 16643	Gi0070837	Methanobacteriales	2.7	37	2467	Draft	
Methanobrevibacter boviskoreani JH1^T	Gi39333	Methanobacteriales	2.1	29	1774	Draft	Lee *et al.*, 2013
Thermoplasmatales archaeon BRNA1	Gi0052360, Gc0052360	Methanomassiliicoccales	1.5	58	1577	Complete	
Methanosarcina sp. CM1	Gi06991	Methanosarcinales	4.5	39	3655	Complete	Lambie *et al.*, 2015

Genomes with N/A (not available) are genome is in early stage of sequencing and no GOLD ID were assigned. Genomes without a GOLD ID were acquired from Morgavi *et al*. (2012).

1.8. Research questions

Genome sequencing of rumen methanogens provides detailed knowledge of their genetic makeup and provides insights into their metabolism via reconstruction of their metabolic pathways from bioinformatic analyses. However, many different types of methanogens exist within the rumen, and ruminant CH_4 mitigation strategies should target all rumen methanogens and not just a few species. Therefore, in order to assess the genomic diversity of methanogens in the rumen, multiple methanogen genomes selected from all available taxonomic levels (order, family, genera, species and strain), should be sequenced and compared. My PhD project seeks to contribute to this effort by acquiring detailed knowledge of rumen methanogens through genome sequencing and comparative bioinformatic analyses of their genomes.

At the beginning of my PhD project, only the *Mbb. ruminantium* $M1^T$ genome had been sequenced and was publicly available (Leahy et al 2010) and my interests were in a particular group of recently discovered methanogens called Rumen Cluster C (RCC, now classified as members of the newly formed methanogen order, Methanomassiliicoccales) and in strain level variations in *Methanobrevibacter* genome sequences. A pure culture of RCC had never been isolated from a ruminant, therefore our knowledge of the rumen RCCs was very limited. The aim of my PhD work was to obtain a wider representation of the rumen methanogen genomes by sequencing the DNA of a CH_4-producing enrichment culture called ISO4-H5, containing a methanogenic archaeon belonging to the RCC group. In addition, the *Methanobrevibacter* sp. D5 sequence was also examined as part of a *Methanobrevibacter* 'pan-genome' analysis to assess strain level genomic variation in these methanogen species.

My specific research questions are:

- What is the genome composition of the methanogenic archaeon RCC ISO4-H5, and how does this differ from other sequenced methanogens?
- What is the metabolic scheme of RCC ISO4-H5, how does it grow in the rumen, and can these features be used to isolate a pure culture of this organism?
- What is the genome composition of *Methanobrevibacter* sp. D5, how does this differ from other sequenced *Methanobrevibacter* spp. strains and how do these differences allow co-existence of multiple *Methanobrevibacter* spp. in the rumen?

Chapter 2
Materials and methods

2.1. Materials

2.1.1. Media

2.2.1.1. Rumen fluid

Rumen fluid was collected from ruminally-fistulated cows fed hay and the rumen fluid was filtered by cheese cloth and stored at -20 ^{o}C. Rumen fluid was thawed, boiled, cooled on ice, and the precipitate was removed by centrifugation at 10,000 gravitational force (× *g*) for 10 minutes (min) at room temperature in a Sorvall Evolution RC centrifuge (Thermo Fisher Scientific, Waltham, MA, USA) prior to media preparation.

2.2.1.2. Salt solution A

Salt solution A consists of (gram (g)/Litre (L)) NaCl (6), KH_2PO_4 (3), $(NH_4)_2SO_4$ (1.5), $CaCl_2 \cdot 2H_2O$ (0.79) and $MgSO_4 \cdot 7H_2O$ (1.2) in distilled water.

2.2.1.3. Salt solution B

Salt solution B consists of (g/L) $K_2HPO_4 \cdot 3H_2O$ (7.86) in distilled water.

2.2.1.4. Trace element solution (SL10)

The trace element solution was prepared by addition of 10 millilitre (mL) 25% HCl, 1.5 g $FeCl_2 \cdot 4H_2O$, 190 milligram (mg) $CoCl_2 \cdot 6H_2O$, 100 mg $MnCl_2 \cdot 4H_2O$, 70 mg $ZnCl_2$, 6 mg H_3BO_3, 36 mg $Na_2MoO_4 \cdot 2H_2O$, 24 mg $NiCl_2 \cdot 6H_2O$ and 2 mg $CuCl_2 \cdot 2H_2O$ sequentially into a final volume of 1 L of distilled H_2O, and was sterilized by autoclaving.

2.2.1.5. Selenite/tungstate solution

The selenite/tungstate solution consists of (g/L) NaOH (0.5), $Na_2SeO_3 \cdot 5H_2O$ (3 mg) and $Na_2WO_4 \cdot 2H_2O$ (4 mg), and was sterilized by autoclaving.

2.2.1.6. BY media

BY medium consists of rumen fluid (30% [volume/volume (v/v)]), Salt Solution A (17% [v/v]), Salt Solution B (17% [v/v]), and (g/L) $NaHCO_3$ (5), yeast extract (1), 0.5 mL of resazurin (0.1% [weight/volume (w/v)]), trace element solution SL10 (0.1% [v/v]) (Widdel and Pfennig 1981), selenite/tungstate solution (0.1% [v/v]) (Tschech and Pfennig 1984), and L-cysteine-HCl (0.5), made up to 100% with distilled water, pH 6.5. All the components except $NaHCO_3$

and L-cysteine-HCl were mixed thoroughly and boiled, bubbled under 100% O_2-free CO_2 for 20 min immediately after boiled while cooled in ice bath, $NaHCO_3$ and L-cysteine-HCl were mixed into the cooled media. The media was dispensed into 100% O_2-free CO_2-flushed anaerobic vessels, the vessel was flushed for a minimum of 30 seconds (s), 15 min and 30 min for Hungate tubes, serum bottles and Schott bottles respectively. For agar-containing media, bacterial agar (1.5% [w/v]) was dispensed into the vessel, flushed with 100% O_2-free CO_2 prior to media dispension. The vessels were capped and sealed with headspace of 100% O_2-free CO_2, and sterilized by autoclaving for 20 min at 121 ^oC, and stored in the dark at room temperature. Agar-containing media were held at 55 ^oC in a water bath post-autoclaving, and additional substrates were added and mixed prior to plating inside an anaerobic hood.

2.1.2. Media additives

Unless otherwise mentioned, all media additives were prepared with boiled distilled water, bubbled under 100% O_2-free N_2 for 10 min.

2.1.2.1. Vitamin 10× concentrate

The Vitamin 10 concentrate stock solution consisted of (mg/L) 4-aminobenzoate (40), d-(+)-biotin (10 mg), nicotinic acid (100) hemicalcium D-(+)-pantothenate (50), pyridoxamine hydrochloride (150), thiamine chloride hydrochloride (100), cyanocobalamin (50), D,L-6,8-thioctic acid (30), riboflavin (30) and folic acid (10). The solution was dispensed anaerobically into sterile, O_2-free, N_2-filled serum bottle through 22-micrometer (µm) pore size sterile Millex GP filter (Merck Millipore, Billerica, MA, USA). The bottles were wrapped in aluminium foil to protect light sensitive vitamins and stored at 4 ^oC for short term or -20 ^oC for long term. The working stock was diluted 10 times with growth medium.

2.1.2.2. Sodium formate (3M)

Sodium formate (204.03 g/L) was dissolved in 100% O_2-free water, bubbled with 100% O_2 free N_2 for 20 min, sealed with 100% O_2-free N_2 headspace, and sterilized by autoclaving.

2.1.2.3. Sodium acetate (1M)

Sodium acetate.$3H_2O$ (136.08 g/L) was dissolved in O_2 free N_2 water, bubbled with 100% O_2 free N_2 for 20 min, sealed with 100% O_2-free N_2 headspace, and sterilized by autoclaving.

2.1.2.4. Methanol (1 M)/ethanol (1 M)/butanol (1 M)

Methanol (41 mL/L, [v/v]) or analytical grade ethanol (58.4 mL/L, [v/v]) or butanol (91.5 mL/L, [v/v]) was mixed in O_2 free N_2 water, bubbled with 100% O_2 free N_2 for 20 min,

dispensed anaerobically into sterile, O_2 free N_2 filled serum bottle through 22 µm pore size sterile Millex GP filter (Merck Millipore).

2.1.2.5. Glucose/Rhamnose/Arabinose/Galactose (1 M)

Glucose (180.16 g/L)/Rhamnose (164.55 g/L)/Arabinose (150.13 g/L)/Galactose (180.16 g/L) was dissolved in O_2-free water, bubbled with O_2-free N_2 for 20 min, dispensed anaerobically into sterile, O_2-free N_2-filled serum bottles through 22 µm pore size sterile filter (Merck Millipore).

2.1.2.6. CoM (10 mM)

CoM (1.41 g/L) was dissolved in O_2-free water, bubbled with O_2-free N_2 for 20 min, dispensed anaerobically into sterile, O_2-free N_2-filled serum bottles through 22 µm pore size sterile filter (Merck Millipore), and stored in the dark at 4 ^oC.

2.1.2.7. Pectin (10% [w/v])

Apple pectin (100 g/L) was dissolved in O_2-free water, bubbled with O_2-free N_2 for 20 min, sealed with 100% N_2 headspace, sterilized by autoclaving and stored at 4 ^oC. The solution was allowed to stand for a period (≥12 h) to dissolve.

2.1.2.8. Cellulose (5%, [w/v])

Cellulose paper was cut to small pieces with scissors and ground to a powder in distilled water in a ball mill at 4 ^oC, bubbled with O_2-free N_2 for 20 min, sealed in serum bottles with 100% N_2 headspace, and sterilized by autoclaving.

2.1.2.9. Antibiotics

Ampicillin (50 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), gentamycin (20 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), kanamycin (50 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), spectinomycin (100 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) and streptomycin (100 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) were dissolved in O_2-free water. Chloramphenicol (50 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), erythromycin (20 mg/mL, Sigma-Aldrich, St. Louis, MO, USA), tetracycline (10 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) was dissolved in N_2-bubbled ethanol. Rifampicin (12 mg/mL, Sigma-Aldrich, St. Louis, MO, USA) was dissolved in N_2-bubbled methanol. Antibiotics were sterilized by filtration through 22 µm pore size sterile filter (Merck Millipore), sealed with 100% N_2 headspace and stored at -20 ^oC, tetracycline and rifampicin were stored in the dark.

2.1.2.10. Dimethyl sulfide (1 M)/Methyl 3-mercaptopropionate (M3SP, 1 M)/methyl-3-methylthiopropionate (M3MSP, 1 M)

Dimethyl sulfide (73.9 mL/L, [v/v], Sigma-Aldrich)/ M3SP (110.8 mL/L [v/v], Sigma-Aldrich)/ M3MSP (124.61 mL/L, [v/v], Sigma-Aldrich) were injected anaerobically into autoclaved BY media with 100% CO_2 headspace.

2.1.3. Methanogen and bacteria strains

2.1.3.1. Methanogenic archaeon ISO4-H5 enrichment culture and Succinivibrio dextrinosolvens H5

An enrichment culture of a representative of the Methanomassiliicoccales was previously obtained from the rumen contents of a nine year old Romney wether sheep in NZ grazing a ryegrass-clover pasture diet (Jeyanathan 2010). This enrichment culture contained a methanogenic archaeon, designated ISO4-H5 and a Gram-negative bacterium, subsequently identified as being most closely related to *Succinovibrio dextrinosolvens*, and designated as strain H5 (Jeyanathan 2010).

2.1.3.2. Methanobrevibacter sp. D5 and enrichment culture H6

A CH_4-forming enrichment culture H6 was originally obtained from the rumen of sheep J472 on pasture diet (G. Henderson, personal communication). A member of the *Mbb. gottschalkii* clade was isolated from the H6 culture and designated as *Methanobrevibacter* sp. D5 (Kim 2012).

2.1.4. General materials

2.1.4.1. PCR reagents

Type I water used for PCR was purified by Milli-Q Water purification system (Merck Millipore, Darmstadt, Germany). All PCR amplifications were performed in Mastercycler^{TM} Pro PCR system (Eppendorf, Hamburg, Germany) using Platinum® PCR supermix or Platinum® PCR supermix high fidelity (Thermo Fisher Scientific, Waltham, Ma, USA) for large products.

2.1.4.2. Primers

Primers were purchased from Integrated DNA Technologies (Integrated DNA technologies Coralville, IA, USA). The stock solutions are made up to 200 micromolar (µM) with Type I water and stored at -20 ^{o}C. The stock solution was diluted to working concentration of 20 µM with Type I water when required.

2.1.4.3. 50× Tris acetate EDTA (TAE) buffer

50×TAE buffer consists of 2 M Tris, 1 M acetic acid and 50 mM pH 8.0 ethylenediaminetetraacetic acid (EDTA). 1× TAE buffer is prepared by dilution with distilled water.

2.1.4.4. 10× Tris borate EDTA (TBE) buffer

10×TBE buffer consists of Tris (54 g/L), boric acid (27.5 g/L) and 10 mM pH 8.0 EDTA. 0.5× TBE buffer was prepared by dilution with distilled water.

2.1.4.5. 100× Tris EDTA (TE) buffer

100× TE buffer consists of 1 M Tris-HCl, 100 mM EDTA, pH adjusted with HCl/NaOH to suit experimental requirement. The pH range used in this thesis includes pH 7.6 and pH 8.0. 1× TE buffer was prepared by dilution with Type I water.

2.1.5. RNA purification materials

Zirconia/Silica beads (dnature Ltd, Gisborne, NZ) of 0.1 mm and 0.5 mm diameter (1:1) and glassware were baked at 180 ^{o}C for two hours (h) to sterilize and deactivate RNase. All surface and equipments were decontaminated by RnaseZap decontamination solution (Thermo Fisher Scientific) prior experiment. RNase free filtered tips were used. Unless otherwise stated, water used is Nuclease-free water, Ambion® (Thermo Fisher Scientific). Diethylpyrocarbonate (DEPC) treated water is prepared by addition of 0.1% [v/v] DEPC to Type I water and incubated for two h at 37 ^{o}C and autoclaved. RNA extraction buffer A consisted of 200 mM NaCl, 20 mM EDTA (pH 8.0) in Type I water and DEPC treated. 20% [w/v] SDS was dissolved in Type I water and DEPC treated. RNase-free ammonium acetate (5 M), Ambion® and phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5) was supplied by Thermo Fisher Scientific. Purification of RNA was carried out with MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific), DNA removal was carried out with Baseline-ZERO DNase (Illumina Netherlands BV, Eindhoven, Netherlands) and TURBO DNA-free Kit (Thermo Fisher Scientific).

2.2. Methods

2.2.1. Anaerobic cultivation

All methanogens used in this study were cultivated anaerobically in BY media. The cultivation and media preparation techniques were introduced by Robert Hungate (Hungate 1966) with

adaptations by Balch (Balch and Wolfe 1976). The medium was dispersed into vessels flushed by O_2-free CO_2. Vessels used in this study include 16 ×125 mm Hungate tubes (BellCo Glass, Vineland, NJ, USA), high pressure Schott bottles with open top screw caps and butyl rubber stoppers and 125 mL serum bottles (BellCo) with butyl rubber stopper. The media was sterilised using a TOMY SX-700E autoclave (Digital Biology, Tokyo, Japan). Filter-sterilized stock solutions were introduced into media anaerobically and aseptically post-autoclaving.

2.2.1.1. Succinivibrio Spent Pectin Growth Medium Supernatant (SSPGMS)

Pectin solution (1%, [w/v]) was added to the pure culture of *Succinivibrio dextrinosolvens* H5 in BY medium, cultivated for two days at 39 ^oC and stored at -20 ^oC. The pectin-incubated culture was thawed and dispensed anaerobically into ISO4-H5 cultures through 22 µm pore size sterile filters (Merck Millipore).

2.2.1.2. Growth measurement

Methanogen growth was monitored by CH_4 production. The head space gas (0.3 mL) was taken by syringe and measured by gas chromatography using an Aerograph 660 (Varian Associates, Palo Alto, Ca, USA) fitted with a Porapak Q 80/100 mesh column (Waters Corporation, Milford, MA, USA) and a thermal conductivity detector operated at 100 ^oC. The column used N_2 as the carrier gas at 12 cm^3/min at room temperature. Standard curves were constructed for H_2 and CH_4 at standard atmospheric pressure. Bacterial growth was monitored by optical density at 600 nm within Hungate tubes via a spectrophotometer (Spectronic 200, Thermo Fisher Scientific).

2.2.1.3. Standard sub-culture conditions

The required substrates and supplements (Table 2.1.) were added to the medium, then inoculated with 10% of culture and pressurized with high-pressure H_2:CO_2 (4:1, 180 kilopascal (kPa)). The maximum period of culture viability is approximately two weeks for methanogenic archaeon ISO4-H5 and *S. dextrinosolvens* H5, and one week for *Methanobrevibacter* sp. D5. Culture viability was achieved by sub-culturing to fresh medium. An additional tube of culture with glucose (10 mM) was used to identify contamination in the ISO4-H5 pure culture and *Methanobrevibacter* sp. D5 culture with microscopy of culture wet mounts.

Table 2.1. The final concentrations or volumes of additives to a 10 mL culture of BY medium in Hungate tubes

Additives	**Methanogenic archaeon ISO4-H5 enrichment culture**	**Methanogenic archaeon ISO4-H5 pure culture**	***Succinivibrio dextrinosolvens* H5**	***Methanobrevibacter* sp. D5**
Sodium formate	60 mM	60 mM	-	60 mM
Sodium acetate	20 mM	20 mM	-	20 mM
Methanol	20 mM	20 mM	-	20 mM
Vitamin mix (working stock)	0.1 mL	0.1 mL	-	0.1 mL
CoM	10 µM	10 µM	-	10 µM
Glucose	-	-	10 mM	-
SSPGMS*	-	1 mL	-	-

* improved growth, not mandatory for culture survival

2.2.2 Microscopy

2.2.2.1 Light microscopy

A drop of culture was placed onto the microscope slide and covered with a cover slip. Excess liquid was removed by pressing the cover slip with a paper towel. The *Methanobrevibacter* sp. D5 was observed by fluorescent microscopy (420 nm) in combination with phase contrast microscopy (DM2500 microscope, Leica Microsystems, Wetzlar, Germany). Methanogenic archaeon ISO4-H5 was observed under phase contrast microscopy only, as ISO4-H5 has not been observed to fluoresce.

2.2.2.2. Transmission electron microscopy

Electron microscopy (Philips CM10 Transmission Electron Microscope with SIS Morada high-resolution digital imaging, Netherlands) was carried out at the MMIC, Massey University, Palmerston North, NZ.

Negative staining transmission electron microscopy procedure was identical to previously described (Jeyanathan 2010) with minor adaptation. Cells were collected by centrifugation (5,000 × *g*, 5 min) and resuspended with anaerobic water. One drop of cell suspension was placed onto carbon-coated formvar film attached to copper grids and left for 2-6 min. Excess sample was drained from grids with filter paper. The grids were placed on a drop of 2% uranyl acetate (in water) resting on parafilm and left for 6-10 min. Excess liquid was drained well and allowed to dry. The whole cells were then observed by electron microscopy at 60 kilovolts (kV).

2.2.2.3. Thin section transmission electron microscopy. Cell pellet was collected by centrifugation (5,000 × *g*, 5 min), the cell pellet was fixed by 3% (v/v) glutaraldehyde in 0.1 M Na_2HPO_4, KH_2PO_4 buffer (pH 7.2) and letting it stand for 2 h at room temperature. The cell suspension was pelleted by centrifugation (4,500 × *g*, 5 min). The cells were rinsed in 0.1 M

Na_2HPO_4, KH_2PO_4 buffer three times for 5, 10 and 15 min, and fixed in 1% (w/v) osmium tetroxide (OsO_4) in the same buffer for 1 h at room temperature. The cells were rinsed again in 0.1 M Na_2HPO_4, KH_2PO_4 buffer three times for 5, 10 and 15 min, followed by dehydration in increasing concentrations of acetone consisting of 25% (v/v), 50% (v/v), 75% (v/v), 95% (v/v) and three changes at 100% (v/v), each step was for 15 min. The dehydrated cells were mixed with 50/50 acetone/resin (Procure 812: ProSciTech, Qld, Australia) (v/v) on a stirrer overnight to allow the acetone to evaporate slowly. Cells were embedded in fresh resin (100%) for 8 h, overnight and then 8 h on a stirrer, and then polymerized at 60 °C for 48 h in an oven. The embedded cell blocks were sectioned with a diamond knife on an Ultra- microtome (Leica Microsystems, Wetzlar, Germany), and collected onto copper grids, stained with saturated uranyl acetate in 50% (v/v) ethanol for 4 min and lead citrate for 4 min. The stained sections were observed by electron microscopy at 60 kV.

2.2.3. Purification of methanogens

2.2.3.1. Minimum inhibitory concentration of antibiotics against Succinivibrio dextrinisolvens H5 within ISO4-H5 enrichment culture

Ampicillin (50 microgram (µg)/mL), chloramphenicol (50 µg/mL), erythromycin (20 µg/mL), gentamycin (20 µg/mL), kanamycin (50 µg/mL), rifampicin (1.2 µg/mL), spectinomycin (100 µg/mL), streptomycin (100 µg/mL) or tetracycline (10 µg/mL) were supplemented into triplicates of ISO4-H5 enrichment culture in the presence of glucose (10 mM), and the growth of *S. dextrinisolvens* H5 was compared to triplicates of negative control omitting antibiotics after overnight incubation (16 h). The CH_4 production was measured after seven days of growth with triplicates of ISO4-H5 enrichment culture under standard cultivation conditions as a positive control. The concentration of antibiotics was increased to identify the minimum inhibitory concentration that did not compromise CH_4 production.

2.2.3.2. Purification of methanogenic archaeon ISO4-H5

Methanogenic archaeon ISO4-H5 cultures supplemented with SSPGMS were treated with ampicillin (200 µg/mL), sub-cultured into BY medium supplemented with glucose (10 mM) to test for surviving *S. dextrinisolvens* H5, or media supplemented with SSPGMS and ampicillin to remove *S. dextrinisolvens* H5. The detection of bacterial contamination is described in Section 2.2.1.3. This procedure was repeated until no bacterial contamination was detected.

2.2.3.3. Purification of Methanobrevibacter sp. D5

Methanobrevibacter sp. D5 cultures with bacterial contamination were subjected to a mixture of ampicillin (10 µg/mL), streptomycin (10 µg/mL) and vancomycin (86.7 µg/mL), then heat-treated at 55 ^{o}C for 30 min followed by streaking on anaerobic BY agar plates. The agar plates were placed in a sealed cannister and pressurised with 1 atm over pressure of H_2:CO_2 (4:1). Plates were incubated at 39 ^{o}C for one week, and a pure *Methanobrevibacter* sp. D5 culture was isolated from a single colony and the subsequent culture purity was verified by 16*S* rRNA gene sequencing.

2.2.4. DNA extraction

2.2.4.1. Genomic DNA extraction

The protocol was adapted from the Genomic-Tip 500/G kit (Qiagen, Silicon Valley, Redwood city, CA, USA). A mortar and pestle were sterilized and stored at -80 ^{o}C. A 1 L culture of *Methanobrevibacter* sp. D5 was grown in Schott bottles or serum bottles. Culture purity was validated by microscopy. Cells were collected by centrifugation (20,200 × *g*, 20 min) in sterile centrifuge bottles and Oakridge tubes. The cell pellet was froze at -80 ^{o}C for a minimum of two h.

The cell pellet was transferred to the pre-chilled mortar and crushed with the pestle whilst kept frozen with liquid nitrogen (N_2). The pellet was ground in a circular motion as it slowly thawed into a slushy consistency, and additional liquid N_2 was added to keep the pellet from thawing completely. RNase A (2 µg/mL, Qiagen) was added to 11 mL of Buffer B1 solution (Qiagen). The ground cells were resuspended with 5 mL of Buffer B1, using a wide-bore pipette tip. The lysate was transfered to a 50 mL centrifuge tube (Falcon^{TM}, Fisher Scientific International Inc. Hampton, USA), and 6 mL of additional buffer B1 was added. Freshly prepared lysozyme (2.5 mg/mL [w/v], Sigma-Aldrich, St. Louis, MO, USA) and proteinase K (8 mg/mL [w/v], Qiagen) was added to the lysate mixture. The remainder of the purification procedures followed the Qiagen Genomic DNA Handbook for bacterial genomic DNA extraction.

The DNA extraction procedure of methanogenic archaeon ISO4-H5 was similar to the procedure used for *Methanobrevibacter* sp. D5, except due to low cell density, 2 L of ISO4-H5 culture was grown. ISO4-H5 cultivation was possible in 125 mL serum bottles but not in 250 mL, 500 mL or 1 L Schott bottles. ISO4-H5 did not require grinding, the addition of SDS (0.1%, [w/v]) to the lysate buffer along with proteinase K and lysozyme was sufficient to release genomic DNA.

2.2.4.2. DNA extraction for 16S rRNA polymerase chain reaction (PCR)

DNA was extracted from pellets of cells using the ZR Fungal/Bacterial DNA MiniPrep^{TM} Kit (Zymo Research Corporation, Irvine, CA, USA) with minor modification to the instruction. The bead beating time was shortened to 90 s.

2.2.5. PCR

Methanogen genome gap regions and partial fragments of 16*S* rRNA gene from both archaea and bacteria were amplified in this study. Each PCR (50 microlitre (µL)) reaction contained 45 µL of platinum supermix, 0.4 µM of each primer, 5 nanogram (ng) of template DNA and Type I water. Each set of PCR contained a positive control with sequence validated template for 16*S* rRNA gene amplification, a negative control replacing template with Type I water was used to identify false positive reactions and primer dimer formation. The primers used in this study are listed in Tables 2.2, A.2.1 and A.2.2.

Table 2.2. Primers used for PCR amplification of 16*S* rRNA gene

Primer pair	**Target gene (fragment size)**	**Sequence (5' – 3')**	**Thermal cycling conditions**	**Reference**
27f	Bacterial 16*S* rRNA (≈1465 bp)	GAG TTT GAT C**M**T GGC TCA G	94ºC 10 min	Weisburg *et al.* 1991
1492r		GG**Y** TAC CTT GTT ACG ACT T	94ºC 30 s 56ºC 30 s 72ºC 90 s (30 cycles) 72ºC 10 min	
915af	Archaeal 16*S* rRNA (≈471 bp)	AGG AAT TGG CGG GGg AGC AC	94ºC 10 min	Watanabe *et al.* 2004, Skillman *et al.* 2004
1386r		GCG GTG TGT GCA AGG AGC	94ºC 30 s 56ºC 30 s 72ºC 30 s (30 cycles) 72ºC 10 min	

Base pair (bp)

Table 2.3. Primers used for PCR amplification of ISO4-H5 putative prophage region

Primer pair	Target gene and fragment size	Sequence (5' – 3')	Thermal cycling conditions	Reference
H5PHX.p1ca	Putative prophage full length (32.8 kb) or ISO4-H5 genome (≥500bp)	GTC CTC CAA GGG TTA TTC GT	94ºC 10 min 94ºC 30 s 56ºC 30 s 68ºC 10 min (30 cycles) 68ºC 10 min	This study
H5PHX.q1ca		TCG TTT CCG GTT AAA AGA CG		
H5PH1.p1ca	Putative prophage internal region (470 bp)	CTG CTA TAC ATA TTG TGT CG	94ºC 10 min 94ºC 30 s 56ºC 30 s 72ºC 60 s (30 cycles)	This study
H5PH1.q1ca		TCA TTT CTT CTT CGT AGA GG		
H5PH2.p1ca	Putative prophage internal region (872 bp)	CTG CAG ACC TGT TCT GAC TA	72ºC 10 min	
H5PH2.q1ca		ATC AAG GTG GAG AAA TTT GG		
H5PH3.p1ca	Putative prophage internal region (924 bp)	TCG GCC TAT TAG CCT CTC TTT CAG		
H5PH3.q1ca		CAA AAG GTC CGG CAA AAT CC		
H5PH4.p1ca	Putative prophage internal region (904 bp)	TCT GGT CAT CAC CTC GAT TG		
H5PH4.q1ca		TTC CGC TTC TGC CTA CGA GA		
H5PH5.p1ca	Putative prophage internal region (850 bp)	GCT GGA TCT CGC CGA CTG AT		
H5PH5.q1ca		TCG AGA TCG TTT CTT CCA CC		

Base paire (bp)

2.2.6. Agarose gel electrophoresis

Amplified genome gap regions or 16*S* rRNA gene products were analysed by agarose gel electrophoresis. Agarose gels (1% [w/v]) were made up in 1× TAE buffer. Amplified PCR products were loaded with SYBR Safe DNA Gel Stain (20%, [v/v], Thermo Fisher Scientific) and run in 1× TAE buffer at 100 volts for 30 min. $1Kb^{+}$ DNA ladder (Thermo Fisher Scientific) was used as molecular weight marker in each gel. The gel was visualised under UV irradiation using a Gel Logic 200 imaging system (Eastman Kodak, New York, NY, USA).

2.2.7. PCR product purification

All PCR products were purified by ExoStar 1-Step (29%, [v/v], GE Healthcare, Little Chalfont, United Kingdom) by incubation at 37 ºC for 15 min and 80 ºC for 15 min.

2.2.8. DNA quantification

The DNA was quantified by using NanoDrop ND-1000 UV-Vis Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) according to the manufacturer's instructions.

2.2.9. Sequencing of PCR products

PCR products were sequenced by Massey Genome Service (Massey University, Palmerston North, NZ). Chromatogram of sequencing result were visualised and edited using Geneious 6.1.5 (Kearse *et al.* 2012).

2.2.10. Genome sequencing

2.2.10.1. Methanogenic archaeon ISO4-H5 genome

Genomic DNA from the methanogenic archaeon ISO4-H5 enrichment culture was extracted as described in Section 2.2.1. The paired-end library with a minimum insert size of 3 kilobase (kb) was prepared and whole-genome sequencing was carried out on a 454 GS-FLX Titanium System pyrosequencing platform (Macrogen, Seoul, South Korea).

2.2.10.2. Methanobrevibacter sp. D5 genome

Genomic DNA from a methanogen enrichment culture H6 containing *Methanobrevibacter* sp. D5 was determined by next-generation sequencing on a 454 GS-FLX Titanium System pyrosequencing platform (Macrogen, Seoul, South Korea). *Methanobrevibacter* sp. D5 was subsequently purified as described in Section 2.2.1, and genomic DNA was extracted as described in Section 2.8.1. Genomic DNA from a pure *Methanobrevibacter* sp. D5 culture was sequenced again using an Illumina MiSeq pair-end reads with a 250 bp insert size (Beijing Genomics Institute, Yantian District, Shenzhen, China).

2.2.11. Quality assessment

The quality of the raw DNA reads were validated by FastQC (Andrews 2010) and poor quality reads were trimmed with FLEXBAR (Dodt *et al.* 2012) using a Phred-scale cut-off of 33 and a minimum read size of 100 bp prior to genome assembly.

"Flexbar –r D5_1.fastq –p D5_2.fastq –n 4 –a adap.all.fa –q 33 –f i1.8 –m 100"

2.2.11.1. Methanogenic archaeon ISO4-H5 genome

Trimmed reads were assembled by the Genome Sequence de novo Assembler (Newbler) version 2.7 (Margulies *et al.* 2005) developed by Roche. The infoseq tool of the EMBOSS software suite was used to assess %G+C of each contigs (Rice *et al.* 2000). The *mcrA* gene of *Mbb. ruminantium* $M1^T$ was used as a reference to identify methanogenesis marker genes via Basic Local Alignment Search Tool (BLAST). The contigs and scaffolds were assigned to ISO4-H5 or *S. dextrinosolvens* H5 based on difference in %G+C and marker genes for methanogenesis. The command used in this study was:
"infoseq contigs.fa > assembly.txt"

2.2.11.2. Methanobrevibacter sp. D5 genome

The trimmed 454 reads of methanogen enrichment culture H6 was assembled by the Genome Sequence de novo Assembler (Newbler) version 2.7 (Margulies *et al.* 2005). The contigs and

scaffolds were assigned to *Methanobrevibacter* sp. D5 based on difference in %G+C, identity of ribosomal proteins, presence of methanogenesis marker genes and the tetra scores calculated by TETRA software in relation to scaffold one (Teeling *et al.* 2004), which was the largest *Methanobrevibacter* sp. D5 scaffold. The reads of *Methanobrevibacter* sp. D5 pure culture was assembled by SPAdes (Bankevich *et al.* 2012). The command used in this study was:

"Python bin/spades.py --pe1-1 ../D52/D5_1.fastq –pe1-2 ../D52/D5_2.fastq –o spades_output"

2.2.12. Genome annotation

The genome annotation was managed by a combination of GAMOLA (Altermann and Klaenhammer 2003) and Artemis software suite (Rutherford *et al.* 2000). The open reading frames (ORFs) and gene identities were predicted by Glimmer (Delcher *et al.* 1999) and BLASTX (Gish and States 1993). The start and stop of predicted ORFs were verified by manual inspection and redefined if necessary. The protein function of ORFs was assigned manually based on results from databases including: BLASTP (Altschul *et al.* 1990) to both non-redundant protein database provided by the National Centre for Biotechnology Information (NCBI) (Sayers *et al*. 2011), clusters of orthologous groups (COG) database (Tatusov *et al.* 2001). The protein motifs were identified by hidden Markov model (HMM) based HMMER (Eddy 1998) search against the Pfam (Finn *et al.* 2016) and TIGRFAM (Haft *et al.* 2003) database.

2.2.13. Primer design

Primers were designed in the Staden Package using the Gap4 software (Staden *et al.* 2000). The primers were designed with a minimum length of 20 bp to ensure uniqueness. The viability of primer sequence was validated in Vector NTI 11.5.2 (Thermo Fisher Scientific), using a cutoff of -4.7 kilocalorie (kcal)/mol and a maximum of 5 bp stem length to prevent primer dimer formation.

2.2.14. Genome gap closure

The genomes were closed by PCR-based gap closure. The gap regions within scaffolds were amplified with primers designed using neighbouring contigs as templates. Small contigs below 1,000 bp were not used as template due to low reliability. The gap regions between scaffolds were amplified with combination PCRs to test all possible connections between scaffolds. The gap closure progress was managed by the Gap4 software of the Staden Package (Staden *et al.*

2000). The potential frameshifts were identified, and assessed by sequencing, and were removed when found.

2.2.15. Pulsed field Gel Electrophoresis (PFGE)

The PFGE was performed using a CHEF-DR II PFGE System (Bio-rad, Hercules, CA, USA) to verify genome assembly and scaffold orientation. The procedure was identical between the *Methanobrevibacter* sp. D5 and that previously described for *Mbb. ruminantium* $M1^T$ PFGE (Leahy *et al.* 2010). Cells were harvested from 50 mL of culture after seven days, by centrifugation (10,000 × *g*, 10 min), frozen with liquid N_2, gently ground by motar and pestle to damage the cell wall, and thawed. The ground material was suspended in 2 mL of 1 M NaCl, 10 mM Tris (pH 7.6), and 300 µL aliquots were mixed with equal volumes of 2% (w/v) Seaken® Gold low melt agarose (Lonza Group AG, Basel, Switzerland) to embed cells. The cells were lysed in lysis buffer (50 mM Tris-HCl, 50 mM EDTA, 1% [w/v] sarkosyl, 0.1 mg/mL Proteinase K (Qiagen)) at 50 ^oC for an h, then washed twice with autoclaved Type I water and three times with TE buffer (pH 8.0), before storing the plug in 10 mM Tris-HCl, 100 mM EDTA (pH 8.0) at 4 ^oC. The restriction enzymes *Apa*I and *Mlu*I (New England Biolabs, Beverly, MA, USA) were used to digest the embedded *Methanobrevibacter* sp. D5 genomic DNA. The embedded DNAs were loaded into 1% (w/v) Seaken Gold agarose, run at 200 V, 20 h, 14 ^oC in 0.5×TBE buffer, The gel was stained with ethidium bromide and visualised using a Gel Doc 1000 system (Kodak Gel Logic 200 Imaging System, Eastman Kodak, Rochester, NY, USA).

2.2.16. Inference of phylogenetic divergence

2.2.16.1. Functional genome distribution (FGD) tree

The similarity between genomes can be inferred from an organism's ORFeome using the unweighted pair group method with arithmetic mean (UPGMA) method (Sneath and Sokal 1962) based on the FGD dissimilarity matrix (Altermann 2012).

2.2.16.2. Phylogenic trees based on 16S rRNA genes

The 16*S* rRNA gene DNA sequence were used to infer the phylogenetic relationship of the methanogens analysed in this thesis. Evolutionary analysis were conducted in MEGA6 (Tamura *et al.* 2013). The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The substitution model was based on the Kimura 2-parameter distance (Kimura 1980). All positions containing gaps or missing data were

eliminated. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the relationship of the taxa analysed.

2.2.16.3. Phylogenic trees based on mcrA/mrtA amino acid sequences

The highly conserved methanogenesis marker gene *mcrA/mrtA* amino acid sequence were used to infer the phylogenetic relationship of methanogens analysed. Evolutionary analysis were conducted in MEGA6 (Tamura *et al.* 2013). The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the MCL approach. The substitution model was based on the Jones-Taylor-Thornton (JTT) distance matrix (Jones *et al.* 1992) at uniform rates. All positions containing gaps or missing data were eliminated. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the relationship of the taxa analysed.

2.2.16.4. Pan-genome tree based on gene families

Pan-genome tree of was inferred based on the absence and presence of gene families by CMG-Biotools (Vesth *et al.* 2013). The relative Manhatten distance were calculated by hierarchical clustering. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the evolutionary history of the taxa analysed.

2.2.17. Prediction of RNAs

2.2.17.1. Prediction of transfer RNAs (tRNAs)

The program tRNAscan-SE 1.3.1 was used to detect tRNA genes in genomic DNA, which utilised the Cove probabilistic RNA prediction package (Eddy and Durbin 1994) with improved algorithm implemented for searching for eukaryotic tRNA promoters (Lowe and Eddy 1997). Cove probabilistic RNA prediction maximized the sensitivity, and combined with the 'archaeal' option allowed the identification of otherwise undetected archaeal tRNAs that contains introns at non-canonical position. The command used in this study was:

"tRNAscan-SE –o output.txt –m outputstats.txt –C –A input.fasta".

2.2.17.2. Prediction of non-coding RNAs

The program Infernal ("INFERence of RNA ALignment") 1.1 was used to predict non-coding RNAs within the genomic DNA. Infernal utilises covariance models to identify RNA homologs with secondary structure conservation to non-coding RNAs in reference databases (Nawrocki and Eddy 2013). The Rfam database was used as reference (Nawrocki *et al.* 2015). A cut off of 1e-05 was applied to remove false positives. The command used in this study was:

"cmscan Rfam.cm intput.fa > output.txt"

2.2.18. Prediction of signal peptides and secretome

Signal peptides were predicted by SignalP 4.1 (Petersen *et al.* 2011) (www.cbs.dtu.dk/services/SignalP). The SignalP-HMM algorithm trained on eukaryotes, Gram positive and Gram negative bacteria were applied to identify signal peptides in Archaeal genomes. The transmembrane helix (TMH) was predicted by TMHMM (Krogh *et al.* 2001) using default parameters. The lipobox was predicted by LIPO (Berven *et al.* 2006) using default parameters. The LPxTG-like motif was predicted using CW-PRED (Fimereli *et al.* 2012) using default paramters.

2.2.19. Prediction of adhesin-like proteins

In order to identify adhesin-like proteins in the genome, an integrated computational approach was used. Adhesin prediction was carried out with SPAAN (Sachdeva *et al.* 2005), a Pad-value was generated as predicted likelihood of genes being adhesin. Additional subcellular localization algorithms LOCTree3, PSORTb and SubLoc were included to minimize false positive (Chen *et al.* 2006; Yu *et al.* 2010; Goldberg *et al.* 2014). Proteins with a SPAAN P_{ad}-value of 0.70 or above and two of the subcellular localization algorithm predicted as extracellular, secreted, plasma membrane or cellwall were considered as adhesin-like proteins.

2.2.20. Prediction of horizontal transfer regions and genes

In order to investigate horizontal gene transfer, Alien_Hunter (Vernikos and Parkhill 2006), an annotation-independent analysis software that utilises interpolated variable order motifs (IVOM) was used. The disadvantage of this parametric method is the high number of false positives (Azad and Lawrence 2007). The regions with abnormal %G+C and codon composition are more likely to be horizontally transferred. The codon composition is indicated by codon adaptation index (CAI) (Sharp and Li 1987) calculated by CAIcal (Puigbo *et al.* 2008) (http://genomes.urv.cat/CAIcal/).

Table 2.4. Bioinformatics software used in this study

Software	Description/Application	Source	Reference
Genome sequence de novo assembler 2.7	Genome assembler for 454 sequencing platform	http://www.454.com/products/analysis-software/	Margulies *et al.* 2005
Velvet de novo assembler	Genome assembler for Illumina sequencing platform	https://www.ebi.ac.uk/~zerbino/velvet/	Zerbino 2010
SPAdes	Genome assembler for Illumina sequencing platform	http://bioinf.spbau.ru/spades	Bankevich *et al.* 2012
FastQC	Quality control of raw sequence data	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/	Andrews 2010
Flexbar	Sequence trimming and adapter removal	http://sourceforge.net/projects/flexbar/	Dodt *et al.* 2012
Artemis Release 8.0	Genome browser and annotation tool	http://www.sanger.ac.uk/science/tools/artemis	Rutherford *et al.* 2000
DNA Plotter	Genome visualization	http://www.sanger.ac.uk/Software/Artemis/circular/	Carver *et al.* 2009
BLAST	Comparison of sequence similarity	http://blast.ncbi.nlm.nih.gov/Blast.cgi	Altschul *et al.* 1990
GAMOLA	Gene calling and automated annotation	http://fsweb2.schaub.ncsu.edu/TRKwebsite/index.htm	Altermann and Klaenhammer 2003
Staden package	Primer design and gap closure	http://staden.sourceforge.net/	Staden *et al.* 2000
Vector NTI Advanced 11.5.2	DNA analysis and design suite	https://www.thermofisher.com/nz/en/home/life-science/cloning/vector-nti-software.html	Lu and Moriyama 2004
Geneious 6.1.15	DNA and protein analysis suite	http://www.geneious.com/	Kearse *et al.* 2012
BLAST Ring Image Generator	Genome identity visualisation	http://brig.sourceforge.net/	Alikhan *et al.* 2011
tRNAscan-SE	tRNA prediction tool	http://lowelab.ucsc.edu/tRNAscan-SE/	Lowe and Eddy 1997
Rfam 12.0	Database of RNA families	http://rfam.xfam.org/	Nawrocki *et al.* 2015
Pfam 29.0	Database of protein families	http://pfam.xfam.org/	Finn *et al.* 2016
Infernal 1.1	RNA prediction tool	http://infernal.janelia.org/	Nawrocki and Eddy 2013
MEGA6	Sequence alignment and phylogeny analysis	http://www.megasoftware.net/	Tamura *et al.* 2013
SignalP 4.1	Signal peptide prediction	http://www.cbs.dtu.dk/services/SignalP/	Petersen *et al.* 2011
LIPO	Lipo-box prediction	http://services.cbu.uib.no/tools/lipo	Berven *et al.* 2006
CW-PRED	LPxTG motif prediction	http://bioinformatics.biol.uoa.gr/CW-PRED/	Fimereli *et al.* 2012
Pfam	Pfam prediction	http://pfam.xfam.org/	Finn *et al.* 2016
TETRA	Tetranucleotide usage pattern analysis	http://www.megx.net/tetra	Teeling *et al.* 2004
SPAAN	Adhesin prediction	http://sourceforge.net/projects/adhesin/files/SPAAN/	Sachdeva *et al.* 2005
PSORTb 3.0	Protein subcellular location prediction	http://www.psort.org/psortb/	Yu *et al.* 2010
SubLoc	Protein subcellular location prediction	http://www.bioinfo.tsinghua.edu.cn/SubLoc/	Chen *et al.* 2006
LocTree3	Protein subcellular location prediction	https://rostlab.org/services/loctree3/	Goldberg *et al.* 2014
ISfinder	Insertion sequence element prediction	https://www-is.biotoul.fr/	Siguier *et al.* 2006

Alien_Hunter	Horizontal transferred region prediction	http://www.sanger.ac.uk/science/tools/alien-hunter	Vernikos and Parkhill 2006
CAIcal	Codon adaption index calculation	http://genomes.urv.cat/CAIcal/	Puigbo *et al.* 2008
PKS/NRPS Analysis Web-site	Identification and domain prediction of NRPS/PKS	http://nrps.igs.umaryland.edu/	Bachmann and Ravel 2009
NRPSsp	NRPS substrate predictor	http://www.nrpssp.com/execute.php	Prieto *et al.* 2012
REPFIND	Repeat identification	http://zlab.bu.edu/repfind/	Betley *et al.* 2002
CMG-biotools	Comparative genomics tools	http://www.cbs.dtu.dk/~dave/CMGtools/	Vesth *et al.* 2013
MestRe Nova 11.0	^{1}H-NMR data analysis	http://mestrelab.com/	Willcott 2009
Chenomx NMR suite 7.0	NMR analysis suite with compound library	http://www.chenomx.com/	Weljie *et al.* 2006
HMDB	Human metabolome database	http://www.hmdb.ca/	Wishart *et al.* 2007
RMDB	Bovine rumen metabolome database	http://www.rumendb.ca/cgi-bin/browse.cgi	Saleem *et al.* 2013
MetaboAnalyst 2.0	Statistical analysis of metabolomics data	http://www.metaboanalyst.ca/	Xia *et al.* 2012
Rockhopper	Transcriptome analysis suite	http://cs.wellesley.edu/~btjaden/Rockhopper/	McClure *et al.*, 2013
EDGE-pro	Transcriptome analysis suite	http://ccb.jhu.edu/software/EDGE-pro/	Magoc *et al.*, 2013
R packages			
mixOmics	Regularised canonical correlation analysis	https://cran.r-project.org/web/packages/mixOmics/index.html	Dejean *et al.* 2011
Predictmeans	Analysis of variance and Kruskal-Wallis test	https://cran.r-project.org/web/packages/predictmeans/index.html	Luo *et al.* 2014
PMCMR	Analysis of variance	https://cran.r-project.org/web/packages/PMCMR/index.html	Pohlert 2014
Vegan	Multivariate analysis	https://cran.r-project.org/web/packages/vegan/index.html	Oksanen *et al.* 2007
Ape	Multivariate analysis	https://cran.r-project.org/web/packages/ape/index.html	Paradis *et al.* 2008
Perm	Multivariate analysis	https://cran.r-project.org/web/packages/perm/index.html	Fay and Fay 2010
MASS	Multivariate analysis	https://cran.r-project.org/web/packages/MASS/index.html	Ripley *et al.* 2015
FactoMineR	Correspondence analysis	https://cran.r-project.org/web/packages/FactoMineR/index.html	Lê *et al.* 2008
doBy	Correspondence analysis	https://cran.r-project.org/web/packages/doBy/index.html	Højsgaard *et al.* 2006
DESeq	Calculate differential gene expression	http://bioconductor.org/packages/release/bioc/html/DESeq.html	Anders and Huber 2010

2.2.21. Prediction of non-ribosomal peptide synthase (NRPS) domains and substrates

The identity of gene encoding NRPSs was predicted during the annotation. The adenylation domains are predicted by the PKS/NRPS Analysis website (Bachmann and Ravel 2009) (http://nrps.igs.umaryland.edu/). The substrates of adenylation domains were predicted by NRPSsp (Non-Ribosomal Peptide Synthase Substrate Predictor, http://www.nrpssp.com/index.php) (Prieto *et al.* 2012).

2.2.22. Prediction of insertion sequence (IS) elements

IS were predicted by IS Finder (Siguier *et al.* 2006) (https://www-is.biotoul.fr/) based on mobile elements from annotation, and REPFIND (Betley *et al.* 2002) (http://zlab.bu.edu/repfind/) was used to verify flanking repeat sequence utilised for mobilisation.

2.2.23. Prediction of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)

The CRISPR elements were predicted by the CRISPR Finder (Grissa *et al.* 2007) (http://crispr.u-psud.fr/Server/). The number and sequence of direct repeats were identified, the spacer sequences were analysed by CRISPRTarget (Biswas *et al.* 2013)(http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html) to identify targets of inhibition.

2.2.24. Metabolic pathway mapping

The metabolic pathway was constructed based on homology to experimentally validated genes and existing pathways in databases including the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) and Metacyc (Caspi *et al.* 2014). The gene homology was compared based on BLOSUM62 (BLOcks SUbstitution Matrix) sequence alignment (Henikoff and Henikoff 1992) within Geneious (Kearse *et al.* 2012).

2.2.25. Codon frequency, amino acid usage, BLAST matrix and Pan-core plot

Codon frequency, amino acid usage were calculated by CMG-biotools (Vesth *et al.* 2013). The genomes were compared by CMG-biotools to produce a BLAST matrix and a Pan-core plot based on a gene family criteria of 50% amino acid identity over 50% of protein length.

2.2.26. Genome similarity analysis

Genome synteny was analysed using the programe PROmer (Delcher *et al.* 2003) in the the MUMmer sequence alignment package. The command used in this study was:

"./promer –prefix=outputname input01.fasta input02.fasta"

"./mummerplot outputname.delta"

The genomes were compared and visualised by BLAST Ring Image Generator (BRIG) to identify similar regions conserved between genomes (Alikhan *et al.* 2011).

2.2.27. Gene homology analysis

The identification of homolog used experimentally validated genes as reference genes, and a BLASTP (Altschul *et al.* 1990) expectation value of e-05 was used as cutoff unless otherwise stated. The coverage and length of the query sequence was checked for truncation and gene disruption, and the surrounding genes were analysed to detect any potential operon structures between functionally connected genes. Conservation of operons between genomes were also analysed.

2.2.28. Metabolite analyses

The proton nuclear magnetic resonance (^{1}H NMR) was used to distinguish the metabolites depleted by growth of methanogenic archaeon ISO4-H5. Uninoculated controls and samples prior and post growth were processed by filtration via Nanosep 3k Omega (PALL Corporation, Port Washington, NY, USA), with 3 kilodalton (kDa) cut-off to remove protein molecules. 440 µL of sample was mixed with 100 µL of $NaHPO_4$ buffer (300 mM, pH 7.4) and 60 µL of 5 mM DSS in deuterated water (D_2O) was used as internal standard.

The ^{1}H NMR was conducted on a superconducting 700 mega hertz (MHz) nuclear magnetic resonance spectrometer (Bruker Corporation, Billerica, MA, USA) at Massey University, Palmerston North, NZ. Each spectra has been adjusted individually with baseline correction via Whittaker Smoother algorithm and manual phase correction in MestReNova 11.0 (Willcott 2009) with the assistance of NMR specialist Dr. Linda Samuelson. Spectras were superimposed and auto-scaled, and aligned by reference peaks. The water peaks, imidazole peaks and reference peaks were removed from aligned spectra. Spectra regions containing peaks between chemical shift of 0.6 and 8.6 parts per million (ppm) were binned by 0.01 ppm via sum, and normalized by total average of 100%. The binned spectra were statistically analysed in

Metaboanalyst 2.0 (Xia *et al.* 2012), data was normalised by log transformation (corrected for heteroscedasticity, the unequal variability of variables) and scaled by Pareto scaling (uses the square root of the standard deviation as the scaling factor) (Brodsky *et al.* 2010). T-tests with *P*-value threshold of 0.05 were used to identify significantly depleted compounds in samples. The compounds were identified in Chenomx 7.0 library (Weljie *et al.* 2006). The identity of compound was verified by HSQC and TOCSY using the Human Metabolome Database as reference (Wishart *et al.* 2013).

2.2.29. Transcriptome analysis of methanogenic archaeon ISO4-H5

2.2.29.1. Experimental design

ISO4-H5 gene expression was investigated with two variables, the H_2 levels and methyl-substrates. Nine different treatment groups were designed to include controls for each treatment (Table 5.3.). Eight biological replicates were conducted in serum bottles, and the CH_4 and H_2 concentration in headspace samples was monitored daily. Four replicates were harvested at 120-h and four replicates were harvested at 144-h post inoculation. The samples were snap frozen in liquid N_2 and stored in -80 ^oC.

2.2.29.2. RNA extraction

All benches, fumehood and equipment were cleansed with RNaseZap decontamination solution (Thermo Fisher Scientific) prior to each RNA extraction experiment. All procedures were performed on ice unless otherwise stated. The frozen sample was thawed in RNAprotect bacteria reagent (1 mL/g of sample, Qiagen) slowly over ice. The cell suspension was centrifuged (15,000 × *g*, 10 min, 4 ^oC), and the supernatant was decanted carefully. Each cell pellet was resuspended in 500 µL of RNA extraction buffer A, 210 µL of 20% SDS, 500 µL of phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5, Thermo Fisher Scientific) with 500 mg of RNase free Zirconia/Silica beads (0.1 mm:0.5 mm, 1:1). The cell suspension was lysed by bead beating twice (45s, 6.5 speed) in a Savant FastPrep FP120 (MP Biomedicals, Santa Ana, CA, USA) with 10 min intervals on ice to prevent samples from overheating. Alternatively, a Mini-Beadbeater (Biospec products INC. OK, USA) was used. Samples were incubated at 4 ^oC overnight. Samples were processed by bead beating three more times. Samples were centrifuged (10,000 × *g*, 5 min, 4 ^oC), and the top layer was transferred to RNase free microcentrifuge tubes. An equal volume of isopropanol and 10% volume of 5 M ammonium acetate (pH 5.5) was mixed and stored at -20 ^oC for a minimum of one h. The RNA was precipitated by centrifugation (10,000 × *g*, 30 min, 4 ^oC). The isopropanol was carefully

removed, and the pellet was washed with 500 µL of 70% RNase free ethanol (Analytical grade ethanol:DEPC treated water, 7:3), centrifuged (14,000 × *g*, 5 min, 25 ^{o}C). The RNA/DNA pellet was dissolved in 50 µL of nuclease free water. The quality of RNA/DNA was checked by gel electrophoresis using 2% agarose.

The tRNA and 5*S* rRNA was removed by MEGAclear Transcription Clean-Up Kit (Thermo Fisher Scientific) as instructed by the manufacturer. The DNA was then removed from samples by two consecutive DNAse treatments; Baseline-ZERO DNase (Illumina Netherlands BV, Eindhoven, Netherlands) treatment was carried out as instructed by the manufacturer, and the TURBO DNA-free Kit (Thermo Fisher Scientific) was also used as the manufacturer instructed. DNA removal was validated by PCR. The RNA was quantified using a Qubit Fluorimeter (Thermo Fisher Scientific) and RNA quality was assessed using a 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA). Three µg of purified RNA was sent to BGI for rRNA depletion and sequenced using the Illumina HiSeq 2000 platform. Libraries with 200 bp inserts were generated and paired-end sequenced to produce 90 bp reads. The adapter sequence were removed by BGI.

2.2.29.3. Sequence read alignment and normalisation

The quality of sequence reads in FASTQ files were assessed by FastQC (Andrews 2010), and reads below a Phred quality score of 28 were trimmed by Flexbar (Dodt *et al.* 2012). Rockhopper (McClure *et al.* 2013) is a programme specifically designed for RNA-seq data analysis. Reference based assembly was carried out with 2% allowed mismatches and default parameters. The reads were aligned to the reference genome based on Bowtie2 (Langmead and Salzberg 2012), where an index was created based on Burrows-Wheeler transform (Burrows and Wheeler 1994) for the reference genome, and reads were aligned exactly to the genome or aligned by seed regions of the read and extended with Smith-Waterman algorithm (Smith and Waterman 1981).

The read count from each sample was normalized by the upper quartile gene expression level excluding genes with zero expression, which was found to provide the most coherent result to the quantitative PCR data (Bullard *et al.* 2010). Rockhopper determines differential expression of each gene. The sum of normalized reads for two treatments were used to calculate differential expression based on the Anders and Huber approach (Anders and Huber 2010), using the statistical test for null hypothesis, negative binomial distributions was used as statistical model to compute two-sided *P*-value. *P*-values are influenced by number of

replicates, size of transcript and variance of transcripts across replicate samples. To reduce false positives, *q*-values were computed from the *P*-value based on the Benjamini-Hochberg correction (Benjamini 2010) with a false discovery rate of <1%.

In addition to Rockhopper, a second software, EDGE-pro (Magoc *et al*., 2013) was used to analyse the transcriptome reads, which follows similar procedure to Rockhopper. The reads were aligned using Bowtie2 (Langmead and Salzberg 2012) to the reference genome, where the raw coverage was converted to reads per kilobase of transcript per Million mapped reads (RPKM) by $R = C \times L/r$, where R is reads mapped to the gene, C is average coverage, L is length of gene, r is read length. The RPKM was then analysed by DESeq package (Anders and Huber 2010) in Rstudio, the differential gene expression was based on the relationship of mean to variance via binomial distribution. The results from EDGE-pro and DESeq was not used.

2.2.29.4. Statistical determination of differential expression

Additional statistical testing was performed in order to select differentially expressed genes. Statistical tests were conducted using R studio 0.98.1049 (Rstudio Team, 2015). Three univariate tests were performed on the normalised reads; one-way analysis of variance (ANOVA), Kruskal-Wallis test (Kruskal and Wallis 1952) and Tukey's test (Tukey 1949) was performed to test the null hypothesis. The result of statistical tests would only be reliable when the data fits the assumption of each test. The two genes were considered differentially expressed between two treatments when both the *q*-value from Benjamini-Hochberg correction and the *P*-value from a reliable univariate test satisfies the criteria of <0.05.

The analyses were based on Bray-Curtis dissimilarity index using relative transcript abundance (Anderson *et al.* 2006), including non-metric multidimensional scaling (MDS), principal coordinate analysis (PCoA) (Legendre and Gallagher 2001), and permutation test for constrained analysis of principal coordinates (Legendre *et al.* 2011). The MDS is a multivariate statistical analysis used to reduce dimensionality and enable visualising similarities and dissimilarities between treatments (Oksanen *et al.* 2007). The PCoA is also used in dimension reduction and data visualization; it differs to MDS by preserving the covariance between data instead of distance (Gower and Legendre 1986). The group dispersion plot determines if one treatment is more variable than the others by permutation distribution under the null hypothesis (Anderson *et al.* 2006). Figures were produced in R studio with vegan, ape, perm and MASS package (Oksanen *et al.* 2007; Paradis *et al.* 2008; Fay and Fay 2010; Ripley *et al.* 2015).

The *R. flavefaciens* FD1 genome was re-annotated to correct for errors in ORF prediction and annotation to serve as an accurate reference genome for the transcriptome analysis. The re-annotation was carried out as described in Section 2.2.12.

Non-parametric multivariate analysis was performed to determine similarity of variation between different treatments. Bray-Curtis dissimilarity measure of distance was used to represent similarity (Bray and Curtis 1957). MDS and principle coordinate analysis were applied to visualize the similarity and clustering of variables between treatments by reducing dimensions, where a low stress value indicates a reliable MDS. Multiple correspondence analysis were used to prioritise and visualise dissimilarity between gene sets presented as heatmaps and network diagrams. Relationships between gene expression of different organisms within the same treatment was assessed by regularised canonical correlation.

Chapter 3

Analysis of the methanogenic archaeon ISO4-H5 genome

3.1. Introduction

Members of the order Methanomassiliicoccales, are methylotrophic methanogens and representatives of this order have been detected in various habitats, including landfills, rice fields, marine thermal vents, fresh water, and in the digestive tracts of termites, millipedes, chickens, ruminants and humans (Großkopf *et al.* 1998; Shinzato *et al.* 1999; Takai and Horikoshi 1999; Egert *et al.* 2003; Huang *et al.* 2003; Dridi *et al.* 2012; Horz *et al.* 2012; Iverson *et al.* 2012; Paul *et al.* 2012; Leahy *et al.* 2013; Padmanabha *et al.* 2013). The Methanomassiliicoccales are considered to be an important group of methanogens in the rumen environment and were originally referred to as RCC methanogens (Tajima *et al.* 2001; Bapteste *et al.* 2005). Their abundance in the rumen is highly variable, according to 16*S* ribosomal RNA gene surveys (Wright *et al.* 2006; Zhou *et al.* 2009; Huang *et al.* 2012), but on average, they constitute around 16% of the rumen archaeal community and are the second most abundant order of rumen methanogens (Janssen and Kirs 2008; Henderson *et al.* 2015). Representatives of these organisms have only recently been isolated in culture, and genomic information on members of the Methanomassiliicoccales are available for isolates from human, termite and bovine sources (Borrel *et al.* 2012; Gorlas *et al.* 2012; Borrel *et al.* 2013; Lang *et al.* 2015; NCBI Reference Sequence: NC_020892.1, (Noel *et al.* 2016; Sollinger *et al.* 2016)).

An enrichment culture of a representative of the Methanomassiliicoccales was previously obtained from the rumen contents of a nine year old Romney wether sheep in NZ grazing a ryegrass-clover pasture diet (Jeyanathan 2010). This enrichment culture contained a methanogenic archaeon, designated ISO4-H5 and a Gram-negative bacterium, subsequently identified as being most closely related to *Succinivibrio dextrinosolvens*, and designated as strain H5. The methanogenic archaeon ISO4-H5 culture was enriched based on serial dilution and heat treatment with CoM supplementation, without antibiotics (Jeyanathan 2010). The ISO4-H5 culture grew slowly and required three to four days to generate detectable CH_4 in the culture headspace. The optical density of cultures after maximal CH_4 formation was very low and the ISO4-H5 cells could not be visualized via fluorescence microscopy at 420 nm due to an apparent lack of the fluorescent 8-hydroxy-5-deazaflavin cofactor, F_{420} (Jeyanathan 2010). To gain an insight into the role of ISO4-H5 in methylotrophic methanogenesis in the rumen

environment, its genome was sequenced. The ISO4-H5 genome represents the first example from a member of the order Methanomassiliicoccales isolated from the ovine rumen.

3.2. Results

3.2.1. Genome sequencing results and assembly

Genome assembly. Whole genome sequencing yielded 1,025,715 reads and a total of 203,994,869 bp. The summary of the assembly statistics is displayed in Table 3.1. The assembly generated 29 scaffolds consisting of 690 contigs with an average size of 6,821 bp, the largest contig was 217,625 bp. The assembled genome of 29 scaffolds totals 4,660,809 bp, this provided 43.8× genome coverage for the combined genome data set containing ISO4-H5 and *S. dextrinosolvens* H5. There were 223 large contigs (> 500 bp) totalling 4,570,501 bp which were used for downstream analyses.

Table 3.1. Genome assembly summary

Genome sequence feature	**Size or %G+C**
Enrichment metagenome	
Total size	4,660,809 bp
Coverage	43.8×
Number of reads	1,025,715
Number of bases (bp)	203,994,879
Number of contigs	690
Number of scaffolds	29
Number of contigs >500 bp	223
G + C content	46.3%
G + C content omiting scaffold 14	40.71%
Scaffold 14	
G + C content	53.98%
Size	1,936,667 bp
Number of reads	356,637
Number of contigs	47
N50 contig size	73,897 bp

ISO4-H5 scaffold assignment. There were two distinct %G+C profiles observed in the enrichment genome (Figure 3.1A.); scaffold 14 was 54%, while scaffolds 01 to 13 and 15 to 29 were 40.7%. The methanogen specific gene, methyl coenzyme-M reductase subunit A (*mcrA*) was identified within scaffold 14, and as a result, this scaffold was assigned to the methanogenic archaeon ISO4-H5. Scaffolds 01 to 13 and 15 to 29 were assigned to the *S. dextrinosolvens* H5 genome, and encompassed 176 contigs totaling 2,724,142 bp. A purified culture of *S. dextrinosolvens* H5 was included in the Hungate 1000 project, which aims to produce a reference set of rumen microbial genome sequences from cultivated rumen bacteria

and methanogenic archaea, together with representative cultures of rumen anaerobic fungi and ciliate protozoa (Creevey *et al.* 2014). The re-sequenced genome of H5 was reassembled into 106 contigs totalling 2,675,466 bp, and is predicted to encode 2,350 protein coding genes, (currently available through the DOE JGI web portal, Project ID 1026020). The *S. dextrinosolvens* H5 genome has not been analysed as part of this thesis.

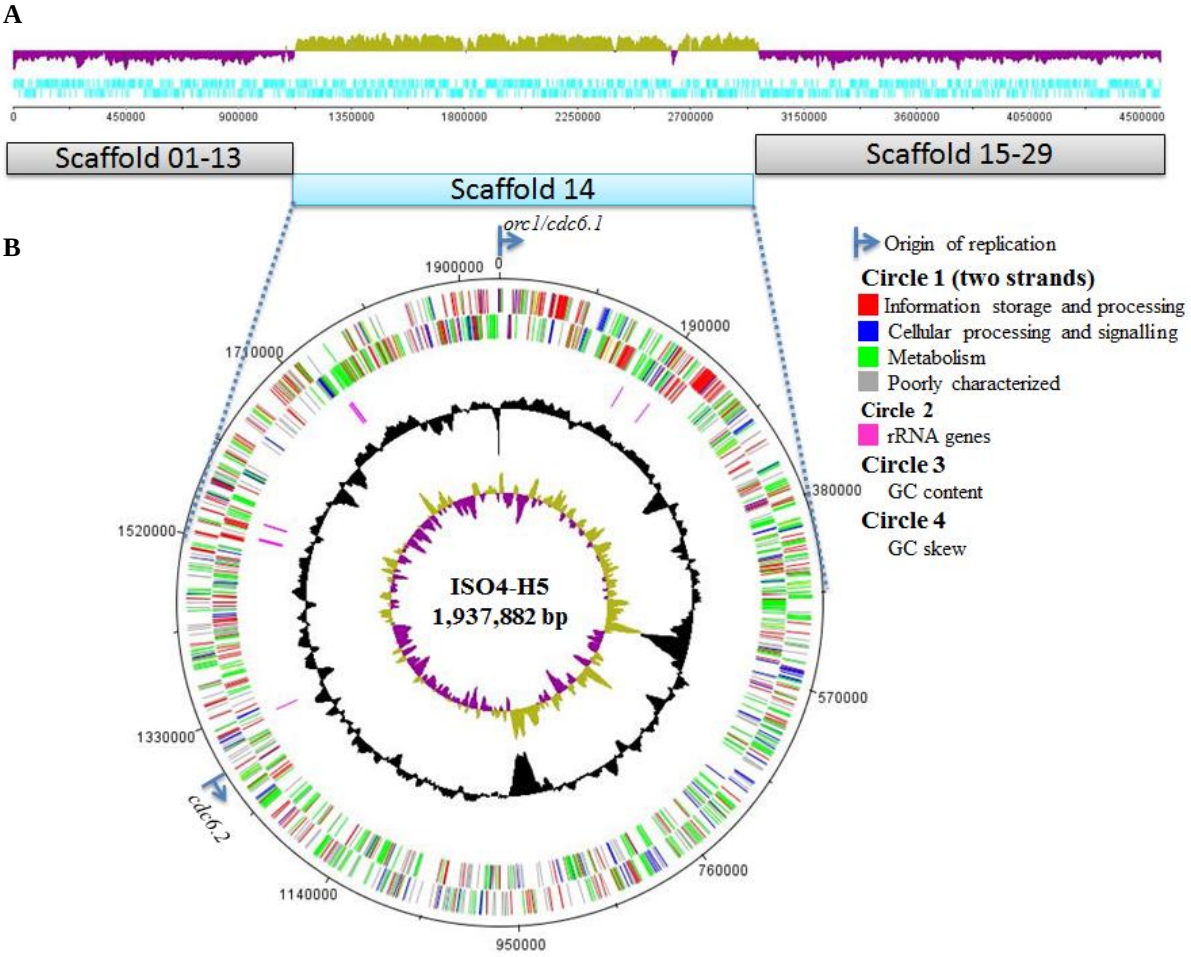


Figure 3.1. ISO4-H5 genome assembly and representation. **A**. The %G+C content of the ISO4-H5 and S. dextrinosolvens H5 genome assembly. The %G+C content is displayed with a window size of 10,000 bp and a step size of 200 bp. The khaki area represents an above average % G+C content, and the purple area represents a below average % G+C content of the sequence data analysed. The cyan represents predicted ORFs in forward (upper) and reverse (lower) strand. B. Circular representation of the ISO4-H5 genome. Circles are numbered 1 (outermost) to 4 (innermost). Circle 1: predicted ORFs on the + and – strands respectively. ORFs are coloured based on the COG categories. Circle 2: location of the rRNA genes. Circle 3: % G+C content. Circle 4: GC skew [(G-C)/(G+C)], khaki coloured regions indicate values >1, purple <1. The identified origins of replication are represented as blue lines perpendicular to the outermost circle, the arrow indicates the 5' to 3' direction of DNA replication.

ISO4-H5 gap closure. To complete the ISO4-H5 genome as a circular chromosome, it was necessary to close the gaps between the identified 47 contigs. The gap closure procedure is described in detail in Section 2.2.14. Nine rounds of gap closure were required to circularize the ISO4-H5 genome as a single contig. This involved a total of 163 PCRs and sequencing

reactions to close gaps and to improve the quality of the genome sequence to ensure correct assembly and to resolve any remaining base conflicts.

3.2.2. Genome properties

The general features of the ISO4-H5 genome are summarised in Table 3.2. Three pseudogenes were predicted in the genome of ISO4-H5, including a tRNA 2'-*O*-methylase (AR505_0044), a site-specific recombinase (AR505_1055) and a tryptophan synthase beta subunit *trpB* (AR505_1151). No plasmids were identified in the genome of ISO4-H5 but a toxin/antitoxin system (AR505_0857, 0858) possibly from a conjugative plasmid or prophage was identified.

Table 3.2. General features of the ISO4-H5 genome

Features	
Source	Ovine rumen
Project status	Complete
Genome size (bp)	1,937,882
%G+C content	54
Number of ORFs	1,845
Coding area (%)	90
Contigs	1
rRNAs (5*S*, 16*S*, 23*S*)	3, 2, 2
tRNAs (with introns)	47 (3)
Non-coding RNAs	6
IS	48
Prophage	1
CRISPR regions	1
Adhesin-like proteins	45

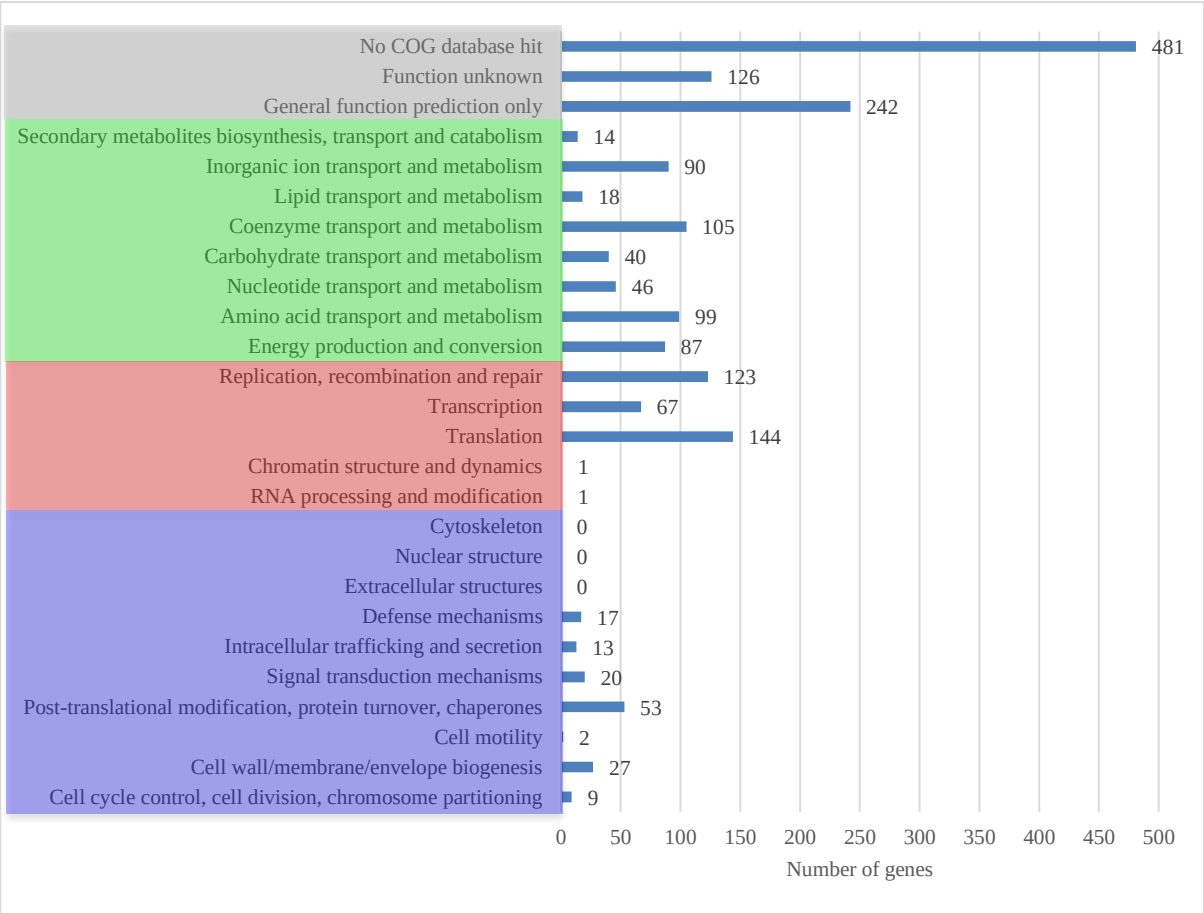


Figure 3.2. Functional classification of the ISO4-H5 predicted genes based on the clusters of orthologous proteins (COGs) database.

The functional categorizations of the predicted genes are summarized in Figure 3.2. Approximately one third of the total genes are annotated as hypothetical proteins with unknown function.

Identification of the ISO4-H5 genome origin of replication. The replication origin (*orc1/cdc6*) of the ISO4-H5 chromosome was identified by GC nucleotide skew [(G-C)/(G+C)] analysis (Figure 3.1B.). Two *orc1/cdc6* genes were predicted in the genome (AR505_0001, AR505_1205). Two Origin Recognition Box (ORB) motifs were identified, 36 bp and 77 bp downstream of AR505_0001, whereas AR505_1205 gene had no ORB identified. Therefore, the *orc1/cdc6* gene (AR505_0001) was predicted to be the true origin of replication of ISO4-H5.

RNA. ISO4-H5 possesses three copies of the 5*S* rRNA gene, two copies of the 16*S* rRNA gene and 2 copies of the 23*S* rRNA gene, but no operon structure was observed for these identified rRNAs (Table 3.3.). Six ncRNAs were predicted to be present in the genome of ISO4-H5 (Table 3.3.).

Table 3.3. Predicted rRNA and ncRNA genes of ISO4-H5

Start	End	Strand	E-value	Identity
rRNA				
161344	161465	+	3.10E-13	5*S* rRNA
200076	199955	-	3.10E-13	5*S* rRNA
1314829	1314950	+	3.10E-13	5*S* rRNA
1742102	1740637	-	0	16*S* rRNA
1552272	1550806	-	0	16*S* rRNA
1530665	1530555	-	1.40E-14	Pseudoknot of the domain G(G12) of 23*S* rRNA
1736778	1736668	-	1.40E-14	Pseudoknot of the domain G(G12) of 23*S* rRNA
ncRNA				
1465321	1465619	+	4.90E-54	Archaeal RNase P
882514	882820	+	4.10E-37	Archaeal signal recognition particle RNA
1892922	1892750	-	9.20E-29	Group II catalytic intron D1-D4-3
908140	908295	+	8.20E-08	Group II catalytic intron D1-D4-7
908845	908921	+	3.80E-06	Group II catalytic intron
1892698	1892603	-	6.30E-06	Group II catalytic intron D1-D4-1

The genome of ISO4-H5 encodes 47 tRNAs, including tRNAs corresponding to all 21 amino acids. Three tRNAs carry introns, namely methionine, tryptophan and tyrosine (Table 3.4.).

Table 3.4. Predicted tRNAs in the ISO4-H5 genome

tRNA	Number of tRNAs	Introns
tRNA-Ala	3	0
tRNA-Arg	4	0
tRNA-Asn	1	0
tRNA-Asp	1	0
tRNA-Cys	1	0
tRNA-Gln	2	0
tRNA-Glu	2	0
tRNA-Gly	3	0
tRNA-His	1	0
tRNA-Ile	1	0
tRNA-Leu	5	0
tRNA-Lys	2	0
tRNA-Met	3	1
tRNA-Phe	1	0
tRNA-Pro	3	0
tRNA-Ser	4	0
tRNA-Thr	3	0
tRNA-Trp	1	1
tRNA-Tyr	1	1
tRNA-Val	3	0
tRNA-Pyl	1	0
Pseudo tRNAs	1	0
Total	47	3

Codon and amino acid usage. The codon usage and the base content at the third codon position were calculated for each predicted coding gene sequence (Table 3.5.) and are summarised in Figure 3.3B, in addition to the CAI (Table A.3.2.). The predicted codon usage of ISO4-H5 has

a strong bias towards C at the third codon position (Figure 3.3A.). The amino acid usage is summarised in Figure 3.3C.

Pyrrolysine usage. ISO4-H5 possesses a complete operon predicted to encode the genes required for the biosynthesis of pyrrolysine and for the aminoacylation of a transfer RNA to pyrrolysine (Figure 3.6.). Table 3.6 lists the predicted pyrrolysine-contaning genes. The likelihood of amber codon read-through and incorporation of pyrrolysine is based on homology via BLASTp. The length of peptide extension beyond amber codon is included to indicate the likelihood of a mis-folded protein product.

IS elements. IS elements were predicted as described in Section 2.2.22. 48 IS elements were identified in the genome of ISO4-H5, and these elements account for 2.6% of the coding DNA sequence. Three different types of IS are found within the ISO4-H5 genome; transposons, group II introns and phage integrases. The IS elements found in ISO4-H5 belong to families IS200/IS605, IS481, IS5, IS1634, IS4, IS110, ISL3, IS91, IS481 and ISKra4, with IS200/IS605 family predicted with the highest copy number (Table 3.7.).

Table 3.5. Codon usage of ISO4-H5

Amino acid	Codon	% Total Codons	% encoded amino acid per codon
Ser (S)	UCA	0.5	7.1
	UCC	2.8	42.8
	UCG	1.0	15.5
	UCU	0.5	8.5
	AGC	1.3	19.4
	AGU	0.4	6.7
Phe (F)	UUC	3.6	92.3
	UUU	0.3	7.7
Thr (T)	ACA	0.6	10.4
	ACC	3.8	66.2
	ACG	0.7	12.4
	ACU	0.6	10.9
Asn (N)	AAC	3.0	76.9
	AAU	0.9	23.1
Lys (K)	AAA	1.7	29.3
	AAG	4.2	70.7
Glu (E)	GAA	2.3	34.1
	GAG	4.4	65.9
Tyr (Y)	UAC	2.6	70.4
	UAU	1.1	29.6
Val (V)	GUA	1.0	13.1
	GUC	3.5	47.5
	GUG	2.1	28.4
	GUU	0.8	11
Gln (Q)	CAA	0.3	10.9
	CAG	2.1	89.1
Met (M)	AUG	3.2	100.0
Cys (C)	UGC	1.2	74.5
	UGU	0.4	25.5
Leu (L)	CUA	0.2	2.7
	CUC	3.5	42.5
	CUG	2.8	34.4
	CUU	1.0	11.6
	UUA	0.2	2.7
	UUG	0.5	6.1

Amino acid	Codon	% Total Codons	% encoded amino acid per codon
Ala (A)	GCA	2.0	25.7
	GCC	3.5	44
	GCG	1.5	18.6
	GCU	0.9	11.8
O (Pyl)	UAG	0.0	100.0
Trp (W)	UGG	1.0	100.0
Pro (P)	CCA	0.2	5.3
	CCC	2.2	56.6
	CCG	0.8	20.1
	CCU	0.7	18.0
His (H)	CAC	1.1	67.4
	CAU	0.6	32.6
Asp (D)	GAC	4.0	61.6
	GAU	2.5	38.4
Arg (R)	CGA	0.0	0.3
	CGC	1.2	23.3
	CGG	0.2	3.3
	CGU	0.7	14.6
	AGA	0.7	14.8
	AGG	2.2	43.6
Ile (I)	AUA	1.0	14.3
	AUC	5.0	73.3
	AUU	0.8	12.4
Gly (G)	GGA	3.2	41.2
	GGC	1.9	25.2
	GGG	0.9	12.0
	GGU	1.7	21.6
Stop codons			
ochre	UAA	13.4	
amber	UAG	1.4 (2.7)*	
opal	UGA	84.1	
Translation initiator			
M	AUG	93.6	
L	GUG	1.7	
V	UUG	4.7	

*the number in brackets indicates total in-frame UAG, without considering its use as a STOP codon or as a Pyl-encoding codon

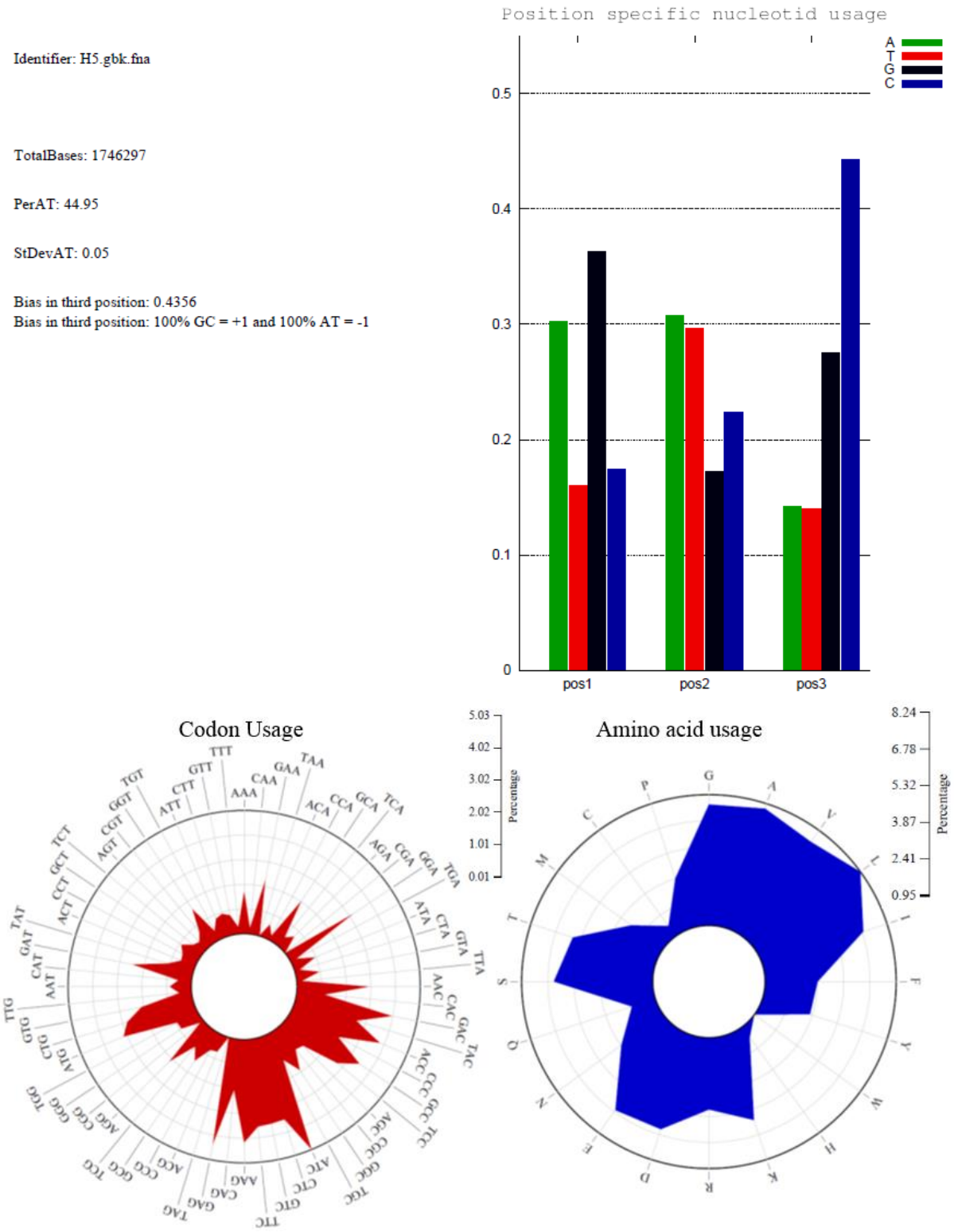


Figure 3.3. The predicted codon and amino acid usage of ISO4-H5. **A**. Base composition of codons in position one (pos1), two (pos2) and three (pos3). **B**. Codon usage percentage of ISO4-H5, scales are displayed on the right. **C**. Amino acid usage percentage of ISO4-H5, scales are displayed on the right.

Table 3.6. Genes within the ISO4-H5 genome predicted to incorporate pyrrolysine in their encoded proteins

Locus tag	Gene name	Number of codons to next STOP*	Predicted class[#]
AR505_0107	Transposase	9	2
AR505_0316	Hypothetical protein	105	3
AR505_0319	Hypothetical protein	104	1
AR505_0344	Hypothetical protein	372	3
AR505_0479	Hypothetical protein	30	1
AR505_0499	2-methylcitrate synthase/Citrate synthase II PrpC	336	1
AR505_0516	Transposase IS605 OrfB family	34	2
AR505_0523	Transposase IS605 OrfB family	47	2
AR505_0544	Phosphoglycolate/pyridoxal phosphate phosphatase family protein	38	1
AR505_0560	Bifunctional phosphoglucose/phosphomannose isomerase	177	1
AR505_0572	Hypothetical protein	25	1
AR505_0573	Hypothetical protein	133	1
AR505_0580	Transposase IS605 OrfB family	51	2
AR505_0639	Transposase IS605 OrfB family	116	2
AR505_0669	Integrase catalytic subunit	9	2
AR505_0670	Hypothetical protein	14, 3	1
AR505_0708	Transposase	9	2
AR505_0772	Trimethylamine:corrinoid methyltransferase MttB	163	1
AR505_0816	Adhesin-like protein	319	1
AR505_0865	Hypothetical protein	84	3
AR505_0872	Hypothetical protein	60	1
AR505_0880	Hypothetical protein	219	1
AR505_0948	Hypothetical protein	266, 23	3
AR505_0952	Methanol corrinoid protein MtaC1	190	1
AR505_1022	Hypothetical protein	41	3
AR505_1089	CRISPR-associate endonuclease Cas3-HD	592	1
AR505_1141	Hypothetical protein	202, 93, 69	3
AR505_1203	Methanogenesis marker 8	72	1
AR505_1213	TPR repeat-containing protein	329	1
AR505_1215	Hypothetical protein	916	3
AR505_1265	Hypothetical protein	98	1
AR505_1289	Nitrogenase	237	1
AR505_1326	Hypothetical protein	5, 1	3
AR505_1327	Monomethylamine methyltransferase MtmB1	258	1
AR505_1328	Monomethylamine methyltransferase MtmB2	259	1
AR505_1332	Dimethylamine:corrinoid methyltransferase MtbB	112	1
AR505_1423	Transposase	9	2
AR505_1433	Geranylgeranyl reductase	158	1
AR505_1571	Hypothetical protein	50	3
AR505_1592	Hypothetical protein	63	3
AR505_1618	Geranylgeranyl reductase	71	1
AR505_1640	Phage integrase	9	2
AR505_1734	Hypothetical protein	57	3
AR505_1776	Group II intron-encoding maturase	80	1
AR505_1784	Adenylate kinase Adk	20	2
AR505_1787	Glycosyl transferase family protein	305	1

*Distance between amber codon and next ochre/opal stop codon. [#]Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes

Table 3.7. IS elements of ISO4-H5

Locus tag	Family	Source of most similar sequence
AR505_0213	IS110	*Shewanella oneidensis*
AR505_0525	IS110	*Haloferax volcanii*
AR505_1703	IS110	*Haloferax volcanii*
AR505_0372	IS1634	*Mycoplasma mycoides*
AR505_0528	IS1634	*Mycoplasma mycoides*
AR505_0574	IS1634	*Mycoplasma bovis*
AR505_1563	IS1634	*Mycoplasma mycoides*
AR505_0010	IS200/IS605	*Methanosarcina mazei*
AR505_0215	IS200/IS605	*Methanosarcina mazei*
AR505_0276	IS200/IS605	*Methanosarcina mazei*
AR505_0290	IS200/IS605	*Methanosarcina mazei*
AR505_0371	IS200/IS605	*Methanosarcina mazei*
AR505_0498	IS200/IS605	*Methanosarcina mazei*
AR505_0516*	IS200/IS605	*Methanosarcina mazei*
AR505_0523*	IS200/IS605	*Methanosarcina mazei*
AR505_0580*	IS200/IS605	*Methanosarcina mazei*
AR505_0639*	IS200/IS605	*Methanosarcina mazei*
AR505_0677	IS200/IS605	*Methanosarcina mazei*
AR505_0860	IS200/IS605	*Methanosarcina mazei*
AR505_0903	IS200/IS605	*Methanosarcina mazei*
AR505_0913	IS200/IS605	*Methanosarcina mazei*
AR505_1219	IS200/IS605	*Methanosarcina mazei*
AR505_1692	IS200/IS605	*Methanosarcina mazei*
AR505_1765	IS200/IS605	*Methanosarcina mazei*
AR505_1775	IS200/IS605	*Methanosarcina mazei*
AR505_1777	IS200/IS605	*Methanosarcina mazei*
AR505_0440	IS4	*Lactobacillus reuteri*
AR505_1562	IS4	*Lactobacillus reuteri*
AR505_1714	IS4	*Lactobacillus reuteri*
AR505_0106	IS481	*Azotobacter vinelandii*
AR505_0107*	IS481	*Archaeoglobus fulgidus*
AR505_0708*	IS481	*Archaeoglobus fulgidus*
AR505_1423*	IS481	*Archaeoglobus fulgidus*
AR505_0669*	IS481	*Archaeoglobus fulgidus*
AR505_1640*	IS481	*Archaeoglobus fulgidus*
AR505_0147	IS5	*Methanosaeta thermophila*
AR505_0151	IS5	*Methanosaeta thermophila*
AR505_0784	IS5	*Methanosaeta thermophila*
AR505_1356	IS5	*Methanosaeta thermophila*
AR505_1458	IS5	*Methanosaeta thermophila*
AR505_0313	IS91	*Shewanella violacea*
AR505_0931	IS91	*Weeksella zoohelcum*
AR505_1543	IS91	*Azoarcus sp.*
AR505_1570	IS91	*Azoarcus sp.*
AR505_1697	IS91	*Azoarcus sp.*
AR505_1239	ISL3	*Acidithiobacillus caldus*
AR505_1574	ISL3	*Acidithiobacillus caldus*
Group II intron maturase		
AR505_1776*	ISKra4	*Legionella drancourtii*

*Contains an in-frame amber codon

CRISPR Elements. ISO4-H5 contains several CRISPR related genes (AR505_1089 – 1095) associated with a CRISPR region. The CRISPR element contains 35 repeats (bases 1153894 to 1155995), 29 bp in length. The consensus sequence of the repeat was GAGTTCCCCACGCATGTGGGGATGAACCG. The presence of the CRISPR associated proteins (AR505_1090, AR505_1091) suggested that the ISO4-H5 CRISPR/Cas system belongs to type I-E of the CRISPR-2 family. A total of 32 predicted spacer sequences with species-specific PAM (protospacer adjacent motif) were used to identify the potential targets of CRISPR RNAs (crRNAs). The predicted crRNA targets are displayed in Table 3.8, spacers with no database matches (spacers 1, 7, 8, 9, 10, 11, 13, 14, 16, 24, 26, 28, 29, 30, 31) were omitted.

Table 3.8. CRISPR spacer homology

Spacer Number	crRNA target*	Accession number	Score#
2	*Legionella pneumophila* str. Lorraine plasmid pLELO	NC_018141	18
3	KJ019151 *Synechococcus* phage ACG-2014f isolate Syn7803C24	KJ019146	18
4	*Burkholderia* sp. RPE64 plasmid p1	NC_021289	16
5	KF669658 *Acinetobacter* phage Presley	KF669658	16
6	*Halobacterium salinarum* R1 plasmid PHS2	NC_010369	16
12	*Silicibacter* sp. TM1040 mega plasmid	NC_008043	24
15	*Pantoea* sp. At-9b plasmid pPAT9B03	NC_014840	16
17	*Azospirillum brasilense* Sp245 plasmid AZOBR_p3	NC_016595	18
18	*Azospirillum lipoferum* 4B plasmid AZO_p2	NC_016586	18
19	*JF974292* Cyanophage S-SSM2	JF974292	16
20	HM152763 *Mycobacterium* phage LeBron	HM152763	16
21	*Klebsiella pneumoniae* strain CAV1344 plasmid pCAV1344-250	NZ_CP011623	18
22	*Klebsiella pneumoniae* subsp. pneumoniae strain 234-12 plasmid pKpn23412-362	NZ_CP011314	16
23	*Pantoea* sp. PSNIH1 plasmid pPSP-3a9	NC_CP010326	20
25	*Rahnella* sp. Y9602 plasmid pRAHAQ01	NC_015062	20
27	*Rhodococcus opacus* PD630 plasmid 1	NZ_CP003950	18
32	*Anabaena cylindrica* PCC 7122 plasmid pANACY.01	NC_019772	16

*spacer sequence identified by BLAST based homology screening (Biswas *et al.* 2013). #Score was calculated by matches (+1) and mismatches (-1) across the whole length of the spacer without gaps.

Horizontal gene transfer. There are 31 regions in the ISO4-H5 genome that are predicted to be horizontally acquired by Alien Hunter based on an atypical sequence composition (Table 3.9.), and these regions account for 17.2% of the genome. A total of 243 genes are predicted across the 31 regions. There are three regions with a high likelihood of being horizontally transferred, regions 4, 7 and 15.

Table 3.9. Predicted horizontal gene transfer regions of ISO4-H5

Region	Base range	IVOM score*	% G+C	Number of genes	Average CAI*	Likelihood
1	1..10000	19.666	50.96	5	0.610	Low
2	62500..67500	19.99	59.69	1	0.891	Low
3	72500..77500	26.576	59.01	2	0.804	Low
4	305000..330000	22.246	47.01	32	0.611	High
5	397500..405000	24.731	47.13	6	0.557	Moderate
6	530000..545000	22.182	46.86	12	0.594	Moderate
7	550000..582500	54.567	41.46	30	0.528	High
8	592500..597500	16.055	47.77	4	0.603	Low
9	657500..670000	27.715	46.51	8	0.585	Moderate
10	687500..695000	20.556	56.3	1	0.757	Low
11	735000..742500	21.367	57.99	1	0.738	Low
12	755000..760000	16.127	48.09	6	0.6215	Low
13	785000..792500	14.62	60.78	1	0.716	Low
14	890000..897500	18.259	47.74	10	0.637	Low
15	907500..950000	52.2	42.13	35	0.546	High
16	1047500..1057500	18.524	56.1	1	0.715	Low
17	1147500..1157500	17.559	49.02	10	0.572	Low
18	1222500..1232500	18.971	52.59	2	0.679	Low
19	1285000..1290000	15.88	48.79	2	0.667	Low
20	1337500..1345000	18.998	48.17	8	0.650	Low
21	1452500..1457500	17.144	48.17	5	0.602	Low
22	1530000..1535000	16.147	52.57	2	0.723	Low
23	1557500..1562500	20.939	46.63	4	0.586	Moderate
24	1567500..1572500	19.414	58.17	4	0.8565	Low
25	1632500..1647500	29.172	46.22	11	0.562	Moderate
26	1670000..1682500	23.294	47.58	5	0.572	Moderate
27	1685000..1692500	14.707	49.38	9	0.612	Low
28	1735000..1747500	19.641	53.71	6	0.775	Low
29	1827500..1845000	21.204	48.25	9	0.615	Moderate
30	1852500..1857500	14.157	48.73	7	0.618	Low
31	1860000..1865000	29.168	45.37	4	0.555	Moderate

*Atypical sequence composition is indicated by IVOM score, average CAI and % G+C. IVOM scores is annotation independent and it implemented variable order *k*-mers as reliable estimates of local sequence composition, as higher order motifs are more likely to capture deviation from genome background. A high regional IVOM score, abnormal % G+C and low average CAI is associated with a high likelihood of horizontal transfer.

Regions 7 and 15 have 42% G+C content in comparison to the average 54% G+C content of the genome. Both regions have atypical codon composition with low CAI, which indicates a bias in codon composition. More than half of the genes within these regions encode hypothetical proteins and the majority of these genes have top BLASTP hits originate from various non methanogenic micro-organisms. Transposases are also identified within, or adjacent to these regions. Additionally, region 7 contains an operon of putatively horizontally transferred genes, containing the dTDP-L-rhamnose biosynthesis genes *rfbA, rfbB, rfbc* (AR505_0552 – 0554) with corresponding transporters downstream. Region 4 has a %G+C content of 47%, 7% below the average genome G+C content of 54%. However, much like region 7 and region 15, a high number of genes within the region are hypothetical proteins. A large cluster of hypothetical genes within this region and extended outside this region (AR505_0333 to AR505_0358) are only present 3' to 5'.

Adhesin-like proteins. There are 45 genes predicted to be adhesin-like proteins and these are displayed in Table 3.10, showing their predicted subcellular location, molecular weight, transmembrane helices, signal peptide and protein domains. Three adhesin-like protein are predicted to contain a quinonprotein alcohol dehydrogenase-like domain (AR505_1521, 1549, 1707) and one adhesin-like protein is predicted to contain a pyrrolo-quinone beta-propeller repeat (AR505_1524). In addition, there are seven adhesin-like proteins with molecular weights above 200 kDa (AR505_0005, 0668, 0704, 0744, 0992, 1155, 1559).

Table 3.10. Adhesin-like proteins

Locus_tag	Size (kDa*)	TMH#	SignalP	SPAAN P_{ad}-value	LOCTree3	PSORTb	SubLoc	Domain
AR505_0005	**209.81**	Yes	Yes	0.91	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0061	197.01	Yes	Yes	0.96	Secreted	Cytoplasmic membrane	Extracellular	
AR505_0353	41.88	Yes	Yes	0.94	Secreted	Extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0354	24.54	Yes	Yes	0.81	Secreted	Unknown	Periplasmic	Listeria_Bacteriodes repeat (PF09479)
AR505_0355	33.34	No	Yes	0.86	Secreted	Extracellular	Extracellular	Putative Ig domain (PF05345)
AR505_0407	76.82	Yes	Yes	0.81	Secreted	Membrane or extracellular	Extracellular	Periplasmic binding protein (PF01497)
AR505_0594	73.45	Yes	Yes	0.81	Secreted	Extracellular	Extracellular	
AR505_0614	38.06	Yes	Yes	0.90	Plasma membrane	Membrane or extracellular	Extracellular	
AR505_0646	34.37	Yes	No	0.77	Secreted	Unknown	Extracellular	FKBP-type peptidyl-prolyl cis-trans isomerase (PF00254)
AR505_0654	24.21	Yes	Yes	0.86	Secreted	Extracellular	Extracellular	
AR505_0657	171.44	No	No	0.89	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0658	22.21	Yes	Yes	0.81	Secreted	Membrane or extracellular	Periplasmic	
AR505_0660	103.94	Yes	Yes	0.86	Secreted	Membrane or extracellular	Extracellular	
AR505_0664	189.16	No	Yes	0.74	Secreted	Membrane or extracellular	Cytoplasmic	
AR505_0666	74.83	Yes	No	0.83	Secreted	Extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0668	**381.43**	Yes	Yes	0.98	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0670	75.21	No	Yes	0.89	Secreted	Extracellular	Extracellular	
AR505_0704	**271.74**	Yes	Yes	0.83	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0744	**382.06**	Yes	Yes	0.94	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0807	74.08	Yes	Yes	0.84	Secreted	Unknown	Cytoplasmic	Listeria_Bacteriodes repeat (PF09479)
AR505_0851	45.14	Yes	Yes	0.77	Secreted	Cellwall	Cytoplasmic	Bacterial surface protein 26-residue repeat (TIGR02167)
AR505_0874	84.25	Yes	Yes	0.87	Secreted	Unknown	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_0985	23.78	Yes	Yes	0.81	Secreted	Extracellular	Extracellular	
AR505_0989	24.16	Yes	Yes	0.84	Secreted	Extracellular	Extracellular	
AR505_0991	34.41	Yes	Yes	0.88	Secreted	Extracellular	Extracellular	
AR505_0992	**436.78**	Yes	Yes	0.97	Secreted	Cellwall	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1032	49.03	Yes	No	0.82	Secreted	Extracellular	Extracellular	
AR505_1033	57.87	Yes	Yes	0.87	Secreted	Extracellular	Extracellular	Bacterial Ig-like domain (group 2) (PF02368)
AR505_1155	**294.67**	Yes	Yes	0.95	Secreted	Cytoplasmic membrane	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1173	43.00	Yes	Yes	0.77	Secreted	Membrane or extracellular	Extracellular	
AR505_1290	146.37	Yes	Yes	0.83	Secreted	Membrane or extracellular	Extracellular	
AR505_1509	77.41	Yes	Yes	0.80	Secreted	Unknown	Extracellular	Bacterial Ig-like domain (group 2) (PF02368)
AR505_1521	95.56	Yes	Yes	0.70	Secreted	Membrane or extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1524	66.24	Yes	Yes	0.75	Secreted	Extracellular	Extracellular	PQQ enzyme repeat (PF01011)
AR505_1532	95.53	Yes	Yes	0.76	Secreted	Membrane or extracellular	Periplasmic	
AR505_1534	91.42	No	Yes	0.87	Secreted	Membrane or extracellular	Extracellular	Listeria_Bacteriodes repeat (PF09479)
AR505_1547	39.47	Yes	Yes	0.74	Secreted	Membrane or extracellular	Periplasmic	
AR505_1549	65.82	Yes	Yes	0.77	Secreted	Extracellular	Extracellular	PQQ enzyme repeat (PF01011)
AR505_1559	**264.61**	Yes	Yes	0.84	Secreted	Cytoplasmic membrane	Cytoplasmic	Listeria_Bacteriodes repeat (PF09479)
AR505_1560	52.10	Yes	Yes	0.83	Secreted	Extracellular	Extracellular	Cohesin domain (PF00963)
AR505_1561	36.76	Yes	Yes	0.84	Secreted	Extracellular	Extracellular	
AR505_1707	76.48	Yes	Yes	0.70	Secreted	Membrane or extracellular	Periplasmic	PQQ enzyme repeat (PF01011)
AR505_1715	39.67	No	Yes	0.82	Secreted	Extracellular	Extracellular	
AR505_1741	41.34	Yes	Yes	0.74	Secreted	Extracellular	Extracellular	
AR505_1761	35.99	Yes	No	0.91	Secreted	Unknown	Extracellular	

*predicted molecular weight in kda, proteins above 200 kDa are displayed in bold. # TMH: transmembrane helices.

Secretome. The ISO4-H5 genome encodes 101 ORFs predicted to contain a signal peptide (Figure 3.4A.). Among the ORFeome predicted to be exported, most ORF functions have been classified under either cell envelope, hypothetical or transporters (Figure 3.4B.).

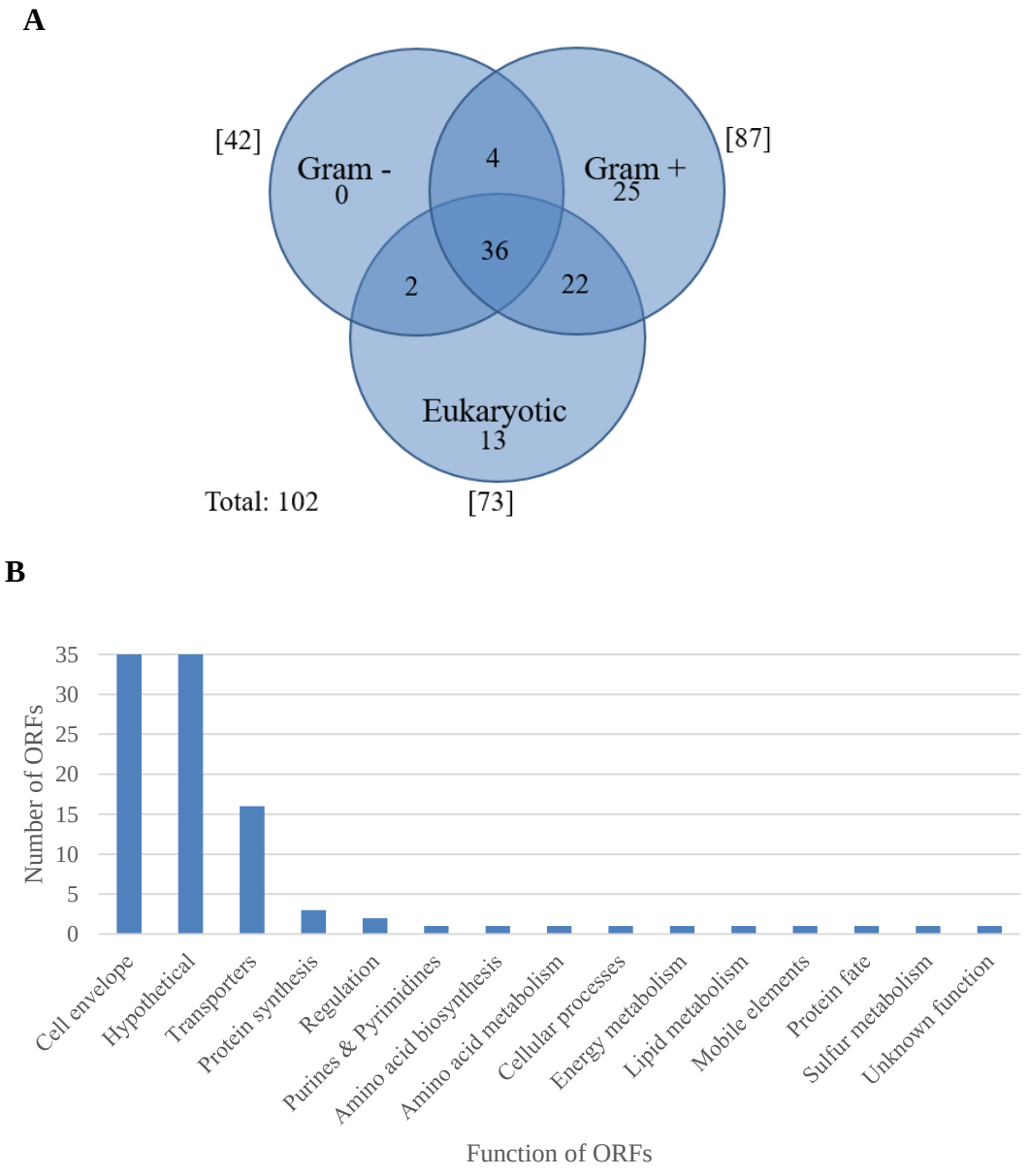


Figure 3.4. Classification of ISO4-H5 signal peptides and their associated ORFs **A**. Signal peptides classified according to SignalP training models. **B**. Functional classification of ISO4-H5 ORFs containing signal peptides

3.2.4. Metabolic pathway reconstruction

ISO4-H5 genes with a predicted metabolic function are listed in Table A.3.1. A total of 716 genes, representing 49% of the genome, have been assigned to a metabolic pathway.

Energy metabolism

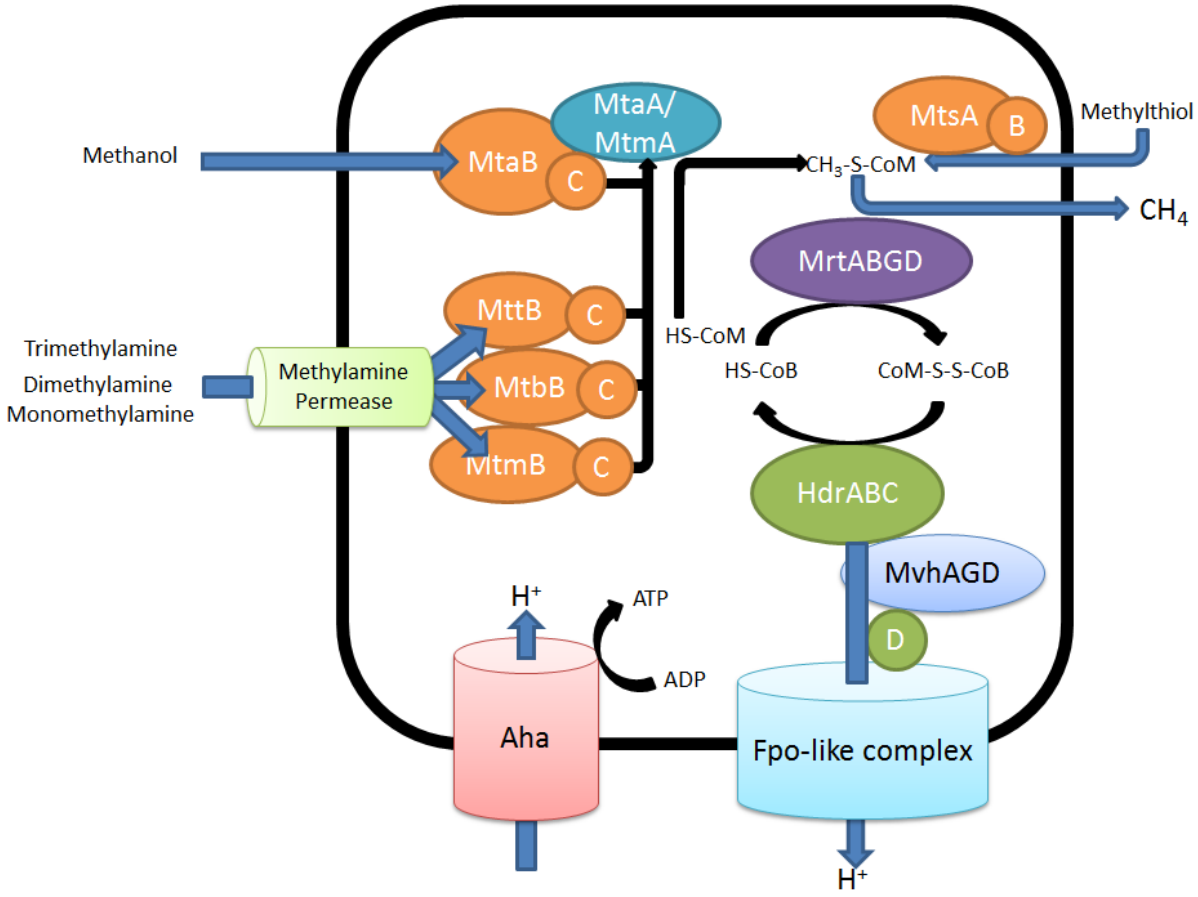


Figure 3.5. Methylotrophic methanogenesis of ISO4-H5 The enzymes involved are: methanol:corrinoid methyltransferase (MtaB), monomethylamine:corrinoid methyltransferase (MtmB), dimethylamine:corrinoid methyltransferase (MtbB), trimethylamine:corrinoid methyltransferase (MttB), methylcobalamin: CoM methyltransferase (MtaA), methylthiol: CoM methyltransferase (MtsA). The corresponding cognate corrinoid protein to each methyltransferase (C) with exception to methylthiol as (B). Methyl CoM reductase (MrtABGD), heterodisulfide reductase (HdrABC and D), methyl viologen hydrogenase (MvhAGD). Components involved in membrane potential generation and energy generation are Fpo-like complex and ATP synthase (Aha)

Methanogenesis. Based on the genome sequence, methanogenesis is the sole energy generating mechanism of ISO4-H5. The genes predicted to be involved in methanogenesis are summarized in Figure 3.5. The ISO4-H5 genome encodes genes required for energy generation via methylotrophic methanogenesis, but lacks the genes encoding hydrogenotrophic methanogenesis prior to the methyl coenzyme reductase (Mrt) step. ISO4-H5 is strictly dependent on direct reduction of methylated compounds for methylotrophic methanogenesis.

The ISO4-H5 genome encodes specific methyltransferases for utilising methylated compounds as substrates, including methanol, methylthiol, mono-, di-, and tri-methylamine (Figure 3.6.). The *mta* operon required for methanol utilisation consists of two methanol:corrinoid methyltransferases, *mtaB1*, *mtaB2* (AR505_0949, AR505_0951), and two associated methanol corrinoid proteins, *mtaC1, mtaC2* (AR505_0950, AR505_0952) (Figure 3.6.). The *mtaC*2 (AR505_0952) gene is predicted to require amber codon read-through.

The monomethylamine and dimethylamine utilisation genes are also present in operons, including two copies each of monomethylamine:corrinoid methyltransferase *mtmB* (AR505_1327 and AR505_1328), and the corresponding monomethylamine corrinoid protein *mtmC* (AR505_1329 and AR505_1330), in addition to a dimethylamine permease *mtbP* (AR505_1331), a dimethylamine:corrinoid methyltransferase *mtbB* (AR505_1332), and a dimethylamine corrinoid protein *mtbC* (AR505_1333) (Figure 3.6.). The trimethylamine utilisation genes, trimethylamine:corrinoid methyltransferase *mttB* and the corresponding trimethylamine corrinoid protein *mttC,* form a small operon (AR505_0772, 0773) far upstream of the other methylamine utilisation genes (Figure 3.6.). The pyrrolysine biosynthesis operon (AR505_1322 – 1325) and *ramA* (AR505_1320) were identified adjacent to the methylamine utilisation operon (Figure 3.6.). Four homologues of methylcobalamin: CoM methyltransferase *mtaA/mtmA* were also found in the ISO4-H5 genome.

ISO4-H5 possesses genes encoding for methylthiol utilisation. AR505_1066 is highly homologous to methylthiol: CoM methyltransferase from *Methanosarcina barkeri*, with 71.6% aa identity, and the corresponding corrinoid protein AR505_1067 shares 50.2% aa identity to the *M. barkeri* methylthiol corrinoid protein. The methylthiol utilising operon is located adjacent to one of the methyltransferase cognate corrinoid genes (although with uncertain substrate specificity) and one of the *mtaA* genes (Figure 3.6.).

The reduction of methyl-CoM to CH_4 is carried out by the methyl-CoM reductase enzyme complex (*mrt*). ISO4-H5 harbors the gene organization of a MCRII/Mrt-type system, with four genes arranged in an operon structure, *mrtBDGA* (AR505_1399 to AR505_1396), with *mrtC* (AR505_1391) and methyl-CoM reductase component A1 *atwA* (AR505_1392) located downstream (Figure 3.6.).

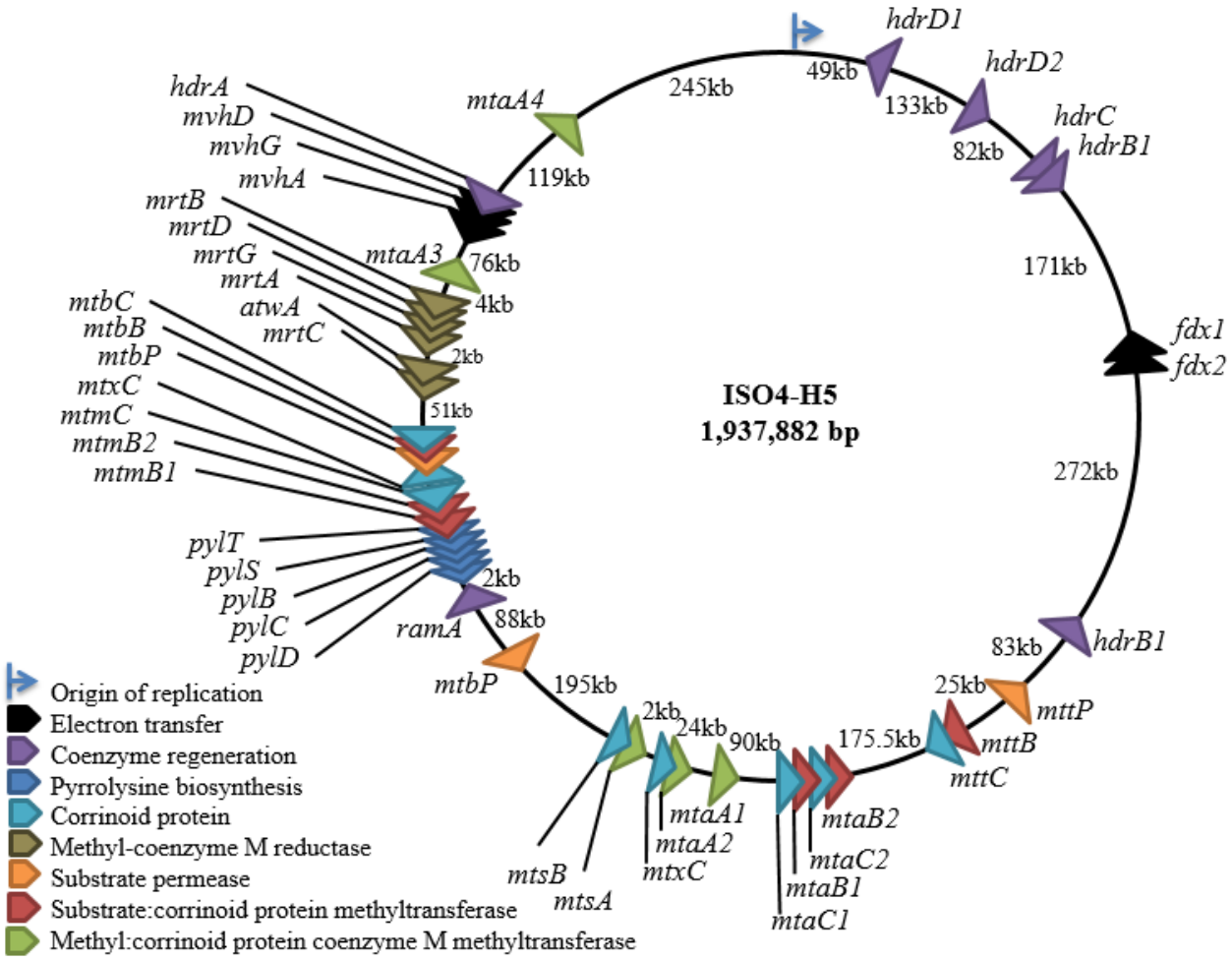


Figure 3.6. Methanogenesis genes of ISO4-H5. Genes predicted to be involved in methanogenesis in the ISO4-H5 genome are displayed with gene name followed by locus tag number. Heterodisulfide reductase (*hdr*), ferredoxin (*fdx*), trimethylamine permease (*mttP*), trimethylamine:corrinoid methyltransferase (*mttB*), trimethylamine corrinoid protein (*mttC*), methanol:corrinoid methyltransferase (*mtaB*), methanol corrinoid protein (*mtaC*), methyl: CoM methyltransferase (*mtaA*), corrinoid protein with unknown substrate (*mtxC*), bifunctional methylthiol:corrinoid methyltransferase (*mtsA*), methylthiol corrinoid protein (*mtsB*), dimethylamine permease (*mtbP*), methylamine methyltranferase corrinoid activation protein (*ramA*), pyrrolysine biosynthesis protein (*pyl*), amber suppressor tRNA (*pylT*), monomethylamine:corrinoid methyltransferase (*mtmB*), monomethylamine corrinoid protein (*mtmC*), dimethylamine:corrinoid methyltransferase (*mtbB*), dimethylamine corrinoid protein (*mtbC*), methyl-CoM reductase (*mrt*), methylviologen hydrogenase (*mvh*). Distance between genes are displayed in kb. Figure not drawn to scale.

An alcohol dehydrogenase gene, *adhE* (AR505_0483), homologous to an alcohol dehydrogenase identified in *Clostridium acetobutylicum* ATCC 824 and with a COG match to the alcohol dehydrogenase class IV Adh7 domain is also present in the ISO4-H5 genome.

Methanogenesis cofactors. During the last step of methanogenesis, coenzyme B (CoB) is used by the Mrt complex for methyl-CoM reduction to produce CH_4 (Figure 3.5.). This process utilises the reducing potential carried by HS-CoB and generates CoM-S-S-CoB (heterodisulfide). Both cofactors need to be recycled into an active state, and this is carried out

by a heterodisulfide reductase (Hdr). ISO4-H5 encodes four Hdr subunits (HdrABCD), with *hdrB* and *hdrC* located in an operon (AR505_0273, AR505_0274), while *hdrA* (AR505_1479) and a second copy of *hdrB* (AR505_0679) are located elsewhere on the genome (Figure 3.6.). There are two copies of *hdrD* (AR505_0040, AR505_0168) in the genome without the usual partner gene *hdrE.* ISO4-H5 encodes the methyl viologen hydrogenase MvhAGD required for coupling H_2 to provide reducing potential for the reducing CoM-S-S-CoB and Fdx via electron bifurcation (*mvhADG*, AR505_1476 – 1478) (Figure 3.6.).

Generation of membrane potential from methanogenesis. The protons produced by regeneration of cofactors involved in methanogenesis is predicted to be removed from the cell by a proton pump, which generates a membrane potential to drive the ATP synthase. Two operons consisting of 11 genes encoding $F_{420}H_2$ dehydrogenase (Fpo) have been found in ISO4-H5 (*fpoABCDHIJKLMN*, include genes, AR505_1622-1625, AR505_1627-1633). Two subunits, *fpoF* and *fpoO*, found in other methanogens are not encoded in the ISO4-H5 genome. Homologues of the nine A_1A_0 ATP synthase subunits are present (*ahaHIKECFABD*, include genes, AR505_1818- 1826) and is predicted to utilise the membrane potential generated by the Fpo-like complex to produce ATP.

Methanogenesis marker proteins. Homologues of methanogenesis marker proteins 1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 15, 16, and 17 are present within the ISO4-H5 genome, while marker proteins 9, 10, 12 and 14 are absent.

Central carbon metabolism

ISO4-H5 is predicted to encode the genes required for a partial TCA cycle, gluconeogenesis and the pentose phosphate pathway.

One-carbon metabolism. ISO4-H5 has an incomplete 1-carbon metabolism and is incapable of producing acetyl-CoA *de novo* from CO and formate, because it does not encode methyltetrahydrofolate:corrinoid methyltransferase, 5,10-methylenetetrahydrofolate reductase or the carbon-monoxide dehydrogenase complex (*cdhABCDE*) (Figure 3.7., Table 3.11.). ISO4-H5 has *folD*, which encodes the bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/5,10-methylene-tetrahydrofolate cyclohydrolase (AR505_1639) used to convert N^{10}-methenyl-tetrahydrofolate to 5,10-methylene-tetrahydrofolate or N^{10}-formyl-tetrahydrofolate.

ISO4-H5 encodes an acetyl-CoA synthetase, *acs*, as well as a holoenzyme of pyruvate synthase *porABCD* (Table 3.11.). Acetate is predicted to be the carbon source of ISO4-H5. The acetyl-CoA synthetase appears to be a single, fused enzyme instead of the usual two subunits found in other organisms. The incorporation of acetyl-CoA into the TCA cycle produces pyruvate, which proceeds towards gluconeogenesis and subsequently the pentose phosphate pathway.

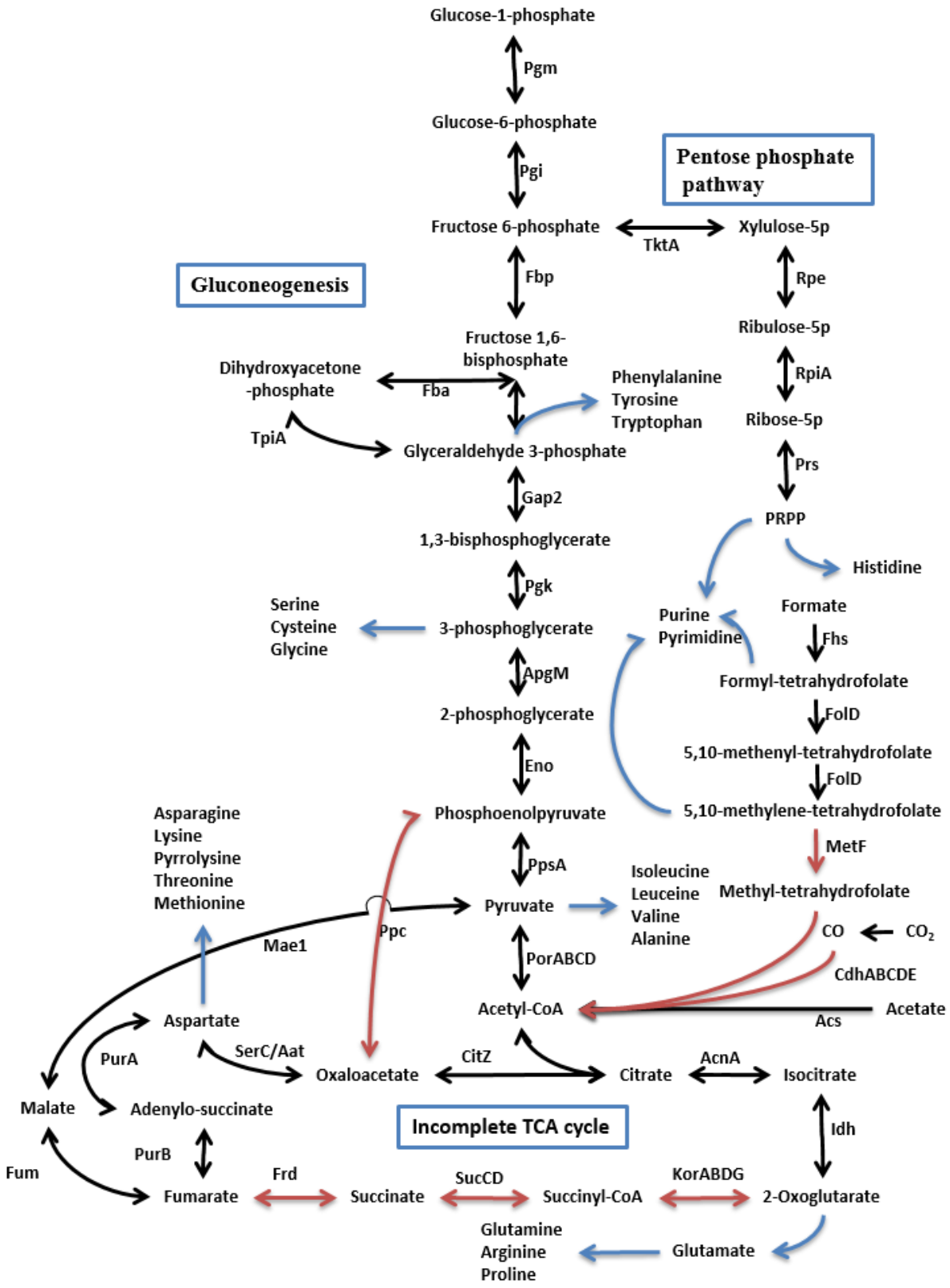

Glucose-1-phosphate
Pgm
Glucose-6-phosphate
Pgi
Pentose phosphate pathway
Fructose 6-phosphate
TktA
Xylulose-5p
Gluconeogenesis
Fbp
Rpe
Ribulose-5p
Fructose 1,6-bisphosphate
RpiA
Dihydroxyacetone -phosphate
Fba
Phenylalanine
Tyrosine
Tryptophan
Ribose-5p
TpiA
Glyceraldehyde 3-phosphate
Prs
Gap2
PRPP
1,3-bisphosphoglycerate
Histidine
Pgk
Serine
Cysteine
Glycine
3-phosphoglycerate
Purine
Pyrimidine
Formate
Fhs
ApgM
Formyl-tetrahydrofolate
2-phosphoglycerate
FolD
Eno
5,10-methenyl-tetrahydrofolate
FolD
Asparagine
Lysine
Pyrrolysine
Threonine
Methionine
Phosphoenolpyruvate
5,10-methylene-tetrahydrofolate
MetF
PpsA
Isoleucine
Leuceine
Valine
Alanine
Methyl-tetrahydrofolate
Pyruvate
Ppc
Mae1
CO
CO2
PorABCD
CdhABCDE
Acetyl-CoA
Acetate
Acs
Aspartate
PurA
SerC/Aat
CitZ
AcnA
Oxaloacetate
Citrate
Isocitrate
Malate
Adenylo-succinate
Incomplete TCA cycle
Idh
PurB
Fum
Frd
SucCD
KorABDG
Fumarate
Succinate
Succinyl-CoA
2-Oxoglutarate
Glutamine
Arginine
Proline
Glutamate

Figure 3.7. ISO4-H5 central carbon metabolism. Black arrow represents functional pathways, red arrow represents pathways absent in ISO4-H5 but present in other members of Methanomassiliicoccales, blue arrows represents connections to other biosynthetic pathways. Formate-tetrahydrofolate ligase (Fhs), NADP-dependent methylene tetrahydrofolate dehydrogenase (FolD), 5,10-methylenetetrahydrofolate reductase (MetF), carbon-monoxide dehydrogenase complex (CdhABCDE), acetyl CoA synthetase (Acs), phosphoglucomutase (Pgm), phosphoglucose isomerase (Pgi), fructose-1,6-bisphosphatatase (Fbp), fructose-bisphosphate aldolase (Fba), glyceraldehyde-3-phosphate dehydrogenase (Gap2), phosphoglycerate kinase (Pgk), phosphoglycerate mutase (ApgM), phosphopyruvate hydratase (Eno), phosphoenolpyruvate synthase (PpsA), pyruvate:fdx oxidoreductase (PorABCD), citrate synthase (CitZ), aconitate hydratase (AcnA), isocitrate dehydrogenase (Idh), 2-oxoglutarate synthase (KorABDG), succinyl-CoA synthetase (SucCD), fumarate reductase (Frd), fumarate hydratase (Fum), adenylosuccinate lyase (PurB), adenylosuccinate synthetase (PurA), phosphoserine aminotransferase/aspartate aminotransferase (SerC/Aat), malate dehydrogenase (Mae1), phosphoenolpyruvate carboxylase (Ppc), transketolase (TktA), ribulose-phosphate 3-epimerase (Rpe), ribose-5-phosphate isomerase (RpiA), ribose-phosphate diphosphokinase (Prs), phosphoribosyl pyrophosphate (PRPP).

Table 3.11. Genes involved in central carbon metabolism

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimental validated homologue
AR505_1217	Formate--tetrahydrofolate ligase	*fhs*	52.1	Rankin *et al.*, 1993
AR505_1639	Bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	*folD*	44.4	Shen *et al.*, 1999
AR505_0076	Thymidylate synthase	*thyX*	27.8	Myllykallio *et al.*, 2002
AR505_1282	Acetyl-CoA synthetase (ADP-forming)	*acs*	36.4, 46.0*	Musfeldt *et al.*, 1999
AR505_0431	Pyruvate synthase subunit α	*porA*	39.9	Smith *et al.*, 1997
AR505_0432	Pyruvate synthase subunit β	*porB*	43.0	Smith *et al.*, 1997
AR505_0429	Pyruvate synthase subunit γ	*porC*	45.1	Smith *et al.*, 1997
AR505_0430	Pyruvate synthase subunit δ	*porD*	37.6	Smith *et al.*, 1997
AR505_0022	Fumarate hydratase	*fumA*	43.5	Shieh & Whitman, 1987
AR505_0023	Fumarate hydratase	*fumB*	45.3	Shieh & Whitman, 1987
AR505_1780	Malate dehydrogenase	*mae1*	12.3	Shieh & Whitman, 1987, Shieh & Whitman, 1987
AR505_0742	Adenylosuccinate lyase	*purB*	25.6	Zhang *et al.*, 2008
AR505_1168	Adenylosuccinate synthetase	*purA*	44.1	Zhang *et al.*, 2008
AR505_0499,	Citrate synthase	*citZ1*,	28.2,	Jongsareejit *et al.*, 1997
AR505_0678		*citZ2*	9.8	Khomyakova *et al.*, 2011
AR505_0592	Aconitate hydratase	*acnA*	51.2	Gruer & Guest, 1994
AR505_0531	Isopropylmalate/isocitrate dehydrogenase	*idh/leuB1*	22.3	Pitson *et al.*, 1999
AR505_1608	Phosphoglucomutase	*pgm*	35.5	Jansen *et al.*, 1982
AR505_0560,	Bifunctional	*pgi1, pgi2*	31.3, 31.6	Hansen *et al.*, 2004
AR505_1769	phosphoglucose/phosphomannose isomerase			
AR505_1149	Fructose-1,6-bisphosphatatase	*fbp*	50.1	Jansen *et al.*, 1982
AR505_0932	Fructose-bisphosphate aldolase	*fba*	55.5	Jansen *et al.*, 1982
AR505_0154	Glyceraldehyde-3-phosphate dehydrogenase	*gap2*	49.3	Jansen *et al.*, 1982
AR505_0155	Phosphoglycerate kinase	*pgk*	41.0	Jansen *et al.*, 1982
AR505_0474	Phosphoglycerate mutase	*apgM*	18.0	Jansen *et al.*, 1982
AR505_0942	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	*gpmA*	60.5	Fraser *et al.*, 1999
AR505_0470	Phosphopyruvate hydratase	*eno*	40.3	Jansen *et al.*, 1982
AR505_0427,	Phosphoenolpyruvate synthase	*ppsA1*	38.9	Narindrasorasak & Bridger,
AR505_1104		*ppsA2*	18.8	1977
AR505_1648	Transketolase subunit A	*tktA*	31.8	Sprenger *et al.*, 1995
AR505_1649	Transketolase subunit B	*tktB*	11.8	Sprenger *et al.*, 1995
AR505_1647	Transaldolase	*talB*	19.5	Sprenger *et al.*, 1995
AR505_1685	Ribose-phosphate riphosphokinase	*prs*	33.4	Hove-Jensen *et al.*, 1986
AR505_0461,	Ribulose phosphate epimerase	*rpe*	40.8	Sprenger, 1995
AR505_1604		*araD*	22.2	
AR505_0014	Bifunctional hexulose-6-phosphate synthase/ribonuclease regulator	*hxlA*	22.2	Yasueda *et al.*, 1999
AR505_0177	6-phospho 3-hexuloisomerase	*hxlB*	30.8	Yasueda *et al.*, 1999

*corresponds to alpha and beta subunit respectively

TCA cycle. ISO-H5 is predicted to have an incomplete TCA cycle (Figure 3.7., Table 3.11.). ISO4-H5 has the *cit*Z gene which encodes citrate synthase required to condense acetyl-CoA and oxaloacetate to produce citrate. Citrate can be further converted to 2-oxoglutarate by aconitate hydratase AcnA and isocitrate dehydrogenase Idh. In absence of malate dehydrogenase, ISO4-H5 encodes adenylosuccinate lyase PurB and adenylosuccinate synthetase PurA which convert fumarate to oxaloacetate via aspartate and the broad specificity phosphoserine/aspartate aminotransferases (SerC/Aat).

Gluconeogenesis. ISO4-H5 encodes all of the genes required for gluconeogenesis from pyruvate to glucose-6-phosphate (Figure 3.7., Table 3.11.). Two homologues of the bifunctional phosphoglucose/phosphomannose isomerase (Pgi1, Pgi2) were found, with Pgi1 (AR505_0560) requiring amber codon read-through to translate the full length protein. The ISO4-H5 genome is predicted to harbour two isozymes of phosphoglycerate mutase; 2,3-bisphosphoglycerate-dependent mutase (GpmA) and 2,3-bisphosphoglycerate-independent mutase (ApgM).

Pentose phosphate pathway. The ISO4-H5 genome encodes all the genes required for non-oxidative pentose phosphate pathway and phosphoribosyl pyrophosphate (PRPP) production (Figure 3.7., Table 3.11.). The *hxlAB* genes do not exist as an operon in ISO4-H5, and only the *N*-terminal domain of the predicted HxlA protein showed homology to characterised proteins.

Amino acid biosynthesis

The ISO4-H5 is predicted to be able to make 19 amino acids based on the genes present in the genome and the genes involved, their predicted enzymatic activities, and homology to characterised enzymes are summarised in Table 3.12. ISO4-H5 is predicted to be unable to produce methionine and glycine, as previously mentioned, ISO4-H5 encodes pyrrolysine biosynthesis.

Table 3.12. Genes encoding amino acid biosynthesis

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
Glutamate/glutamine				
AR505_0145	Glutamate dehydrogenase	*gdhA*	55.4	Sakamoto *et al.* 1975
AR505_0091	Glutamine synthetase	*glnA*	44.3	Amaya *et al.* 2005
AR505_0205	Ammonium transporter	*atmB*	38.2	Soupene *et al.* 1998
Arginine				
AR505_0674	Ornithine acetyltransferase	*argJ*	38.2	Hoch, Losick, & Sonenshein, 1993
AR505_0675	Acetylglutamate kinase	*argB*	27.5	Hoch *et al.*, 1993
AR505_0673	*N*-acetylglutamylphosphate reductase	*argC*	38.0	Hoch *et al.*, 1993
AR505_0676	Acetylornithine aminotransferase	*argD*	39.1	Hoch *et al.*, 1993
AR505_1401	Ornithine carbamoyltransferase	*argF*	45.4	Hoch *et al.*, 1993
AR505_0672	Argininosuccinate synthase	*argG*	51.3	Hoch *et al.*, 1993
AR505_0671	Argininosuccinase	*argH*	45.1	Hoch *et al.*, 1993
AR505_0501, AR505_0504	Carbamoyl-phosphate synthetase	*carA*	38.6, 40.5	Hoch *et al.*, 1993
AR505_0500, AR505_0503	Carbamoyl-phosphate synthetase	*carB*	44.8, 44.2	Hoch *et al.*, 1993
AR505_0268	Pyruvoyl-dependent arginine decarboxylase	*pdaD*	29.6	Fukuda *et al.* 2008
AR505_1607	Agmatinase	*speB*	31.0	Sekowska *et al.* 1998
Alanine				
AR505_1199	Alanine aminotransferase	*alt*	52.3	Ward, Kengen *et al.*, 2000
AR505_1429	Aspartate/tyrosine/aromatic aminotransferase	*aat*	36.0	Porat *et al.*, 2004.
AR505_1666	Aspartate/tyrosine/aromatic aminotransferase	*aat*	37.3	Porat *et al.*, 2004.
Aspartate/asparagine				
AR505_0736	Asparagine synthase B1	*asnB1*	19.2	Scofield *et al.*, 1990
AR505_1303	Asparagine synthase B2	*asnB2*	18.7	Scofield *et al.*, 1990
Lysine				
AR505_0160	Aspartate kinase	*lysC*	27.7	Yoshioka *et al.*, 2001
AR505_0491	Aspartate-semialdehyde dehydrogenase	*asd*	22.9	Paris *et al.*, 2002
AR505_0161	Dihydrodipicolinate synthase	*dapA*	29.6	Silk *et al.*, 1994
AR505_0728	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	38.5	Devenish *et al.*, 2010
AR505_0157	L,L-diaminopimelate aminotransferase	*dapL*	29.7	Liu *et al.*, 2010
AR505_0158	Diaminopimelate epimerase	*dapF*	37.5	Hor *et al.*, 2010
AR505_0159	Diaminopimelate decarboxylase	*lysA*	28.7	Andre'O *et al.*, 2005
Pyrrolysine				
AR505_1325	Pyrrolysine--tRNA ligase	*pylS*	28.6	Srinivasan *et al.*, 2002
AR505_1324	Methylornithine synthase	*pylB*	39.4	Quitterer *et al.*, 2012
AR505_1323	(2*R*,3*R*)-3-methylornithyl-N^6-lysine synthase	*pylC*	32.4	Gaston *et al.*, 2011b
AR505_1322	Pyrrolysine synthase	*pylD*	31.3	Quitterer *et al.*, 2013
Methionine				
AR505_0293	Homoserine dehydrogenase	*metL*	22.9	Belfaiza *et al.*, 1984
AR505_0694	Homoserine *O*-succinyltransferase	*metA*	45.4	Born & Blanchard, 1999
AR505_0738	*O*-acetylhomoserine sulfhydrylase	*met17*	43.4, 40.7	Yamagata *et al.*, 1994, Moore & Thompson, 1967
AR505_0466, AR505_0579	*S*-adenosylmethionine synthetase	*metK1*, *metK2*	14.4, 63.4	Markham *et al.*, 1980
AR505_0327, AR505_0339, AR505_0340	DNA-cytosine methyltransferase	*dcm1*, *dcm2*, *dcm3*	26.2, 20.5, 18.8	May & Hattman, 1975
AR505_0547, AR505_1788	*S*-adenosyl-L-homocysteinase	*ahcY1*, *ahcY2*	39.9, 39.9	Ogawa *et al.*, 1987
AR505_1118	*S*-adenosylhomocystein nucleosidase	*mtn*	37.6	Cornell & Riscoe, 1998
Threonine				
AR505_0610	Homoserine kinase	*thrB*	34.7	White, 2003
AR505_0611	Threonine synthase	*thrC*	54.2	Bult *et al.*, 1996

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
Chorismate				
AR505_1148	Triose-phosphate isomerase	*tpiA*	44.3	White & Xu, 2006
AR505_0508, AR505_0932	Fructose 1,6-bisphosphate aldolase	*fba1, fba2*	23.8, 21.0	White & Xu, 2006
AR505_0509	Dehydroquinate synthase	*aroB*	41.8	White, 2004
AR505_0510	Shikimate dehydrogenase	*aroE*	35.1	Padyana & Burley, 2003
AR505_0511	Shikimate kinase	*aroL*	14.7	Millar, Lewendon, Hunter, & Coggins, 1986
AR505_0512	3-phosphoshikimate-1-carboxyvinyltransferase	*aroA*	29.7	Anderson, Sikorski, & Johnson, 1988
AR505_1449	Chorismate synthase	*aroC*	36.3	White, Millar, & Coggins, 1988
Tryptophan				
AR505_1158	Anthranilate synthase component I	*trpE*	38.0	Tang *et al.*, 1999
AR505_1159	Anthranilate synthase component II	*trpG*	39.7	Tang *et al.*, 1999
AR505_0997, AR505_1160	Anthranilate phosphoribosyltransferase	*trpD1*, *trpD2*	39.2, 40.4	Tang *et al.*, 1999
AR505_1161	Indole-3-glycerol phosphate synthase	*trpC*	35.1	Tang *et al.*, 1999
AR505_1162	Phosphoribosyl anthranilate isomerase	*trpF*	33.3	Tang *et al.*, 1999
AR505_1163	Tryptophan synthase beta subunit	*trpB*	62.3	Tang *et al.*, 1999
AR505_1164	Tryptophan synthase alpha subunit	*trpA*	35.4	Tang *et al.*, 1999
Phenylalanine and tyrosine				
AR505_0490, AR505_1450	Chorismate mutase	*aroH1*, *aroH2*	25.8, 20.7	Porat *et al.*, 2004
AR505_1745	Prephenate dehydrogenase	*tyrA*	18.1	Porat *et al.*, 2004
AR505_0515	Prephenate dehydratase	*pheA*	27.1	Fischer & Jensen, 1987
Proline				
AR505_0108	Glutamate-5-semialdehyde dehydrogenase	*proA*	56.1	Zhang *et al.*, 2002
AR505_0109	Glutamate 5-kinase	*proB*	49.6	J. K. Zhang *et al.*, 2002
AR505_1634	Pyrroline-5-carboxylate reductase	*proC*	33.5	Kenklies *et al.*, 1999
Histidine				
AR505_0494	ATP phosphoribosyl transferase	*hisG*	32.5	Alifano *et al.*, 1996
AR505_0486	Phosphoribosyl-ATP pyrophosphatase/phosphoribosyl-AMP cyclohydrolase	*hisI*	38.9	Alifano *et al.*, 1996
AR505_0497	[*N*-(5-phosphoribosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase	*hisA*	33.5	Margolies & Goldberger, 1966
AR505_0496	IGP synthase glutamine amidotransferase subunit	*hisH*	28.4	Demin *et al.*, 2004
AR505_0487	IGP synthase cycloligase subunit	*hisF*	41.4	Demin *et al.*, 2004
AR505_0488	IGP dehydratase	*hisB*	33.8	Brilli & Fani, 2004
AR505_0495	Histidinol-phosphate aminotransferase	*hisC*	26.5	Haruyama *et al.*, 2001
AR505_1770	Histidinol phosphate phosphatase	*hisK*	13.5	Omi *et al.*, 2004
AR505_1073	Histidinal dehydrogenase	*hisD*	40.3	Andorn & Aronovitch, 1982
Serine				
AR505_1664	Phosphoglycerate dehydrogenase	*serA*	39.2	Ho & Saito, 2001
AR505_0073, AR505_0757	Phosphoserine phosphatase	*serB1*, *serB2*	21.7, 12.1	Neuwald & Stauffer, 1985
AR505_1665	Phosphoserine aminotransferase	*serC*	29.0	Helgadóttir *et al.*, 2007
Cysteine				
AR505_1192	Serine acetyltransferase	*cysE*	36.9	Denk & Bock, 1987
AR505_0799	*O*-acetylserine sulfhydrylase	*cysM*	40.2	Zhao *et al.*, 2006
AR505_0695, AR505_0800	Cysteine synthase subunit A	*cysK1*, *cysK2*	41.5, 53.6	Flint *et al.*, 1996

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
Branched chain amino acids				
AR505_1768	Acetolactate synthase small subunit	*ilvN*	22.1	Xing & Whitman, 1991
AR505_0152, AR505_0153	Acetolactate synthase large subunit	*ilvB1*, *ilvB2*	43.1, 42.7	Xing & Whitman, 1991
AR505_0150	2,3-dihydroxy-isovalerate:$NADP^+$ oxidoreductase	*ilvC*	46.5	Xing & Whitman, 1991
AR505_1452	Dihydroxy-isovalerate dehydratase	*ilvD*	56.2	Xing & Whitman, 1991
AR505_1767	Branched-chain amino acid aminotransferase	*ilvE*	37.2	Xing & Whitman, 1991
AR505_0631	2-isopropylmalate synthase	*leuA*	36.2	Vartak *et al.*, 1991
AR505_0531	Isopropylmalate/isocitrate dehydrogenases	*idh/leuB1*	23.4	Vartak *et al.*, 1991
AR505_0634	3-isoprolypmalate dehydrogenase	*leuB2*	29.7	Vartak *et al.*, 1991
AR505_0632	Isopropylmalate dehydratase large subunit	*leuC*	30.6	Vartak *et al.*, 1991
AR505_0633	Isopropylmalate dehydratase small subunit	*leuD*	28.1	Vartak *et al.*, 1991

Glutamate/glutamine. The ISO4-H5 genome is predicted to encode all the genes necessary for glutamate and glutamine biosynthesis using ammonium (Table 3.12.).

Arginine. Arginine biosynthesis via the acetyl cycle pathway is summarized in Figure 3.8 and Table 3.12. The majority of the genes involved in arginine biosynthesis form an operon consisting of *argHGCJBD* (AR505_0671 – 0676).

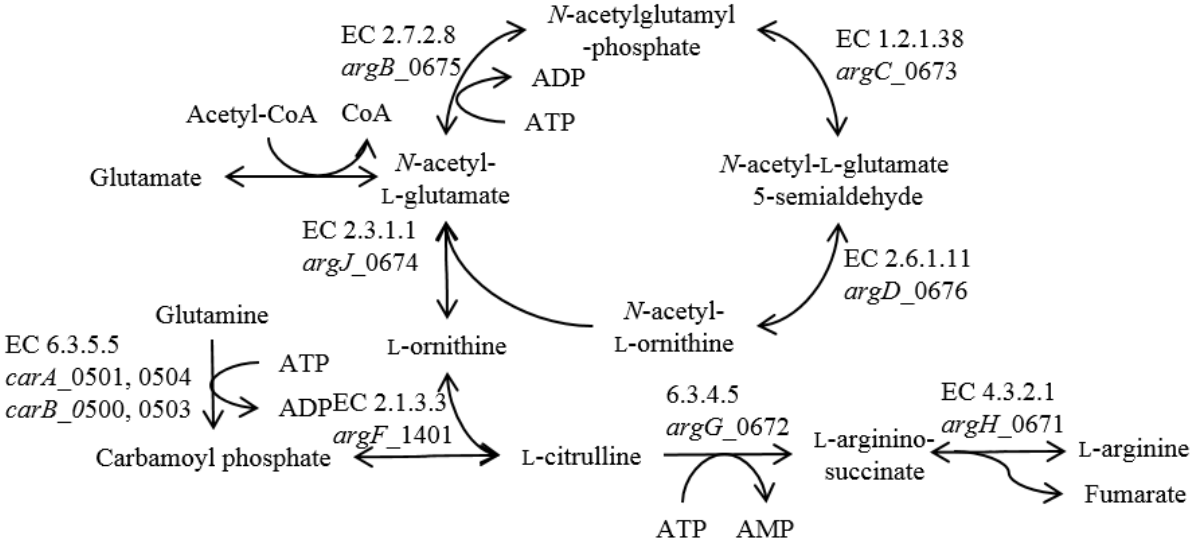


Figure 3.8. Biosynthesis of arginine in ISO4-H5. Enzymatic reactions are identified by E.C number, gene name followed by locus number. Bifunctional ornithine acetyltransferase/*N*-acetylglutamate synthase (*argJ*). Acetylglutamate kinase (*argB*). *N*-acetylglutamylphosphate reductase (*argC*). Acetylornithine aminotransferase (*argD*). Carbamoyl-phosphate synthetase (*carAB*). Ornithine carbamoyltransferase (*argF*). Argininosuccinate synthase (*argG*). Argininosuccinate lyase (*argH*).

Genes involved in arginine catabolism are present; arginine decarboxylase (AR505_0268), which converts L-arginine to agmatine, and agmatinase (AR505_1607), which breaks down agmatine to putrescine and urea. ISO4-H5 does not encode urea carboxylase gene, which is required in order to recycle urea.

Alanine/aspartate/asparagine. The biosynthesis of alanine involves the reversible transfer of an amino group from glutamate to pyruvate, producing alanine and 2-oxoglutarate (Table 3.12.). It is predicted to proceed via the alanine aminotransferase Alt (*alt*, AR505_1199), a gene specific to alanine, as well as via two broad substrate specificity aspartate aminotransferases, Ast (*aat*, AR505_1429 and AR505_1666). The broad substrate specificity aspartate amino transferases produce aspartate by transamination of oxaloacetate. ISO4-H5 encodes asparagine synthase B, AsnB (*asn*B1, B2; AR505_0736, 1303) which produces asparagine from aspartate.

Lysine. The ISO4-H5 genome is predicted to encode all the genes required for lysine biosynthesis via diaminopimelic acid (DAP) as an intermediate. This pathway branches to homoserine and pyrrolysine biosynthesis downstream.

Pyrrolysine. The ISO4-H5 genome is predicted to encode all the genes necessary for pyrrolysine biosynthesis and incorporation.

Methionine. Although ISO4-H5 is predicted not to make methionine, the genome encodes a partial methionine biosynthesis pathway, including homoserine dehydrogenase *metL* (AR505_0293), homoserine *O*-succinyltransferase *metA* (AR505_0694), and a possible *O*-acetylhomoserine sulfhydrylase (AR505_0738) gene (Table 3.12., Figure 3.9.). Expression of these genes would allow biosynthesis of *O*-succinyl-L-homoserine from L-aspartate semialdehyde, however, the ISO4-H5 genome lacks genes encoding the enzymes required to catalyse the remaining steps of methionine biosynthesis. The *metL* in ISO4-H5 only possesses the C-terminal domain of homoserine dehydrogenase (AR505_0293) and is predicted not to be bifunctional.

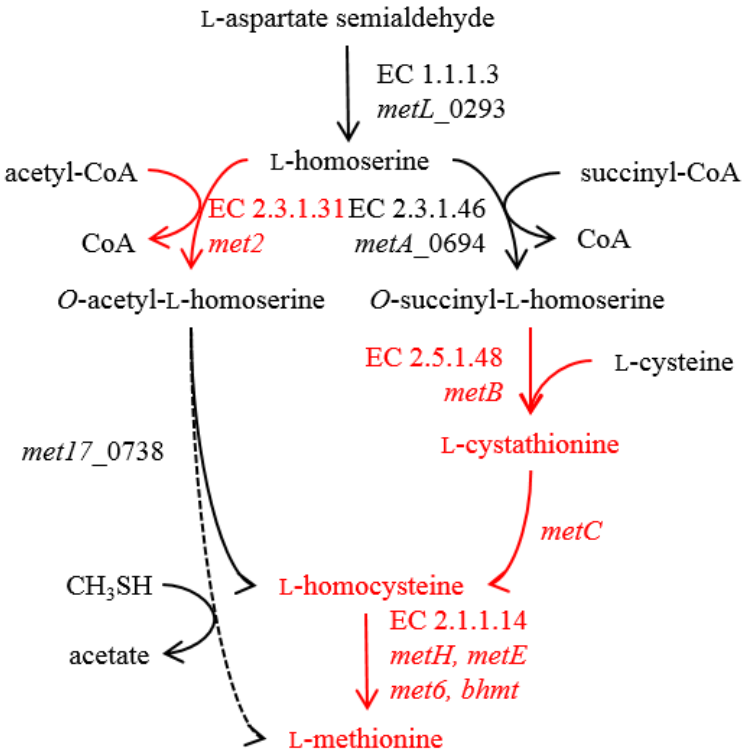


Figure 3.9. Incomplete biosynthesis pathway of methionine in ISO4-H5 Enzymatic reactions are identified by EC number if known, gene name followed by locus number. Pathways and enzymes absent in ISO4-H5 are represented in red text. The dotted line represents a proposed pathway. Homoserine dehydrogenase (*metL*). Homoserine *O*-succinyltransferase (*metA*). Homoserine *O*-acetyltransferase (*met2*). *O*-acetylhomoserine sulfhydrylase (*met17*). *O*-succinylhomoserine(thiol)-lyase (*metB*). Cystathionine β-lyase (*metC*). Cobalamin-independent methionine synthase (*metE*). Methionine synthase (*metH*). Homocysteine methyltransferase (*met6*). Betaine-homocysteine *S*-methyltransferase (*bhmt*).

Threonine/proline/histidine/serine. The ISO4-H5 genome is predicted to encode all the genes required for threonine, proline, histidine and serine biosynthesis (Table 3.12.).

Phenylalanine/tyrosine/tryptophan/cysteine. Homologues of the seven genes required for stepwise synthesis of chorismate (a common precursor in aromatic amino acid biosynthesis) are present in the ISO4-H5 genome (Table 3.12.). Based on the genes identified within the genome, ISO4-H5 is predicted to be capable of phenylalanine, tyrosine and tryptophan biosynthesis. The enzymes required for biosynthesis of tryptophan from chorismate are present (Table 3.12.). A pseudogene made up of two truncated homologues of *trpB* (AR505_1151), upstream of the tryptophan biosynthesis operon is also present.

Homologues of enzymes involved in biosynthesis of phenylalanine and tyrosine from chorismate are scattered throughout the ISO4-H5 genome (Table 3.12.). The *pheA* gene in ISO4-H5 carries the N-terminal prephenate dehydratase domain without a C-terminal

regulatory ACT domain, which is named after three proteins containing ACT domain, aspartate kinase, chorismate mutase and TyrA prephenate dehydrogenase.

Glycine. Although there are several different ways to produce glycine, (via 5,10-methylenetetrahydrofolate:glycine hydroxymethyltransferase, threonine aldolase, alanine-glyoxylate aminotransferase), no homologues for any of these enzymes were found within the ISO4-H5 genome.

Isoleucine/leucine/valine. The pathways of branched chain amino acid biosynthesis in ISO4-H5 appear complete and are similar to what has been found in other organisms, but with minor differences (Table 3.12.). ISO4-H5 is unlikely to utilise threonine for branched chain amino acid biosynthesis due to the lack of a homologue to threonine ammonia-lyase. However, the genome contains all the genes necessary for valine biosynthesis from pyruvate (Table 3.12.), with 3-methyl-2-oxobutanoate as an intermediate.

Vitamins and cofactors

F_{430}. The biosynthesis of tetrapyrrole cofactors such as F_{430}, siroheme, and cobalamin all share a common intermediate, uroporphyrinogen. Homologues of genes required for uroporphyrinogen biosynthesis from glutamate are present in the ISO4-H5 genome as a single operon (AR505_1040 – 1046) in addition to the glutamyl-tRNA synthetase (AR505_1121) (Table 3.13.). The homologue of the uroporphyrinogen-III C-methyltransferase gene *corA* (AR505_1041) is present in the same operon as the uroporphyrinogen biosynthesis genes in the ISO4-H5 genome. CorA converts uroporphyrinogen-III to precorrin-1, which in turn is converted to precorrin-2 using *S*-adenyl methionine (SAM) as a cofactor.

Table 3.13. Genes involved in cofactor biosynthesis

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
Tetrapyrrole common pathway				
AR505_1040	Uroporphyrinogen III synthase	*hemD*	18.1	Sasarman *et al.*, 1987
AR505_1042	Porphobilinogen deaminase	*hemC*	36.7	Jordan *et al.*, 1988
AR505_1043	Glutamate-1-semialdehyde-2,1-aminomutase	*hemL*	44.1	Ilag *et al.*, 1991
AR505_1044	Delta-aminolevulinic acid dehydratase	*hemB*	54.0	Bhosale *et al.*, 1995
AR505_1045	Glutamyl-tRNA reductase	*hemA*	25.8	Schauer *et al.*, 2002
AR505_1121	glutamyl-tRNA synthetase	*gltX*	21.8	Breton *et al.*, 1986
AR505_1046	Precorrin-2 dehydrogenase	*pc2_dh*	28.8	Lobo *et al.*, 2009
F_{430}				
AR505_1041	Uroporphyrin-III C-methyltransferase	*corA*	43.3	Blanche *et al.*, 1991
Riboflavin, FMN, FAD				
AR505_0210	GTP cyclohydrolase	*arfA*	14.1	Morrison *et al.*, 2008
AR505_1180	3,4-dihydroxy-2-butanone-4-phosphate synthase	*ribB*	46.8	Fischer *et al.*, 2002
AR505_1181	FAD synthetase	*ribL*	37.6	Mashhadi *et al.*, 2010
AR505_1182	Riboflavin synthase	*ribC*	68.4	Fischer *et al.*, 2004
AR505_1183	6,7-dimethyl-8-ribityllumazine synthase	*ribH*	55.8	Haase *et al.*, 2003
AR505_1679	Riboflavin kinase	*ribK*	38.8	Mashhadi *et al.*, 2008
$NAD^+/NADP^+$				
AR505_0143	L-Aspartate dehydrogenase	*aspDH*	28.5	Yang *et al.*, 2003
AR505_1155	Quinolinate synthase	*nadA*	33.3	Ceciliani *et al.*, 2000
AR505_0471	Quinolinate phosphoribosyltransferase	*nadC*	16.8	Bhatia & Calvo, 1996
AR505_0377	NH_3-dependent NAD^+ synthase	*nadE*	20.9	Ozment *et al.*, 1999
AR505_0616	Nicotinamide-mononucleotide adenylyltransferase	*nadM*	33.3	Raffaelli *et al.*, 1997
AR505_0121	NAD kinase	*nadK*	26.9	Magni *et al.*, 2006
AR505_1654	Nicotinate phosphoribosyltransferase	*pncB*	12.0	Wubbolts *et al.*, 1990

Cobalamin/thiamine. The sirohydrochlorin intermediate produced from F_{430} biosynthesis can be used for cobalamin (vitamin B12) biosynthesis, (Appendix, Table A.3.1.). The ISO4-H5 genome is predicted to encode all but 3 genes required for the anaerobic pathway of cobalamin biosynthesis, the missing genes are *cbiE*, *cbiT* and cobirinic acid-a,c-diamide reductase.

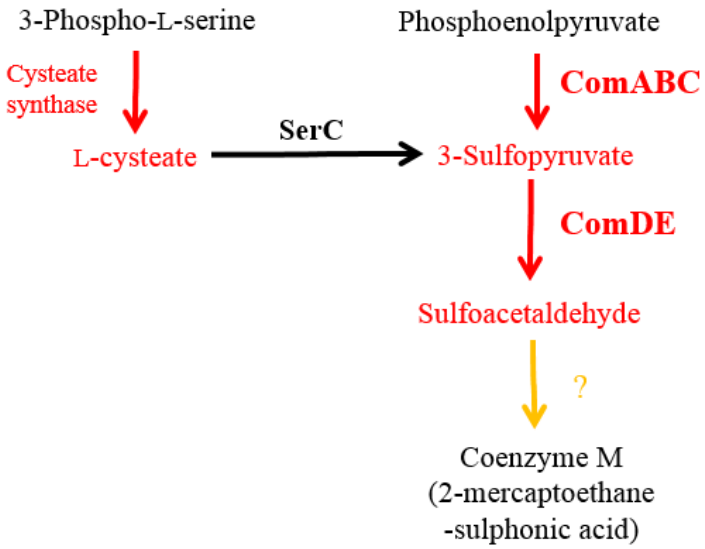


Figure 3.10. Absence of a complete CoM biosynthesis pathway in ISO4-H5. Red arrow or names: predicted to be absent in ISO4-H5. Black arrows or names: predicted to be present in ISO4-H5. Yellow arrow or names: unknown enzyme and pathway. Broad specificity phosphoserine aminotransferase (SerC), (*R*)-phosphosulfolactate synthase (ComA), phosphosulfolactate phosphatase (ComB), sulfolactate dehydrogenase (ComC), sulfopyruvate decarboxylase (ComDE).

Riboflavin/FMN/FAD. Vitamin B2, also known as riboflavin, is a precursor to flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The ISO4-H5 genome is predicted to be missing two genes necessary for the biosynthesis of riboflavin: 2-amino-5-formylamino-6-ribosylaminopyrimidin-4(3H)-one 5'-monophosphate deformylase, ArfB, and pyrimidine nucleotide reductase, ArfC, which produces the 5-amino-6-(D-ribitylamino) uracil intermediate (Table 3.13.).

Coenzyme M. Coenzyme M is an essential cofactor required for methanogenesis, its biosynthesis from phosphoenolpyruvate requires the (*R*)-phosphosulfolactate synthase gene *comA*, phosphosulfolactate phosphatase gene *comB*, sulfolactate dehydrogenase gene *comC*, sulfopyruvate decarboxylase genes *comDE* and an unidentified enzyme (Figure 3.10.). However, homologues to any of the known CoM biosynthesis genes *comABCDE* are absent from the ISO4-H5 genome.

NAD^+/$NADP^+$. The biosynthesis of nicotinamide adenine dinucleotide (NAD^+) and nicotinamide adenine dinucleotide phosphate ($NADP^+$) are important cofactors for a large number of enzymes. NAD^+ production in ISO4-H5 is predicted to proceed via two different pathways; from aspartate or from nicotinamide/nicotinic acid (Table 3.13.). The ISO4-H5 genome is predicted to encode all but two genes required for the first biosynthesis pathway; aspartate oxidase NadB, which converts aspartate to iminoaspartate, and nicotinate-nucleotide adenylyltransferase NadD, that catalyses the formation of NAD. The proposed NAD biosynthesis pathway via nicotinamide involves NadM, which condenses nicotinamide mononucleotide with ATP to form NAD^+ directly without the requirement of NadE. The nicotinamide mononucleotide is predicted to be produced by a homologue of the nicotinate phosphoribosyltransferase gene (AR505_1654) from nicotinamide and 5-phospho-D-ribose-1-diphosphate. A homologue of NAD kinase (AR505_0121) converts NAD^+ to $NADP^+$ is also found withint the ISO4-H5 genome.

Biotin. ISO4-H5 lacks the genes required to synthesise biotin.

Lipids and cell envelope

In ISO4-H5, glycerol-1-phosphate (G1P) is predicted to be produced from dihydroxyacetone phosphate (DHAP) by EgsA, where the dihydroxyacetone phosphate is derived from the gluconeogenesis pathway by the breakdown of fructose 1,6 phosphate (Figure 3.11.). There is one missing step in the conversion of mevalonate-5-phosphate to isopentenyl phosphate due to the lack of phosphomevalonate decarboxylase in the ISO4-H5 genome.

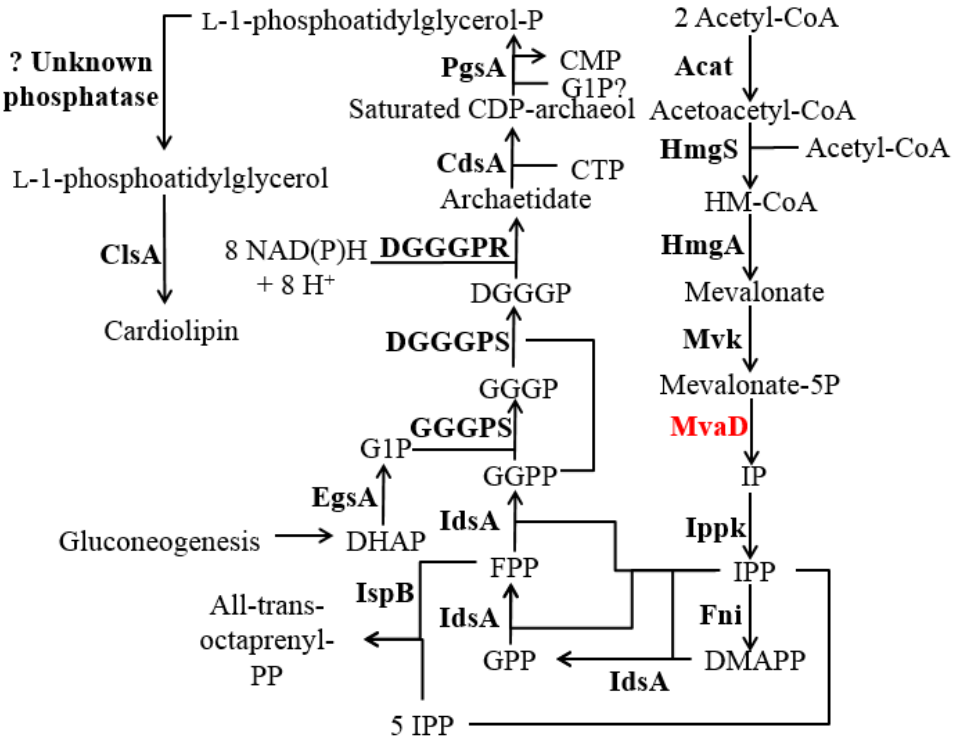


Figure 3.11. Phospholipid biosynthesis in ISO4-H5. Enzyme names are displayed in bold. Red enzyme names indicate absence in the ISO4-H5 genome. Acetyl-CoA acetyltransferase (Acat), hydroxymethylglutaryl-CoA synthase (HmgS), hydroxymethylglutaryl-CoA reductase (HmgA), hydroxymethylglutaryl-CoA (HM-CoA), mevalonate kinase (Mvk), phosphomevalonate decarboxylase (MvaD), isopentenyl phosphate (IP), isopentenyl diphosphate (IPP), isopentenyl-diphosphate δ-isomerase (Fni), 3,3-dimethylallyldiphosphate (DMAPP), bifunctional short chain isoprenyl diphosphate synthase (IdsA), geranyl diphosphate (GPP), farnesyl diphosphate (FPP), octaprenyl-diphosphate synthase (IspB), geranylgeranyl diphosphate (GGPP), dihydroxyacetone phosphate (DHAP), NAD(P)-dependent glycerol-1-phosphate dehydrogenase (EgsA), glycerol-1-phosphate (G1P), geranylgeranylglyceryl phosphate synthase (GGGPS), geranylgeranylglyceryl phosphate (GGGP), digeranylgeranylglyceryl phosphate synthase (DGGGPS), digeranylgeranylglyceryl phosphate (DGGGP), digeranylgeranylglycerophospholipid reductase (DGGGPR), CDP-diglyceride synthetase (CdsA), cytidine triphosphate (CTP), cytidine monophosphate (CMP), CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (PgsA), cardiolipin synthase (ClsA).

The proposed *de novo* phospholipid biosynthesis pathway in ISO4-H5 is displayed in Figure 3.11. Based on the presence of the digeranylgeranylglycerylphospholipid reductase (DGGGPR) gene and a CDP-diglyceride synthetase (CdsA) gene, ISO4-H5 is predicted to use archaetidate and saturated CDP-archaeol as intermediates in phospholipid biosynthesis. The

genes in ISO4-H5 suggest that CDP-archaeol proceeds towards cardiolipin biosynthesis, with cardiolipin likely constituting part of the ISO4-H5 cell envelope.

In addition to cardiolipin, ISO4-H5 possesses three geranylgeranyl reductase family proteins, two phosphoglucosamine mutase genes, a cell wall biosynthesis protein Mur ligase family gene, and a NAD-dependent epimerase/dehydratase gene (Table 3.14.).

The ISO4-H5 genome does not encode UDP-*N*-acetylglucosamine diphosphorylase, glucosamine-1-phosphate *N*-acetyltransferase and glucosamine-fructose-6-phosphate aminotransferase involved in pseudomureine production. There is a cell wall glycopolymer cluster predicted to be involved in cell envelope biosynthesis (AR505_0539 - 0561), including 3 genes involved in dTDP-L-rhamnose biosynthesis (AR505_0552 – 0554). In addition, four genes have been predicted to produce glycerol-manno-heptose (AR505_1754, AR505_1770 – 1772).

Table 3.14. Genes involved in lipid biosynthesis

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimentally validated homologue
G1P biosynthesis				
AR505_0626	NAD(P)-dependent glycerol-1-phosphate dehydrogenase	*egsA*	47.4	Nishihara & Koga, 1997
Isoprenoid biosynthesis				
AR505_0004, AR505_0602	Acetyl-CoA acetyltransferase	*acat1*, *acat2*	40.4, 19.9	Middleton, 1974
AR505_0601	HMG-CoA synthase	*hmgS*	23.1	Wilding *et al.*, 2000
AR505_0768	HMG-CoA reductase	*hmgA*	47.3	Bochar *et al.*, 1997
AR505_1431	Mevalonate kinase	*mvk*	32.8	Primak *et al.*, 2011
AR505_0192	Isopentenyl phosphate kinase	*ippk*	31.4	Chen & Poulter, 2010
AR505_0191	Isopentenyl-diphosphate δ-isomerase	*fni*	36.4	Yamashita *et al.*, 2004
AR505_1619	Octaprenyl-diphosphate synthase	*ispB*	31.3	Asai *et al.*, 1994
AR505_0190	Bifunctional short chain isoprenyl diphosphate synthase IdsA/GGPS	*idsA*	39.0	Chen & Poulter, 1994
Phospholipid biosynthesis				
AR505_1588	Geranylgeranylglyceryl phosphate synthase GGGPS	*gggps*	43.4	Hemmi *et al.*, 2004
AR505_1587	Digeranylgeranylglyceryl phosphate synthase	*dgggps*	30.8	Zhang *et al.*, 1990
AR505_1433	Digeranylgeranylglycerophospholipid reductase	*dgggpr*	45.4	Nishimura & Eguchi, 2006
AR505_1616	CDP-diglyceride synthetase	*cdsA*	26.7	Sparrow & Raetz, 1985
AR505_1783	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	*pgsA*	15.9	Hirabayashi *et al.*, 1976
AR505_0933	Cardiolipin synthase	*clsA*	27.7	Ohta *et al.*, 1985
Cell surface glycopolymer				
AR505_1754	D-sedoheptuloase 7-phosphate isomerase	*gmhA*	38.8	Kneidinger *et al.*, 2001
AR505_1771	D-glycero-D-manno-heptose-7-phosphate kinase	*hddA*	42.4	Kneidinger *et al.*, 2001
AR505_1770	D-glycero-D-manno-heptose 1,7-bisphosphate 7-phosphatase	*gmhB*	30.8	Kneidinger *et al.*, 2001
AR505_1772	D-glycero-D-manno-heptose 1-phosphate guanylyltransferase	*hddC*	30.8	Kneidinger *et al.*, 2001
Others				
AR505_0123, AR505_1608	Phosphoglucosamine mutase	*glmM1*, *glmM2*	17.7, 30.7	Mengin-Lecreulx & van Heijenoort, 1996
AR505_0892, AR505_1618, AR505_1646	Geranylgeranyl reductase family protein			
AR505_0361	Cell wall biosynthesis protein Mur ligase family			
AR505_0541	NAD-dependent epimerase/dehydratase			

Nucleic acid biosynthesis

The biosynthesis of nucleotides requires phosphoribosyl pyrophosphate (PRPP) produced from central carbon metabolism and a formyl group. Other than the two genes involved in dGTP biosynthesis, the full complement of genes required for purine and pyrimidine biosynthesis have been predicted in the ISO4-H5 genome (Figure 3.12., Tables 3.15., 3.16.). The IMP dehydrogenase GuaB that produces xanthosine monophosphate (XMP) and guanylate kinase Gmk that phosphorylates GMP to GDP is predicted to be absent.

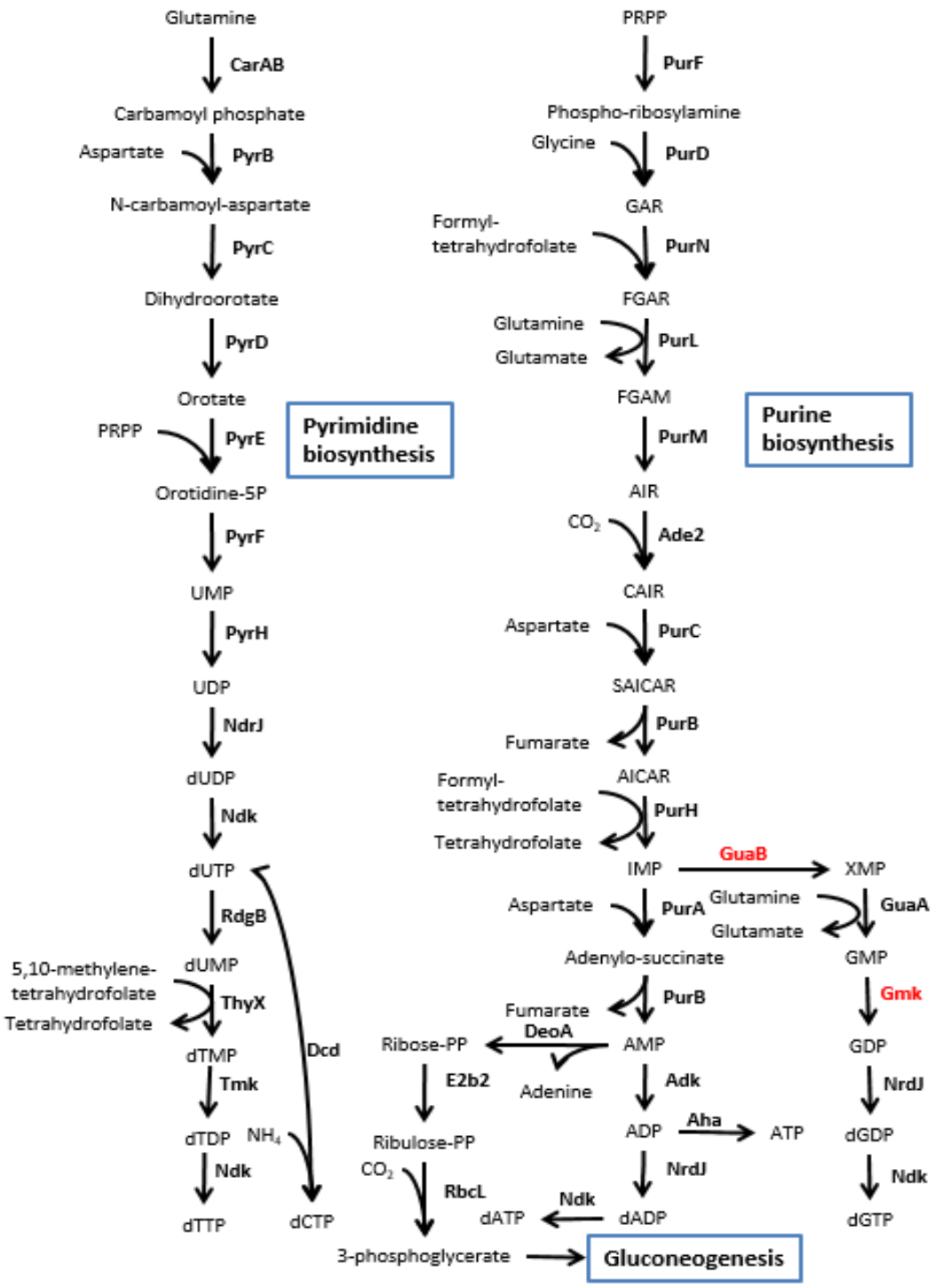

Glutamine
CarAB
Carbamoyl phosphate
Aspartate
PyrB
N-carbamoyl-aspartate
PyrC
Dihydroorotate
PyrD
Orotate
PRPP
PyrE
Orotidine-5P
PyrF
UMP
PyrH
UDP
NdrJ
dUDP
Ndk
dUTP
RdgB
dUMP
5,10-methylene-tetrahydrofolate
ThyX
Tetrahydrofolate
dTMP
Tmk
dTDP
NH4
Ndk
dTTP
Dcd
dCTP
Pyrimidine biosynthesis
PRPP
PurF
Phospho-ribosylamine
Glycine
PurD
GAR
Formyl-tetrahydrofolate
PurN
FGAR
Glutamine
Glutamate
PurL
FGAM
PurM
AIR
CO2
Ade2
CAIR
Aspartate
PurC
SAICAR
Fumarate
PurB
AICAR
Formyl-tetrahydrofolate
PurH
Tetrahydrofolate
IMP
GuaB
XMP
Aspartate
PurA
Glutamine
Glutamate
GuaA
Adenylo-succinate
GMP
Fumarate
PurB
Gmk
Ribose-PP
DeoA
AMP
GDP
E2b2
Adenine
Adk
NrdJ
ADP
Aha
ATP
dGDP
Ribulose-PP
CO2
NrdJ
RbcL
Ndk
Ndk
dATP
dADP
dGTP
3-phosphoglycerate
Gluconeogenesis
Purine biosynthesis

Figure 3.12. *De novo* nucleotide biosynthesis in ISO4-H5. Enzyme names are displayed in bold. Red enzyme names indicate absent in ISO4-H5 genome. Carbamoyl phosphate synthase (CarAB), aspartate transcarbamylase (PyrB), dihydroorotase (PyrC), dihydroorotate dehydrogenase (PyrD), phosphoribose pyrophosphate (PRPP), orotate phosphoribosyltransferase (PyrE), deoxy (d), [adenine (A); cytidine (C); uridine (U); thymidine (T); guanine (G); inosine (I); xanthosine (X)] (Y), [monophosphate (MP); diphosphate (DP); triphosphate (TP)] (ZP) (dYZP), UMP kinase (PyrH), broad substrate specificity nucleoside diphosphate kinase (Ndk), ribonucleoside-diphosphate reductase (NrdJ), dITP/XTP pyrophosphoatase (RdgB), flavin-dependent thymidylate synthase (ThyX), thymidylate kinase (Tmk), CMP/dCMP deaminase (Dcd), amidophosphoribosyl transferase (PurF), phosphoribosylamineglycine ligase (PurD), (5-phospho-β-D-ribosyl)glycinamide (GAR), phosphoribosylglycinamide formyltransferase (PurN), formyl-(5-phospho-β-D-ribosyl)glycinamide (FGAR), phosphoribosylformylglycinamidine synthetase (PurL), formamido-5-phospho-β-D-ribosyl)acetamidine (FGAM), phosphoribosylformylglycinamide cyclo-ligase (PurM), amino-(5-phospho-ribosyl)imidazole (AIR), phosphoribosylaminoimidazole carboxylase (Ade2), amino-(5-phospho-ribosyl)imidazole-4-carboxylate (CAIR), phosphoribosylaminoimidazole-succinocarboxamide synthase (PurC), 5-phosphoribosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole (SAICAR), bifunctional enzyme adenosuccinate lyase (PurB), amino-1-(5-phosphoribosyl)imidazole-4-carboxamide (AICAR), bifunctional AICAR transformylase/IMP cyclohydrolase (PurH), adenylosuccinate synthetase (PurA), adenylate kinase (Adk), A_1A_0 ATP synthase (Aha), IMP dehydrogenase (GuaB), GMP synthase (GuaA), guanylate kinase (Gmk), AMP phosphorylase (DeoA), ribulose bisphosphate carboxylase (RbcL), ribose-1,5-bisphosphate isomerase (E2b2).

Table 3.15. Genes involved in pyrimidine biosynthesis and interconversion

Locus tag	Predicted product	Gene	Amino acid identity (%)	Experimental validated homologue
Pyrimidine biosynthesis				
AR505_0501, AR505_0504	Carbamoyl phosphate synthase subunit A	*carA1, carA2*	38.5, 33.4,	Thoden *et al*., 1999
AR505_0500, AR505_0503	Carbamoyl phosphate synthase subunit B	*carB1, carB2*	48.6, 48.1	Thoden *et al*., 1999
AR505_0437	Aspartate transcarbamylase PyrB	*pyrB*	48.4	Wang *et al.*, 2005
AR505_0436	Aspartate transcarbamylase regulatory subunit	*pyrI*	31.9	Wang *et al.*, 2005
AR505_0281	Dihydroorotase	*pyrC*	18.7	Thoden *et al*., 2001
AR505_1029	Dihydroorotate dehydrogenase	*pyrD*	21.5	Norager *et al.*, 2002
AR505_1030	Dihydroorotate dehydrogenase electron transfer subunit	*pyrK*	28.3	Nielsen *et al*., 1996
AR505_1615	Orotate phosphoribosyltransferase	*pyrE*	26.8	Aghajari *et al.*, 1994
AR505_0060	Orotidine-5'-phosphate decarboxylase	*pyrF*	24.9	Lee & Houk, 1997
AR505_0251	UMP kinase	*pyrH*	22.1	Serina *et al.*, 1995
AR505_0187	dITP/XTP pyrophosphatase	*rdgB*	37.6	Chung *et al.*, 2002
AR505_1756	Broad substrate specificity nucleoside diphosphate kinase	*ndk*	45.5	Bernard *et al*., 2000
AR505_0964	Ribonucleoside-diphosphate reductase	*nrdJ*	25.8	Kolberg *et al.*, 2004; Torrents *et al.*, 2000
AR505_0076	Flavin-dependent thymidylate synthase	*thyX*	26.6	Myllykallio *et al*., 2002
Interconversion				
AR505_1102, AR505_1690	Thymidylate kinase	*tmk1, tmk2*	19.0, 31.8	Daws & Fuchs, 1984
AR505_1827	CMP/dCMP deaminase	*dcd*	37.3	McIntosh & Haynes, 1986

ISO4-H5 contains an operon for recycling of AMP, consisting of the genes *deoA, rbcL*, and *e2b2* (Table 3.16.). The presence of the carbonic anhydrase Cab homologue in the ISO4-H5 genome (AR505_1245) adds weight to the functionality of this *RubisCO* operon, as Cab mediates the interconversion of H_2 carbonate and CO_2, which usually feeds towards oxaloacetate production.

Table 3.16. Genes involved in purine biosynthesis and interconversion

Locus tag	Predicted gene product	Gene	Amino acid identity (%)	Experimental validated homologue
IMP biosynthesis				
AR505_0284	Amidophosphoribosyl transferase	*purF*	37.6	Messenger & Zalkin, 1979
AR505_0066	Phosphoribosylamineglycine ligase	*purD*	31.7	Zhang *et al.*, 2002
AR505_0141	Phosphoribosylglycinamide formyltransferase	*purN*	17.7	Marolewski *et al.*, 1994
AR505_1680	Phosphoribosylformylglycinamidine synthetase subunit Q	*purQ*	37.1	Hoskins *et al.*, 2004
AR505_1681	Phosphoribosylformylglycinamidine synthetase subunit L	*purL*	36.5	Hoskins *et al.*, 2004
AR505_1682	Phosphoribosylformylglycinamidine synthetase subunit S	*purS*	29.1	Hoskins *et al.*, 2004
AR505_0037	Phosphoribosylformylglycinamide cyclo-ligase	*purM*	32.8	Smith & Daum, 1986
AR505_1804	Phosphoribosylaminoimidazole carboxylase	*ade2*	40.0	Stotz & Linder, 1990
AR505_1595	PurE-related protein		10.9	Meyer *et al.*, 1992
AR505_0622	Phosphoribosylaminoimidazole-succinocarboxamide synthase	*purC*	19.0	Meyer *et al.*, 1992
AR505_0742	Bifunctional enzyme adenosuccinate lyase	*purB*	25.6	Green *et al.*, 1996
AR505_1656	Bifunctional AICAR transformylase/IMP cyclohydrolase	*purH*	50.7	Flannigan *et al.*, 1990
Adenosine deoxyribonucleotide biosynthesis				
AR505_1168	Adenylosuccinate synthetase	*purA*	44.1	Honzatko & Fromm, 1999
AR505_0262, AR505_1784	Adenylate kinase	*adk1, adk2*	44.5, 10.5	Huss & Glaser, 1983
AMP recycling				
AR505_1641	AMP phosphorylase	*deoA*	47.0	Sato *et al.*, 2007; Watson *et al.*, 1999
AR505_1642	Ribulose bisphosphate carboxylase RubisCO	*rbcL*	34.5	Sato *et al.*, 2007; Watson *et al.*, 1999
AR505_1643	Ribose-1,5-bisphosphate isomerase	*e2b2*	42.7	Sato *et al.*, 2007; Watson *et al.*, 1999
AR505_1245	Carbonic anhydrase	*cab*	23.8	Smith & Ferry, 1999
Guanosine deoxyribonucleotide biosynthesis				
AR505_0706, AR505_1802, AR505_1803	GMP synthase	*guaA1*, *guaAa*, *guaAb*	44.8, 11.7, 40.1	Tesmer *et al.*, 1994

Cell cycle

ISO4-H5 encodes two origin of replication *orc1/cdc6* genes (AR505_0001, 1205) and an OB-fold nucleic acid binding domain protein (AR505_1668) which might act as the single-strand binding protein (SSB) that protects single stranded DNA during replication and recombination. DNA polymerase genes found in the genome include DNA polymerase family D subunits *polD1,2* (AR505_1438, 1816), family B *polB* (AR505_0130), and family C *polC* (AR505_0374). The polymerase is regulated by the sliding clamp, proliferating cell nuclear antigen PcnA (*pcnA*, AR505_1650) and replication factor C RfcLS genes (*rfcL,S*, AR505_0828, 1209). Genes required for joining of Okazaki fragments are also present, including flap endonuclease *fen* (AR505_0460), ATP-dependent DNA ligase *dnlI* (AR505_1671) and ribonuclease HII *rnhB* (AR505_0209). Ribonuclease H (*rnhH*, AR505_0420) degrades the RNA-DNA hybrid. The tyrosine recombinase XerCD is responsible for termination of replication in bacteria, however no homologous genes are predicted to be in the ISO4-H5 genome.

The genome segregates in the D phase, and in bacteria the ParAB protein functions to bind the chromosomes and separate them. The ISO4-H5 genome does not harbour ParAB homologs, instead ISO4-H5 relies on chromosome segregation and condensation protein ScpBA (*scpB,A,* AR505_1609, 1610) and chromosome segregation protein SMC (*smc*, AR505_1611). There are two additional SMC domain containing proteins (AR505_0890, 0891) that could also contribute to genome segregation. DNA gyase GyrAB (*gyrA,B,* AR505_0033, 0034) and DNA topoisomerase TopAB (*topA,B*, AR505_0388, 0389, 1299) required for genome segregation has also been found in the genome.

The ISO4-H5 genome is predicted to encode several genes involved in cell division, including two cell division protein genes, *ftsZ* (AR505_0975, 1792) and the cell division ATPase gene *minD* (AR505_1412). Cell division is regulated by cell division control proteins Cdc48 (*cdc48-1*, *-2*, AR505_0052, AR505_1810), and two additional genes which might also be related to this function; the AAA family ATPase of Cdc48 subfamily (AR505_0445) and the Cdc-48 related protein (AR505_0854).

Nucleic acid metabolism

The ISO4-H5 genome is predicted to encode an archaeal histone (AR505_1236) and an associated histone acetyltransferase (AR505_0051) that modifies histones post-translationally. In addition, there are also five Gcn5-related *N*-acetyltransferase (GNAT) family acetyltransferases and eight SAM-dependent methyltransferase in the ISO4-H5 genome that can possibly modify the histones. The genes predicted to be involved in DNA repair in the ISO4-H5 genome is listed in Table 3.17.

Table 3.17. DNA recombination and repair

Locus_tag	Gene product
AR505_0027	8-oxoguanine DNA-glycosylase Ogg
AR505_0028	endonuclease IV
AR505_0117	ssDNA exonuclease RecJ
AR505_0125	replication-associated recombination protein A
AR505_0178	endonuclease IV
AR505_0189	DNA repair and recombination protein RadB
AR505_0329	DNA mismatch endonuclease Vsr
AR505_0391	formamidopyrimidine-DNA glycosylase MutM
AR505_0448	6-*O*-methylguanine DNA methyltransferase Ogt
AR505_0693	endoribonuclease L-PSP
AR505_0798	ssDNA exonuclease RecJ
AR505_1055	site-specific recombinase
AR505_1056	site-specific recombinase
AR505_1069	DNA repair photolyase
AR505_1084	exodeoxyribonuclease VII small subunit XseB
AR505_1085	exodeoxyribonuclease VII large subunit XseA
AR505_1139	exonuclease
AR505_1154	archaeal Holliday junction resolvase Hjc
AR505_1278	hef nuclease
AR505_1305	endonuclease III Nth
AR505_1309	DNA alkylation repair enzyme
AR505_1402	ssDNA exonuclease RecJ
AR505_1578	excinuclease ABC A subunit UvrA
AR505_1585	DNA repair and recombination protein RadA
AR505_1614	uracil-DNA glycosylase Ung
AR505_1729	resolvase domain-containing protein

Protein synthesis

Transcription. The transcription factors required to initiate and terminate transcription, including the TATA-box binding protein Tbp (*tbp*, AR505_0740), transcription initiation factor TFIIB (*tfb1*, *2*, AR505_0810, AR505_1807), transcription factor S, Tfs (*tfs*, AR505_1651) and NusA-like protein (AR505_0058). The ISO4-H5 genome encodes 11 subunits of the DNA-directed RNA polymerase (Table 3.18.). In addition, there are four RNA-binding proteins (AR505_0137, 1411, 1425, 1617) that might be involved in transcription regulation.

Table 3.18. ISO4-H5 RNA polymerase genes

Locus_tag	Gene product
AR505_0053	DNA-directed RNA polymerase subunit H RpoH
AR505_0054	DNA-directed RNA polymerase subunit B RpoB
AR505_0055	DNA-directed RNA polymerase subunit A' RpoA1
AR505_0056	DNA-directed RNA polymerase subunit A'' RpoA2
AR505_0134	DNA-directed RNA polymerase subunit F RpoF
AR505_0182	DNA-directed RNA polymerase subunit N RpoN
AR505_0922	DNA-directed RNA polymerase subunit K RpoK
AR505_1373	DNA-directed RNA polymerase subunit E'' RpoE2
AR505_1374	DNA-directed RNA polymerase subunit E RpoE
AR505_1439	DNA-directed RNA polymerase subunit D RpoD
AR505_1488	DNA-directed RNA polymerase subunit P RpoP

Maturation of tRNA. The tRNA precursor molecules that are transcribed contain 5' and 3' extensions which need to be removed to produce functional tRNAs. The ISO4-H5 genome encodes genes requiring this function, including RNAse P subunits P29 (rpp29, AR505_0227)

and RPR2 (*rpr2*, AR505_1424), ribonuclease Z (*rnz*, AR505_0464), tRNA nucleotidyltransferase Cca (AR505_1583), tRNA intron endonuclease EndA (*endA*, AR505_1808), 5' RNA ligase (AR505_1021) and RNA-splicing ligase RtcB (*rtcB*, AR505_1406). The aminoacyl-tRNA synthetases are responsible for linking amino acids to corresponding tRNAs for amino acids and these are listed in Table 3.19.

Table 3.19. ISO4-H5 genes involved in aminoacyl-tRNA biosynthesis

Locus_tag	Gene product
Aminoacyl-tRNA synthetases	
AR505_0198	Alanyl-tRNA synthetase AlaS1
AR505_0200	Aspartate-tRNA synthetase AspS1
AR505_0247	Tryptophanyl-tRNA synthetase TrpS
AR505_0248	Phenylalanyl-tRNA synthetase PheS
AR505_0253	Arginine-tRNA synthetase ArgS
AR505_0387	Lysyl-tRNA synthetase LysS
AR505_0394	Histidyl-tRNA synthetase HisS
AR505_0397	Threonyl-tRNA synthetase ThrS
AR505_0455	Isoleucyl-tRNA synthetase IleS
AR505_0709	Methionyl-tRNA synthetase MetG1
AR505_0710	Methionyl-tRNA synthetase MetG2
AR505_0767	Phenylalanyl-tRNA synthetase PheT
AR505_0972	Leucyl-tRNA synthetase LeuS
AR505_1121	Glutamyl-tRNA synthetase GltX
AR505_1150	Aspartate-tRNA synthetase AspS2
AR505_1191	Cysteinyl-tRNA synthetase CysS
AR505_1325	Pyrrolysine-tRNA synthetase PylS
AR505_1339	Valyl-tRNA synthetase ValS
AR505_1341	Glycyl-tRNA synthetase GlyS
AR505_1426	Seryl-tRNA synthetase SerS
AR505_1448	Alanyl-tRNA synthetase AlaS2
AR505_1494	Tyrosyl-tRNA synthetase TyrS
AR505_1684	Prolyl-tRNA synthetase ProS
ATP dependent transamination	
AR505_0068	Glutamyl-tRNA(gln) amidotransferase GatD
AR505_0069	Glutamyl-tRNA(gln) amidotransferase GatE
AR505_0201	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatB
AR505_0202	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatA
AR505_0203	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatC

Maturation of rRNA. The endonuclease and ribonuclease involved in tRNA maturation are also involved in rRNA maturation. In addition, the ISO4-H5 genome encodes a ribosome maturation protein SBDS (AR505_1483) and four genes encoding exosomes subunits (AR505_1358), RNA-binding protein Rrp4 (*rrp4*, AR505_1484), exonuclease Rrp41 (*rrp41*, AR505_1485) and RNA-binding protein Rrp42 (*rrp42*, AR505_1486).

RNA modifications. Genes involved in RNA post-transcriptional modifications, including tRNA modification to wyosine, queuosine and archaeosine are shown in Table 3.20. There are no functional small nuclear RNA (snRNA) nor RNA components of C/D box methylation guide ribonucleoprotein complex identified within the ISO4-H5 genome, although an aNOP56 subunit (AR505_1605) and fibrillarin (AR505_0120) are found.

Table 3.20. ISO4-H5 genes involved in tRNA modification

Locus_tag	Gene product
Wyosine, queuosine and archaeosine	
AR505_0438	Archaeosin tRNA-ribosyltransferase TgtA1
AR505_0637	Archaeosin tRNA-ribosyltransferase TgtA2
AR505_0827	Wyosine biosynthesis protein TYW1
AR505_1414	7-cyano-7-deazaguanosine biosynthesis protein QueE
AR505_1415	Queuosine biosynthesis protein QueC
AR505_1416	6-pyruvoyl tetrahydropterin synthase QueD
Others	
AR505_0038	Ribosomal RNA large subunit metyltransferase J RrmJ
AR505_0044	tRNA 2'*O*-methylase (predicted pseudogene)
AR505_0065	tRNA pseudouridine synthase A TruA
AR505_0132	Dimethyladenosine transferase KsgA
AR505_0243	tRNA pseudouridine synthase B TruB
AR505_0307	SAM-dependent methyltransferase
AR505_0536	SAM-dependent methyltransferase
AR505_0809	tRNA-dihydrouridine synthase DusA
AR505_0973	SAM-dependent methyltransferase
AR505_0981	SAM-dependent methyltransferase
AR505_1008	SAM-dependent methyltransferase
AR505_1146	Ribosomal-protein-alanine acetyltransferase RimI
AR505_1196	SAM-dependent methyltransferase
AR505_1203	SAM-dependent methyltransferase
AR505_1233	Pseudouridylate synthase
AR505_1275	tRNA pseudouridine synthase D TruD
AR505_1285	SAM-dependent methyltransferase
AR505_1460	MiaB-like tRNA modifying enzyme
AR505_1586	N_2, N_2-dimethylguanosine tRNA methyltransferase Trm
AR505_1658	Ribosomal protein L11 methyltransferase PrmA
AR505_1683	tRNA methyltransferase subunit
AR505_1791	RNA methyltransferase TrmH family

Ribosomal proteins. The gene encoding proteins in the large (50*S*) and small (30*S*) subunits of the ribosome are listed in Table A.3.1.

Translation factors. Translation factors are essential in facilitation of translational initiation, peptide elongation and termination; those found in ISO4-H5 are listed in Table 3.21.

Table 3.21. Translation factors in the ISO4-H5 genome

Locus_tag	Gene product
Translation initiation factors	
AR505_0139	Translation initiation factor aIF-1A
AR505_0226	Translation initiation factor aSUI1
AR505_0782	Translation initiation factor aIF-6
AR505_1459	Translation initiation factor aIF-2 alpha subunit
AR505_1516	Translation initiation factor aIF-2 alpha subunit
AR505_1603	Translation initiation factor aIF-2 beta subunit
AR505_1606	Translation initiation factor aIF-5A
AR505_1752	Translation initiation factor aIF-2 gamma subunit
AR505_1755	Translation initiation factor aIF-2
Translation elongation factors	
AR505_1011	Elongation factor Tu domain
AR505_1454	Translation elongation factor aEF-1 alpha subunit
AR505_1455	Translation elongation factor aEF-2
AR505_1774	Translation elongation factor aEF-1 beta
Translation termination factors	
AR505_0252	Peptide chain release factor aRF1

Degradation of messenger RNA (mRNA). In addition to the aforementioned exosome complex that facilitates degradation of defective mRNAs, a NMD3 family protein (AR505_0020) involved in nonsense-mediated mRNA decay is predicted in the ISO4-H5 genome.

Protein fate

Secretion. The components of ISO4-H5 secretion systems are listed in Table 3.22.

Table 3.22. ISO4-H5 genes involved in secretion

Locus_tag	Gene product
AR505_0006	Signal recognition particle receptor FtsY
AR505_0758	Signal recognition particle protein SRP54
AR505_1473	Signal recognition particle protein SRP19
AR505_0386	Preprotein translocase subunit SecG
AR505_0241	Preprotein translocase subunit SecY
AR505_1799	Preprotein translocase subunit SecE
AR505_0667, AR505_1612, AR505_1828	Signal sequence peptidase SPI

Protein folding. A variety of chaperones are encoded by ISO4-H5 (Table 3.23.). A nascent polypeptide-associated complex protein (alpha-NAC homolog, AR505_0195) involved in stabilizing newly synthesised polypeptides is predicted. Two subunits of prefoldin PfdAB (*pfdAB*, AR505_0007, 1488) are predicted and are thought to act to transfer non-native polypeptides to chaperones. The molecular chaperones then hold the polypeptides in place and assist it to fold into its native state. Another class of enzymes that assist protein folding are the peptidyl-prolyl isomerases (AR505_0673, AR505_1028, AR505_1235). A class of chaperonin specific to archaea, the thermosome, was also seen. This high molecular weight complex is formed from only two subunits (AR505_0092, AR505_1062), which binds denatured polypeptides and functions as a chaperonin.

Table 3.23. ISO4-H5 genes involved in protein folding and stability

Locus_tag	Gene product
Prefoldin	
AR505_0007	Prefoldin alpha subunit PfdA
AR505_1488	Prefoldin beta subunit PfdB
Chaperones	
AR505_0093	Molecular chaperon GrpE
AR505_0094	Chaperon protein DnaK1
AR505_0095	Chaperon protein DnaJ
AR505_0286	Heat shock protein Hsp90
AR505_1228	Chaperon protein DnaK2
AR505_1232	Chaperon protein DnaK3
AR505_1248	ATP-dependent chaperon protein ClpB
Peptidyl-prolyl isomerase	
AR505_0685	Peptidyl-prolyl cis-trans isomerase FKBP-type
AR505_1028	Peptidyl-prolyl cis-trans isomerase
AR505_1235	Peptidyl-prolyl cis-trans isomerase
Thermosome	
AR505_0092	Thermosome subunit
AR505_1062	Thermosome subunit

Protein degradation. Misfolded, denatured proteins and spent regulatory proteins need to be degraded, as they could cause aggregation within the cell. The degradation of proteins in ISO4-H5 is mediated by an array of peptidases and a proteasome (Table 3.24.)

Table 3.24. ISO4-H5 genes involved in protein degradation

Locus_tag	Gene product
Proteasome	
AR505_0967	Proteasome endopeptidase complex
AR505_1482	Proteasome alpha subunit PsmA
AR505_1674	Proteasome-activating nucleotidase
Peptidase (transmembrane)	
AR505_0013	Peptidase M48 family
AR505_0264	Peptidase M50 family
AR505_0615	ATP-dependent protease S16 family
AR505_0702	Peptidase C1A papain
AR505_0721	CAAX amino terminal protease family protein
AR505_0785	Peptidase M50 family
AR505_1813	CAAX amino terminal protease family protein
Peptidase (cytoplasmic)	
AR505_0122	Metal-dependent protease
AR505_0144	Peptidase M24 family
AR505_0270	Peptidase M16 family
AR505_0449	Aminoacyl-histidine dipeptidase PepD
AR505_0907	Peptidase U62, modulator of DNA gyrase
AR505_0908	Peptidase U62, modulator of DNA gyrase
AR505_0919	Proline-specific peptidase
AR505_1001	Peptidase U32 family
AR505_1151	ATP-dependent protease
AR505_1318	Peptidase M18 family
AR505_1443	Methionine aminopeptidase Map

Regulation

The ISO4-H5 genome is predicted to encode 37 predicted transcriptional regulators, most with unknown functions (Table A.3.1.).

Table 3.25. ISO4-H5 genes involved in regulatory protein interaction

Locus_tag	Gene product
AR505_0138	Serine/threonine protein kinase RIO1 family
AR505_0206	Nitrogen regulatory protein P-II
AR505_0259	Phosphate uptake regulator PhoU1
AR505_0260	Phosphate uptake regulator PhoU2
AR505_0402	Nitrogen regulatory protein P-II
AR505_0609	Low molecular weight phosphotyrosine protein phosphatase
AR505_1176	Regulatory protein LytS
TPR repeat containing proteins	
AR505_0204	TPR repeat-containing protein
AR505_0275	TPR repeat-containing protein
AR505_0366	TPR repeat-containing protein
AR505_0367	TPR repeat-containing protein
AR505_0368	TPR repeat-containing protein
AR505_0370	TPR repeat-containing protein
AR505_0373	TPR repeat-containing protein
AR505_0419	TPR repeat-containing protein
AR505_0536	TPR repeat-containing protein
AR505_0557	TPR repeat-containing protein
AR505_0641	TPR repeat-containing protein
AR505_0793	TPR repeat-containing protein
AR505_0795	TPR repeat-containing protein
AR505_0898	TPR repeat-containing protein
AR505_0899	TPR repeat-containing protein
AR505_0917	TPR repeat-containing protein
AR505_1020	TPR repeat-containing protein
AR505_1048	TPR repeat-containing protein
AR505_1145	TPR repeat-containing protein
AR505_1212	TPR repeat-containing protein
AR505_1213	TPR repeat-containing protein
AR505_1214	TPR repeat-containing protein
AR505_1216	TPR repeat-containing protein
AR505_1225	TPR repeat-containing protein
AR505_1273	TPR repeat-containing protein
AR505_1334	TPR repeat-containing protein
AR505_1351	TPR repeat-containing protein
AR505_1463	TPR repeat-containing protein

Nitrogen metabolism

ISO4-H5 is capable of using nitrogen via ammonium uptake and glutamate/glutamine biosynthesis, as mentioned previously, but also has an array of genes associated with nitrogen fixation (Table 3.26.). However the involvement of these genes in nitrogen fixation activity is currently unknown.

Table 3.26. ISO4-H5 genes potentially involved in nitrogen fixation

Locus_tag	Gene product
AR505_0186	Nitroreductase family protein
AR505_0207	Nitrilase/cyanide hydratase and apolipoprotein N-aceyltransferase
AR505_0359	Nitrogenase iron protein NifH1
AR505_0591	Nitroreductase family protein
AR505_0822	Nitroreductase family protein
AR505_0956	Hydroxylamine reductase Hcp
AR505_1059	FeS cluster assembly scaffold protein NifU
AR505_1060	Cysteine desulfurase NifS
AR505_1268	Dinitrogenase iron-molybdenum cofactor biosynthesis protein
AR505_1288	Oxidoreductase/nitrogenase component 1
AR505_1289	Nitrogenase
AR505_1413	Dinitrogenase iron-molybdenum cofactor biosynthesis protein
AR505_1447	Nitrogenase cofactor biosynthesis protein NifB
AR505_1501	Nitrogenase iron protein NifH2
AR505_1502	Oxidoreductase/nitrogenase component 1

Oxidative stress

Under anaerobic conditions, oxygen (O_2) is not the terminal electron acceptor, so the formation of reactive oxygen species is not problematic; which explains the lack of genes such as catalases and superoxide dismutases (SOD) in ISO4-H5. However, rumen methanogens require genes to deal with oxidative stress (Table 3.27.). ISO4-H5 encodes a desulfoferrodoxin (*dfx*, AR505_0630) which detoxifies O_2 to hydrogen peroxide, and a FAD linked oxidase domain-containing protein (AR505_0263), which may also detoxify O_2. The hydrogen peroxide is detoxified by rubrerythrin (*rbr*, AR505_1353) to water via NADH peroxidase activity. Like most archaea, the ISO4-H5 genome lacks genes encoding the glutathione pathway, but it harbors the thiol redox pathway of thioredoxin (*trxA*, AR505_0047) and thioredoxin-disulfide reductase TrxB (*trxB*, AR505_0048).

Table 3.27. ISO4-H5 genes involved in oxidative stress

Locus_tag	Gene product
AR505_0212	CoA-disulfide reductase cdr
AR505_0263	FAD linkex oxidase domain-containing protein
AR505_0630	Desulfoferrodoxin Dfx
AR505_0700	Rubredoxin
AR505_0766	Peroxiredoxin AhpC
AR505_0970	Oxidoreductase
AR505_1047	Rubredoxin Rbr
AR505_1353	Rubrerythrin Rbr
Thioredoxin	
AR505_0047	Thioredoxin TrxA
AR505_0048	Thioredoxin-disulfide reductase TrxB1

3.3. Discussion

This chapter describes the sequencing, gap closing and bioinformatics analysis of the ISO4-H5 genome. ISO4-H5 is the first reported member of the Methanomassiliicoccales isolated from the ovine rumen, and its genome sequence represents an important contribution to the knowledge accumulating about this newly discovered group of rumen methanogens. Its 1.9 Mb genome and the analysis of its gene expression under different growth conditions has provided new insight into the lifestyle and cellular processes of this rumen methanogen.

A reported feature of genomes from members of the Methanomassiliicoccales order is the lack of organisation of rRNA genes into an operon (Borrel *et al.* 2014). rRNA genes are typically ordered 16*S*, 23*S* and 5*S*, and organised in an operon structure (Brosius *et al.* 1981). Archaeal rRNA operons vary in both structure and frequency (Dennis 1997), with 50% of sequenced archaeal genomes containing only one rRNA operon (Yip *et al.* 2013). However, the number of copies of rRNA operons within archaeal genomes can be as high as four (Yip *et al.* 2013), with some rRNA operons containing one or two tRNAs within spacer sequences, including $tRNA^{Glu}$, $tRNA^{Ile}$, $tRNA^{Ala}$ and $tRNA^{Cys}$ (Lund *et al.* 1976; Ikemura and Nomura 1977; Morgan *et al.* 1977; Keller *et al.* 1980; Chant and Dennis 1986). Some archaeal genomes are found with the 5*S* rRNA gene existing outside the rRNA operon, as observed in the *Desulfurococcus mobilis*, *Sulfolobus solfataricus* and *Methanococcus jannaschii* genomes (Kjems and Garrett 1987; Bult *et al.* 1996; She *et al.* 2001). Rumen methanogens typically have their rRNA genes organised in an operon, however, the ISO4-H5 genome lacks an rRNA operon. This is similar to the *Thermoplasma acidophilum* genome (Tu and Zillig 1982) from the order Thermoplasmatales, to which members of the Methanomassiliicoccales were originally assigned (Tajima *et al.* 2001). While the separation of the 16*S*, 23*S* and 5*S* rRNA genes appears to be a trait shared by all members of the Methanomassiliicoccales (Borrel *et al.* 2014) the reason for this is not clear. The cellular rRNA content determines the rate of protein synthesis and cytoplasmic mass accumulation, the location of rRNA operon is on average 10 min from the origin of replication in an *E. coli* with a doubling time of 100 min (Bremer and Dennis 2008). It is conceivable that the closer the rRNA genes to the origin of replication, the earlier it would produce rRNA and facilitate protein synthesis, therefore the location of rRNA genes may be correlated with the growth rate of the organism (Bremer and Dennis 2008). There are seven rRNA genes in ISO4-H5, and their average distance to the origin of replication is 16% of the genome size, the longer distance may suggest the ISO4-H5 could possibly be a

slow growing organism. The genome of ISO4-H5 encodes 47 tRNAs encoding for all amino acids. Introns in transfer RNAs are found in all three kingdoms of life and present on average in 15% of archaeal tRNAs (Yoshihisa 2014). Three of the ISO4-H5 tRNAs; $tRNA^{Met}$, $tRNA^{Trp}$ and $tRNA^{Tyr}$, contain an intron. The tRNA intron endonuclease *endA* (AR505_1808) required for tRNA intron removal (Thompson *et al.* 1989) has been identified within the ISO4-H5 genome. A difference observed in the $tRNA^{Met}$ intron in ISO4-H5 is that it does not encode a C/D box snRNA, which is thought to be involved in *O*-methylation of rRNA in *Nanoarchaeum equitans* (Randau *et al.* 2005). This suggests that either *O*-methylation of rRNA does not occur in ISO4-H5 or it might be carried out by other means. The ISO4-H5 genome is predicted to encode only single copies of $tRNA^{Trp}$ and $tRNA^{Tyr}$, therefore, intron removal is essential to allow the synthesis of their mature tRNAs needed for translation. The pseudouridine modification of $tRNA^{Tyr}$ is dependent on its intron in *Saccharomyces cerevisiae* (Johnson and Abelson 1983), therefore the intron in $tRNA^{Tyr}$ of ISO4-H5 may perform a similar role, because ISO4-H5 encodes genes required for pseudouridylation within its genome. The $tRNA^{Trp}$ intron has been observed to perform a catalytic function in *Pyrococcus abyssi*, where it acts as a small nucleolar RNA (snoRNA) and guides the ribose methylation of tryptophan pre-tRNA exons (D'Orval et al. 2001). However, the ISO4-H5 $tRNA^{Trp}$ intron sequence does not match any known sequence in the Rfam database (Griffiths-Jones *et al.* 2005) and whether the ISO4-H5 intron in $tRNA^{Trp}$ harbors catalytic activity remains unknown.

Horizontal gene transfer occurs commonly amongst prokaryotes, and continuous acquisition and removal of genes in the genome is an integral part of genome evolution (Ochman *et al.* 2000; Koonin *et al.* 2001), assisting with environmental adaptation (Gilbert and Cordaux 2013). The genes acquired from horizontal transfer often differ to other genes in the genome in base composition, codon usage and bias of di- and tri-nucleotides (Ochman *et al.* 2000). Two regions in the ISO4-H5 genome are considered horizontally acquired based on their atypical base composition (Table 3.9.). The first region (Region 7) contains genes involved in dTDP-L-rhamnose biosynthesis (AR505_0552 – 0554). In Gram-negative bacteria, rhamnose is a component of the outer membrane exopolysaccharide (EPS) as well as the large outer membrane lipopolysaccharide known as the *O*-antigen (Samuel and Reeves 2003; Jofre *et al.* 2004). Rhamnose has been detected in EPS in the methanogens *Methanobacterium formicum* and *Methanosarcina mazeii* (Veiga *et al.* 1997). This suggests that rhamnose may play a role in the cell envelope structure of Methanomassiliicoccales. A second region (Region 4) contains a large cluster of hypothetical genes, but a closer examination identifies a 25 Kb section with

25 predicted ORFs (AR505_0333 to AR505_0358) which show some characteristics of a prophage (Canchaya *et al.* 2003). Region 4 contains a phage integrase (AR505_0313) which is required for phage integration, and three DNA-cytosine methyltransferases genes (AR505_0327, AR505_0339, AR505_0340) required to circumvent host restriction-modification barriers which increase the likelihood of successful infection (Buhk et al. 1984; Yasmin et al. 2010). However, no phage structural components were discernable within this region, so it is unclear whether Region 4 represents an integrated prophage, phage remnant or other mobile element.

Insertion sequence elements and other mobile elements contribute to plasticity and evolution of archaeal genomes, and are capable of facilitating horizontal transfer (Brügger *et al.* 2002). IS elements are segments of DNA, less than 2.5 Kb in length and capable of inserting into a target sequence (Frost *et al.* 2005). IS elements often encode a transposase flanked with short inverted repeats necessary for transposase recognition and cleavage (Mahillon and Chandler 1998). A region of DNA flanked by two IS elements may be mobilized as a composite transposon (Siguier *et al.* 2015). IS elements constitute 2.3% of the ISO4-H5 genome, whereas IS elements were found to constitute only 1.2% of the Mx1201 genome (Borrel *et al.* 2012). Higher number of IS elements may act as a driving force of genome evolution (Biemont and Vieira 2006), aiding ISO4-H5 to persist and define its niche within the competitive habitat of the rumen environment.

The ISO4-H5 genome is predicted to encode two toxin and four antitoxin system genes (Table A.3.1.). Toxin/antitoxin systems are typically identified within plasmids, where they trigger the death of daughter cells that fail to inherit the plasmids (Hayes 2003). Toxin/antitoxin genes have also been identified within chromosomes where they function in programmed cell death and in metabolic control during nutritional stress (Gerdes *et al.* 2005; Allocati *et al.* 2015). The ISO4-H5 genome has one example of toxin and antitoxin genes being co-located which suggests they function together as a system (Gerdes *et al.* 2005). The predicted toxin gene (AR505_0858) is 34.4% homologous to the predicted toxin gene *relE* (WP_011992585.1) from the lambdaSa04 prophage found in *Campylobactur curvus* 525.92 (NCBI Reference Sequence NC_009715.1). Similarly, the antitoxin gene is 29.8% homologous to the prevent-host-death protein (WP_009650570.1) from the same prophage and organism. The function of the RelE protein in bacteria and archaea is to cleave the mRNA at the ribosome and trigger non-lethal arrest of translation that inhibits cell growth, and its function can be triggered by nutrient starvation (Christensen and Gerdes 2003). There is no genomic evidence to suggest the

presence of a plasmid in the ISO4-H5, which in turn suggests the toxin/antitoxin may function as a regulator to halt cellular growth. There is a TetR family regulator with a predicted transmembrane helice immediately downstream of the *relE* gene. The TetR family regulator HapR in *Vibrio cholera* regulates virulence genes, HapR has been found to be regulated by quorum-sensing signals (Miller *et al.* 2002). Quorum sensing may converge with starvation-sensing (Lazazzera 2000). The TetR family regulator gene downstream of *relE* gene may function similary, regulated by quorum/starvation-sensing signals and regulating the toxin/antitoxin system.

The ISO4-H5 genome is predicted to contain a CRISPR element. CRISPR elements typically confer acquired immunity against phage and conjugative plasmids (Marraffini and Sontheimer 2010). The CRISPR elements function in a manner similar to RNA interference; the spacer sequence between repeat elements contains DNA sequence from former invading genetic elements. Spacer regions can uptake new material to match new foreign genetic elements and prevent conjugation and transduction of foreign genetic elements (Marraffini and Sontheimer 2008). The spacer sequences of the identified CRISPR element were used to identify foreign genetic elements that ISO4-H5 may have encountered in the past (Biswas *et al.* 2013; Shariat *et al.* 2015). Out of 32 spacer sequences analysed, 17 matched known phage and plasmid sequences. Plasmid hosts included *Azospirillum* spp. and *Rahnella* spp., both genera are nitrogen-fixing bacteria found in the rhizosphere of grass, soybeans or wheat (Berge *et al.* 1991; Steenhoudt and Vanderleyden 2000; Weidner *et al.* 2012). It is conceivable that within the rumen environment ISO4-H5 may have come into contact with similar organisms colonising forage plant material, and may have been exposed to their mobile genetic elements.

The defining metabolism of the Methanomassiliicoccales is their formation of CH_4, and like others members of this order (Borrel *et al.* 2014), ISO4-H5 possesses a complete set of genes required to carry out methylotrophic methanogenesis using methyl substrates. The methyl groups are transferred onto CoM by methyltransferases and then reduced to CH_4 by the Mrt complex (Borrel *et al.* 2014). This particular pathway of methylotrophic methanogenesis appears to be unique to the Methanomassiliicoccales order (Borrel *et al.* 2014). Other methylotrophic methanogens within the rumen, such as *Methanosarcina* spp. disproportionate methanol by electron bifurcation, oxidizing it to produce CO_2 while generating reducing potential to reduce the methyl group to CH_4 (Welander and Metcalf 2005). The methanogenesis pathway in ISO4-H5 lacks the genes encoding the enzymes required to oxidize methanol to CO_2, and is predicted to only reduce methylated compounds directly to CH_4. Functionally, this

is similar to *Methanosphaera stadtmanae* MCB-3, which encodes all the genes for the enzymes needed to oxidize methanol to CO_2 but does not use this pathway due to the lack of genes encoding synthesis of molybdopterin, a cofactor required for formation of an active formylmethanofuran dehydrogenase (Fricke *et al.* 2006). The Methanomassiliicoccales is the second most abundant order of rumen methanogens (Janssen and Kirs 2008), which suggests that ISO4-H5 and other rumen Methanomassiliicoccales compete successfully against *Methanosarcina* spp. and *Methanosphaera* spp. for methyl substrates. ISO4-H5 strictly depend on methylotrophic methanogenesis and direct reduction of methyl-substrates, theoretically ISO4-H5 can produce one mole of CH_4 from one mole of H_2, whereas hydrogenotrophic methanogens require four moles of H_2 to produce one mole of CH_4, this suggests ISO4-H5 likely only requires a small amount of H_2 to survive, which could provide the rumen Methanomassiliicoccales an unique advantage and allowed establishment of their niche within the rumen environment. While *Methanosphaera* spp. also utilises methylotrophic methanogenesis, they lack the specialization of ISO4-H5 as they carry the full array of hydrogenotrophic methanogenesis genes, which implies additional cellular expenditure, which may explain why members of Methanomassiliicoccales order are more abundant than *Methanosphaera* spp. in the rumen.

Based on the genome sequence, ISO4-H5 is predicted to utilise methanol, mono-, di-, tri-methylamine and methylthiol as substrates for methanogenesis. The ability to utilise five different substrates for methanogenesis offers ISO4-H5 a competitive advantage over the methylotrophic rumen methanogen *Methanosphaera*, which is limited to the use of methanol as a methyl substrate (Jeyanathan, 2010). Three methyltransferase genes are typically required to transfer the methyl group from the substrate onto CoM, and to utilise methanol ISO4-H5 contains the genes, *mta*A, *mta*B, and *mta*C. MtaB is a methanol-specific corrinoid methyltransferase which transfers the methyl group from methanol onto the corresponding corrinoid protein component, MtaC. MtaA is a methylcobalamin: CoM methyltransferase responsible for transferring the methyl group from the corrinoid protein onto the CoM for methanogenesis. Similar sets of genes are identified within the ISO4-H5 genome for the use of methylamines. One of the more interesting features of the genomes of members of the Methanomassiliicoccales is the requirement of pyrrolysine for methylamine utilisation (Borrel *et al.* 2014). The ISO4-H5 methylamine:corrinoid methyltransferase genes *mtmB1, mtmB2*, *mtbB* and *mttB* all require amber codon read-through to be fully transcribed and incorporate

pyrrolysine. The methanol corrinoid protein gene *mtaC2* also requires amber codon read-through to function.

ISO4-H5 encodes four methylcobalamin:CoM methyltransferase *mtaA/mtmA* genes which allow for the use of either methanol and methylamines as substrates for methanogenesis. Experimental evidence is required to assess the substrate specificity, such as conducted in *M. acetivorans* (Bose *et al.* 2006; Bose *et al.* 2008). The methylthiol:CoM methyltransferase, MtsA, required for the utilisation of methylthiol as a substrate, has been identified as a bifunctional enzyme capable of transferring the methyl group from substrates onto a coenzyme corrinoid protein, MtsB, as well as transferring the methyl group onto CoM (Tallant *et al.* 2001). While methyltransferase systems for methanol and methylamines have three components, the methylthiol methyltransferase only requires the MtsA itself and the corresponding cobalt and iron containing corrinoid Fe-S protein MtsB (Tallant *et al.* 2001). The methylthiol utilisation genes *mtsA* and *mtsB* in ISO4-H5 are homologous to enzymes in *Methanosarcina barkeri* (Tallant and Krzycki 1997), and one of the substrates of the ISO4-H5 MtsA, methyl-3-methylthiopropionate, has been experimentally verified in Chapter 5.

ISO4-H5 is predicted to encode an alcohol dehydrogenase (AR505_0483), and ethanol is known to be a methanogenic substrate for methanogens such as *Methanofollis ethanolicus* (Imachi *et al.* 2009). Ethanol has also been observed to enhance methanogenesis in *Mbb. ruminantium* $M1^T$, presumably by the oxidation of ethanol by its NADP-dependent alcohol dehydrogenases and capture of the reducing potential as $F_{420}H_2$ by the action of their NADPH-dependent F_{420} dehydrogenase (Leahy *et al.* 2010). This metabolism is predicted to spare some of the H_2 or formate normally used to produce $F_{420}H_2$ and thus stimulate growth (Leahy *et al.* 2010). Since neither F_{420} biosynthesis nor NADPH-dependent F_{420} dehydrogenase genes were predicted within the ISO4-H5 genome, ISO4-H5 is unlikely to utilise ethanol as a supply of reducing potential for methanogenesis. The *Methanosphaera* sp. WGK6 has been observed to use ethanol as a source of reducing potential via alcohol dehydrogenase and aldehyde dehydrogenase (Hoedt *et al*. 2016). The role of the alcohol dehydrogenase in ISO4-H5 ethanol metabolism requires additional experimental investigation.

For all methanogenic substrates, the methyl group is attached to CoM and reduced by methyl-CoM reductase to produce CH_4 (Thauer 1998). Two different isozymes of the methyl-CoM reductase have been observed in nature; MCR I, abbreviated as Mcr, and MCR II, abbreviated as Mrt. The thermophilic methanogen, *Methanobacterium thermoautotrophicum* harbours both isozymes and appear to be differentially regulated at different stages of growth (Rospert *et al.*

1990; Bonacker *et al.* 1993; Pihl *et al.* 1994; Reeve *et al.* 1997; Friedrich 2005). The two isozymes display different operon structures within a genome, *mcrBDCGA* and *mrtBDGA*. The gene organization of the methyl-CoM reductase in ISO4-H5 suggests it belongs to the MCRII/Mrt isozyme, similar to what was found in other strains of Methanomassiliicoccales (Lang *et al.* 2015). The metagenome and metatranscriptome has been characterised from rumen contents of high and low CH_4 emitting sheep, it was found the increased CH_4 emission was primarily associated with increases in the expression of *mcr/mrt* operon (Shi *et al.* 2014). As mentioned above, some methanogens harbour both the Mcr and Mrt, such as the *Methanobrevibacter* sp. D5 described in Chapter 6, while some Methanobrevibacters spp. only possess Mcr, as described in *Mbb. ruminantium* $M1^T$ (Leahy *et al.* 2010). The *mcr/mrt* operon were classified into three groups, and high expression of two groups were associated with increased CH_4 emission, the group extended from the *mcr* operon of *Mbb*. spp, and the group that extended from the *mrt* operon os ISO4-H5 (Shi *et al.* 2014). This suggests ISO4-H5 and similar methanogens play an important role in conferring CH_4 emission of the rumen (Shi *et al.* 2014), which must be taken into consideration when formulating strategies to lower CH_4 emission.

The catalytic action of Mrt depends on the corphin, cofactor F_{430}, the nickel-containing prosthetic group in the active site of methyl-CoM reductase (Scheller *et al.* 2010). The biosynthetic pathway for F_{430} is proposed to involve seven steps, but only the enzyme involved in the first step, uroporphyrin-III C-methyltransferase, CorA, has been experimentally proven (Pfaltz *et al.* 1987; Thauer and Bonacker 1994). The enzymes involved in the remaining steps of the pathway have yet to be identified (Pfaltz *et al.* 1987). ISO4-H5 is predicted to encode a *cor*A gene (AR505_1041), which is homologous to the *cor*A gene of *Methanobacterium ivanovii* sharing 43.3% aa identity (Blanche *et al.* 1991). To date, the F_{430} has been uniquely conserved across methanogens (Diekert *et al.* 1981), therefore it is likely ISO4-H5 possesses F_{430}.

CoM is required for the methanogenesis pathway as all methyl groups are transferred to CoM before being reduced to CH_4 (Thauer *et al.* 2008). The ISO4-H5 genome does not contain the *comABCDE* genes required for CoM biosynthesis (Graupner *et al.* 2000). An alternative method of CoM biosynthesis utilising sulfopyruvate biosynthesis has been proposed, involving the genes, threonine synthase-like L-cysteate synthase and an aspartate aminotransferase (Graham *et al.* 2009). Unlike *Methanosarcina acetivorans*, whose cysteate synthase was discovered originally as threonine synthase (Graham *et al.* 2009), only one threonine synthase

is present within the ISO4-H5 genome, and as such, is unlikely to be a L-cysteate synthase. This means both known pathways of CoM biosynthesis are predicted to be absent in the ISO4-H5 genome, suggesting ISO4-H5 likely requires an external supply of CoM to survive. This is not uncommon within the rumen environment, *Mbb. ruminantium* $M1^T$ also requires external CoM to survive (Leahy *et al.* 2010). The lack of CoM biosynthesis genes is consistent with the original enrichment condition of ISO4-H5, which included 10 mM CoM in the medium (Jeyanathan, 2010), and CoM has been used routinely for *in vitro* cultivation of this methanogen.

The heterodisulfide reductase Hdr is essential for cofactor regeneration during methanogenesis. Hdr has been observed in two classes; a cytoplasmic HdrABC complex that is found across many methanogens, and a transmembrane HdrED found in *Methanosarcina* and *Archaeoglobus* (Mander *et al.* 2004; Buan and Metcalf 2010). HdrE is a membrane-anchored cytochrome, and HdrD contains motifs for iron sulphur cluster binding and two conserved cysteine motifs important in reoxidation of CoM-S-S-CoB (Kaster *et al.* 2011). ISO4-H5 encodes the cytoplasmic HdrABC genes, as well as two HdrD genes without a HdrE gene counterpart. In *Methanosarcina*, HdrDE is coupled to a $F_{420}H_2$ dehydrogenase (Fpo) complex during methylotrophic methanogenesis (Baumer *et al.* 2000). In order for Hdr to function and recycle the CoB and CoM cofactors, it requires electron donors. In other methanogens the electron donor is reduced F_{420}, which is reduced by the F_{420} reducing hydrogenase (Frh), with H_2 being the electron donor (Kulkarni *et al.* 2009; Thauer *et al.* 2010). However, ISO4-H5 lacks the genes encoding Frh. Morphological evidence suggests the ISO4-H5 cell lacks F_{420}, as it does not possess the typical fluorescence emitted by F_{420} under 420 nm (Jeyanathan 2010). *Methanomassiliicoccus luminyensis* was initially reported to be autofluorescent (Dridi *et al.* 2012) but later reported as non-fluorescent (Lang *et al.* 2015), while both strains of Methanomassiliicoccales isolated from the termite and millipede gut have been reported to lack autofluorescence (Paul *et al.* 2012). The methyl viologen hydrogenase, Mvh, has been known to utilise H_2 and supply electrons to the Hdr complex using Fdx (Thauer *et al.* 2010). Therefore, it is predicted that ISO4-H5 acquires electrons via Mvh.

The Fpo enzyme is a large transmembrane complex that can oxidize $F_{420}H_2$ and pump protons out of the cell (Deppenmeier 2004). It is highly similar to the prokaryotic NADH:quinone oxidoreductase and the subunits have been named accordingly. It is comprised of 13 subunits, FpoA,B,C,D,F,H,I,J,K,L,M,N,O, and all the genes encoding these subunits, apart from *fpoF*, are within an operon (Deppenmeier 2004). The *fpoF* and *fpoO* genes are absent in the ISO4-

H5 genome. FpoF in other methanogens contains a FAD binding site and an iron-sulfur cluster binding motif, and is thought to act as a receptor of $F_{420}H_2$ (Welte and Deppenmeier 2011). FpoO also contains an iron-sulfur cluster binding motif, and is likely responsible for reduction of methanophenazine. ISO4-H5 is predicted to lack both methanophenazine and F_{420}, which would make *fpoF* and *fpoO* genes unnecessary, this may explain the absence of these genes from the ISO4-H5 genome. The absence of F_{420} from ISO4-H5 means another electron carrier must be used. *Methanosaeta* spp. also lack the *fpoF* gene; it has been proposed that a Fdx can act as an alternative electon carrier (Welte and Deppenmeier 2011). Fdx has been observed to couple Hdr and Mvh electron bifurcation between methanogenesis and energy generation (Kaster *et al.* 2011). ISO4-H5 is predicted to encode three *fdx* genes. In another member of Methanomassiliicoccales, *Candidatus* Methanoplasma termitum", the HdrD that contains an iron-sulfur binding cluster has been proposed to physically interact with the Fpo-like complex forming an energy-converting Fdx:heterodisulfide oxidoreductase (Lang *et al.* 2015). The electron bifurcation proposed to occur at the "*Ca.* Methanoplasma termitum" Mvh/HdrABC complex means that only the electrons of every second H_2 oxidized by the Mvh/HdrABC complex will go to the Fpo-like/HdrD complex. This will reduce the ATP production and growth yield of this organism, but will mean its threshold for H_2 will be lower (Lang et al 2015). The energy conservation system in ISO-H5 is predicted to be identitical to "*Candidatus* Methanoplasma termitum" and this has several consequences for its existence in the rumen. ISO4-H5 is likely to have a low ATP yield and its growth will therefore be slower. However, ISO4-H5 is predicted to have a lower H_2 threhshold, which may offer a competitive advantage against other rumen methylotrophic methanogens. *Methanosphaera stadtmanae* for example, has the same general metabolic stoichiometry for CH_4 formation from methanol, but pumps two ions per CH_4 formed (Thauer *et al.* 2008) and is likely to grow faster, with a higher H_2 threshold. While the hydrogenotrophic and methylotrophic methanogens may undergo cycles of growth and decline as ruminal H_2 levels rise and fall, ISO4-H5 and other rumen Methanomassiliicoccales are likely to grow more consistently, albeit at a slower rate.

A number of protein-coding genes have been found as genetic markers for methanogenic archaea. The function of these proteins remain unknown but they are hypothesised to be methanogenesis related, and as such have been termed methanogenesis marker proteins (MMP) (Gao and Gupta 2007). There are 17 MMPs currently described, all of which are conserved among methanogens (Gao and Gupta 2007). ISO4-H5 lacks MMPs 9, 10, 12, 14. It is possible these four MMPs are associated with parts of the methanogenesis pathways that are absent in ISO4-H5.

The central carbon metabolism pathway of ISO4-H5 is depicted in detail in Figure 3.7. The genes present suggest acetate is the primary carbon source of ISO4-H5, and this is consistent with the requirement of acetate for growth (Jeyanathan, 2010). Supplementation of the acetate metabolism inhibitor fluoroacetate (Cappenberg 1974) in culture may clarify the usage of acetate in ISO4-H5. In methanogens, one carbon metabolism begins with the conversion of NADPH to formate by formate dehydrogenase (Andreesen and Ljungdahl 1974), and the incorporation of formate into tetrahydrofolate via formate-tetrahydrofolate ligase. Downstream reactions lead to the production of methyl-tetrahydrofolate, which then condenses with CO to form acetyl-CoA by the CO dehydrogenase complex (Stupperich *et al.* 1983). Acetyl-CoA can enter into the citric acid cycle or be converted to pyruvate by pyruvate synthase and enter gluconeogenesis. ISO4-H5 utilises acetate as the main carbon source via acetyl-CoA synthetase. The acetyl-CoA synthetase appears to be a single, fused enzyme instead of two subunits commonly found in other organisms, such as the archaeal genome of *Pyrococcus furiosus* (Musfeldt *et al.* 1999). The ISO4-H5 genome is predicted to lack a methylenetetrahydrofolate reductase, a methyltetrahydrofolate:corrinoid methyltransferase and a CO dehydrogenase complex, which suggests ISO4-H5 is unable to produce acetyl-CoA from formate, unlike B10 and Mx1 (Lang *et al.* 2015). Although, ISO4-H5 can utilise formate to produce 5,10 methylene-tetrahydrofolate via the formate-tetrahydrofolate ligase (*fhs*) and bifunctional NADP-dependent methylene tetrahydrofolate dehydrogenase (*folD*) genes, it is not predicted to incorporate CO. The 5,10 methylene-tetrahydrofolate can be used for pyrimidine salvage and biosynthesis via the thymidylate synthase gene. It is interesting to note that in an environment of limited acetate, other Methanomassiliicoccales, and indeed *Methanosphaera* spp., may outcompete ISO4-H5 due to their ability to use formate as a carbon source. However, acetate is usually abundant in the rumen and the acetate requirement of ISO4-H5 may reflect its adaptation to this environment.

During central metabolism acetyl-CoA may proceed into the TCA cycle or gluconeogenesis pathway. Genomes of members of the Methanomassiliicoccales have been reported to possess an incomplete TCA cycle (Lang *et al.* 2015). The incomplete TCA cycle of ISO4-H5 begins with fumarate produced from arginine biosynthesis and ends with 2-oxoglutarate, which feeds into glutamate biosynthesis (Pitson *et al.* 1999). Fumarate is one of the endpoints of the TCA cycle, it is recycled into purine ribonucleotide by PurB (Zhang *et al.* 2008). The gluconeogenesis pathway paves the way to the pentose phosphate pathway. The pentose phosphate pathway connects glyceraldehyde-3-phosphate from gluconeogenesis to 5-phospho-

D-ribose 1-diphosphate (PRPP) for purine biosynthesis (Sprenger 1995; Kato *et al.* 2006). The ISO4-H5 pathway is predicted to rely heavily on the broad substrate specificity of the enzymes transketolase and transaldolase (Cunin *et al.* 1986). The additional D-xylulose-5-phosphates generated by the process can be recycled by Rps, HxlA, HxlB back to D-fructose-6-phosphate (Kato *et al.* 2006). The *hxlA* and *hxlB* genes do not form an operon, unlike in *Bacillus subtilis* (Yasueda *et al.* 1999).

The ISO4-H5 genome is predicted to contain all the genes required to synthesize 19 amino acids, with the exception of glycine and methionine. Therefore ISO4-H5 may require an external supply of glycine; but a transporter for glycine has yet to be identified within the genome of ISO4-H5. The methionine biosynthesis pathway is depicted in Figure 3.9. Several genes required for the synthesis of methionine are not found in the genome of ISO4-H5 and whether ISO4-H5 is capable of methionine biosynthesis remains unclear. There are two potential routes of methionine biosynthesis: a three step route that uses *O*-acetyl-L-homoserine as intermediate lacks the *met2* gene in second step, and a five step route that uses *O*-succinyl-L-homoserine as an intermediate lacks all the enzymes from the third step onwards (Figure 3.9.). The first step is universally carried out by MetL (Belfaiza *et al.* 1984) and ISO4-H5 encodes a *metL* gene. This gene is predicted to encode a homoserine dehydrogenase domain at its C-terminal end (Belfaiza *et al.* 1984). MetL from *E. coli* is a bifunctional enzyme with the N-terminal end acting as an aspartate kinase (Belfaiza *et al.* 1984). Without the N-terminal aspartate kinase domain in the the ISO4-H5 *metL* gene, it is predicted to carry out only one function - homoserine biosynthesis from aspartate-semialdehyde. The methionine biosynthesis pathway branches at L-homoserine. The Met2, homoserine *O*-acetyltransferase, is involved in producing *O*-acetyl-L-homoserine (Yamagata 1987), and ISO4-H5 lacks a *met2* gene. The gene *O*-acetyl-L-homoserine sulfhydrylase (AR505_0738) is also known as *O*-acetylhomoserine aminocarboxypropyltransferase, Met17. This gene has been observed to convert the *O*-acetyl-L-homoserine to L-methionine using methanethiol in *Neurospora* (Moore and Thompson 1967). It has been shown to react with other thiols to form homocysteine, an intermediate for methionine biosynthesis in many organisms (Shimizu *et al.* 2001). There are four enzymes capable of converting homocysteine to L-methionine, MetH, MetE, Met6 and Bhmt (Figure 3.9.), but none of these genes have homologues within the ISO4-H5 genome. In the other route, ISO4-H5 is predicted to encode MetA that can produce *O*-succinyl-L-homoserine but there are no enzymes encoded in the genome to process this substrate further. The enzymes present in both routes could complement each other to produce methionine if the Met17 gene could utilise

O-succinyl-L-homoserine as a substrate instead of *O*-acetyl-L-homoserine. Methionine is also an important part of the *S*-adenosyl-L-methionine (SAM) cycle (Loenen 2006) and SAM is an important cofactor involved in methylation reactions. Homologues of enzymes required for the SAM cycle and SAM dependent DNA methylation are predicted within the ISO4-H5 genome, apart from methionine synthase (Table 3.12.). ISO4-H5 is unable to produce spermidine from SAM due to the lack of a spermidine synthase gene. It is possible that ISO4-H5 has adapted a novel mechanism of L-methionine production, as it is unlikely for an organism to carry incomplete versions of both biosynthesis and utilisation of methionine, as well as the SAM cycle, if they are not being utilised in some way by the organism.

An interesting aspect of ISO4-H5 amino acid biosynthesis was that although it is predicted to be able to produce asparagine, the amino source of the enzymatic reaction remains elusive. The asparagine synthase B AsnB converts aspartate to asparagine (Scofield *et al.* 1990). The ISO4-H5 genome is predicted to encode two AsnB genes (predicted molecular weight, 35 kD and 28kD), which are smaller than the experimentally verified *E. coli* AsnB (63kD) (Scofield *et al.* 1990). A multiple alignment reveals that the ISO4-H5 *asnB* genes only possess the C-terminal domain that is responsible for binding of ATP and aspartate, but not the N-terminal domain that hydrolyses the glutamine (Larsen *et al.* 1999). This suggests the ISO4-H5 AsnB does not utilise glutamine as the amino source. Further study on the enzyme kinetics of this AsnB would be required to identify the amino source.

As mentioned earlier, the amino acid pyrrolysine is required for expression of enzymes involved in methylamine utilisation. Pyrrolysine was discovered as the 22nd amino acid in 2002 in the crystal structure of monomethylamine methyltransferase MtmB from *Methanosarcina barkeri* (Hao *et al.* 2002). Pyrrolysine is encoded by the UAG codon, also known as the amber codon (Hao *et al.* 2002). The pyrrolysine usage extends beyond the order Methanomassiliicoccales and is found amongst other anaerobic archaea as well as bacteria (Herring *et al.* 2007; Prat *et al.* 2012). Methanogenic archaea that utilise pyrrolysine include *Methanosarcina acetivorans* (Gaston *et al.* 2011a), *Methanococcoides burtonii* (Gaston *et al.* 2011a), *Methanohalobium evestigatum* (Gaston *et al.* 2011a), and *Methanohalophilus mahii* (Gaston *et al.* 2011a). Four genes are required for the biosynthesis of pyrrolysine from lysine, as well as for the biosynthesis of the pyrrolysine-tRNA (Table 3.12.), *pylS, pylB, pylC, pylD* (AR505_1325 - 1322). The amber suppressor tRNA^{CUA} gene, *pylT* is also required. The organisation of the pyrrolysine biosynthesis operon in ISO4-H5 (Figure 3.6.) is identical to what has been found in the *Methanosarcina barkeri* genome (Zhang *et al.* 2005) and being

highly similar to what has been described for other members of the Methanomassiliicoccales order (Srinivasan *et al.* 2002; Borrel *et al.* 2014).

The recoding of the amber codon for pyrrolysine is not the only alternative coding of a stop codon observed in nature. The opal stop codon UGA has been observed to encode selenocysteine in many organisms (Johansson *et al.* 2005), and tryptophan has been observed to be encoded by UGA in *Mycoplasma capricolum* (Yamao *et al.* 1985). Recent advancements in single cell sequencing technology has acquired genome sequences of uncultivated micro-organisms, and it revealed the recoding of UGA to glycine in *Gracilibacteria* (Rinke *et al.* 2013). In ISO4-H5 the opal codon acts as the major stop codon and accounts for 84.1% of all predicted stop codons, while the ochre stop codon accounts for 13.4% (Table 3.5.). While read-through of the amber codon may be essential for the complete translation of functional methylamine methyltransferase(s), it can also produce what is effectively a 'nonstop' mutation to other proteins carrying the in-frame amber codon (Wong and Schwartzbach 2015). The 'nonstop' mutation has been studied in humans, where 119 identified 'nonstop' mutations in 87 proteins were found to be associated with an inherited disease, as the extended amino acid often caused peptides to fold differently, which caused alteration or loss of functionality. The longer the peptide extension the more likely the resulting peptide was found to fold differently and lose the original function (Ameri *et al.* 2007; Hamby *et al.* 2011).

The recoding of stop codons leads to an interesting point on how ISO4-H5 regulates pyrrolysine incorporation. Based on sequence homology, there are 10 genes predicted to utilise the amber codon as a stop codon rather than pyrrolysine incorporation (Table 3.8.). This includes nine IS elements and one adenylate kinase gene. The importance of pyrrolysine to some of these genes is unclear, however it is obvious that without pyrrolysine, ISO4-H5 would be unable to utilise methylamines as a source of methyl-compounds for methanogenesis. The abundance of methylamines in the rumen environment (Saleem *et al.* 2013) combined with the low number of genes utilising the amber codon as a stop codon, (including the uncertain genes only accounts for 1.5% of all genes), suggests that the pyrrolysine operon is constitutively expressed and no regulation mechanism is in place within ISO4-H5. Therefore, the additional 10 genes predicted to use the amber codon as a stop codon would encode extended and likely misfolded, non-functional proteins. The loss of function to the nine IS elements means the nine IS elements may extend and potentially lose function, however, the inverted repeats flanking the IS element is not influenced by transcription itself, and the inverted repeats may potentially influence the expression of surrounding genes. With regard to the adenylate kinase gene, if the

20 extended amino acids caused the adenylate kinase to misfold, ISO4-H5 does carries a second copy of an adenylate kinase (AR505_0262) which does not have a pyrrolysine codon. ISO4-H5 may lose some energy through producing a non-functional adenylate kinase, but it would not lose this cellular function completely.

The ISO4-H5 is predicted to possess a near complete pathway of *de novo* nucleotide biosynthesis, as illustrated in Figure 3.12, the purine biosynthesis pathway is similar to what has been reported for other members of Methanomassiliicoccales (Borrel *et al.* 2014). The enzyme required for conversion of IMP to XMP remains to be identified. One of the steps, the conversion of amino-(5-phospho-ribosyl) imidazole (AIR) to amino-(5-phospho-ribosyl) imidazole-4-carboxylate (CAIR) is helpful to ISO4-H5 energy conservation. Instead of the ATP-dependent two-step reaction catalysed by carboxyaminoimidazole ribonucleotide synthetase (PurK) and carboxyaminoimidazole ribonucleotide mutase (PurE), AIR is converted to CAIR by phosphoribosylaminoimidazole carboxylase (Ade2) which incorporates CO_2 (Stotz and Linder 1990). This reduces the ATP requirement of purine *de novo* biosynthesis in ISO4-H5 and suggests ISO4-H5 may have a preference to enzymes that do not consume ATP.

The predicted *deoA*, *rbcL* and *e2b2* operon is an interesting feature of the ISO4-H5 genome involved in purine degradation and carbon fixation. DeoA recycles AMP to adenine and ribose-1,5-bisphosphate, which can be isomerized to ribulose-1,5-bisphosphate, which is supplied to RubisCO for CO_2 fixation as well as re-entry into the pentose phosphate pathway via 3-phospho-D-glycerate (Watson *et al.* 1999; Sato *et al.* 2007). ISO4-H5 lacks a detectable phosphenolpyruvate carboxylase and the hydrogen carbonate is likely not utilised for oxaloacetate production (Smith and Ferry 1999). This allows ISO4-H5 to recycle nucleotides as well as acquire an additional carbon source. However, the carbon fixation via RubisCO may not sustain the cell as the sole carbon source.

The archaeal phospholipids differentiate themselves from phospholipids from other domains of life, by being composed of isoprenoid chains and glycerol-1-phosphate, joined by ether linkages (Nishimura and Eguchi 2006). The predicted ISO4-H5 lipid biosynthesis pathway is illustrated in Figure 3.11, and shows that the biosynthesis of isoprenoids from the mevalonate pathway is similar to what was previously established in *Methanocaldococcus jannaschii* (Kuzuyama 2002; Grochowski *et al.* 2006). The ISO4-H5 genome is predicted to lack phosphomevalonate decarboxylase, MvaD, however it has been proposed that conversion of mevalonate-5-phosphate to isopentenyl phosphate could be carried out by an unidentified ATP-

independent decarboxylase (Vinokur *et al.* 2014). In methanogens such as *Methanothermobacter thermoautotrophicus* and *Mbb. ruminantium* M1^{T}, the CDP-archaeol proceeds towards production of archaetidylserine and archaetidylethanolamine (Morii *et al.* 2000; Morii and Koga 2003; Leahy *et al.* 2010), which ISO4-H5 is unable to produce due to lack of an archaetidylserine synthase and decarboxylase. Instead, ISO4-H5 is predicted to produce cardiolipin, a phospholipid identified within the inner mitochondrial membrane and in bacterial cell envelope (Schlame 2008). It has been proposed to function as a proton trap by bicyclic resonance between phosphates and a central hydroxyl group, capable of protein binding with chaperone-like activity and stabilising the curved region of a phospholipid membrane (Haines and Dencher 2002; Schlame and Ren 2009). The properties of cardiolipin may be beneficial to the survival of ISO4-H5 within the rumen and contribute to its unusual cell morphology. ISO4-H5 is devoid of cell wall protection, therefore the presence of cardiolipin may make the ISO4-H5 cell envelope more flexible, however it remains physically fragile. It is predicted that from exposure to physical stresses, methanogens equipped with a pseudomurein and S-layer-containing cell envelopes will have a better chance of survival than ISO4-H5-like organisms.

In methanogens, the cell envelope often consists of pseudomurein polysaccharides enclosed in a S-layer (Arbing *et al.* 2012). The absence of UDP-*N*-acetylglucosamine diphosphorylase, glucosamine-1-phosphate *N*-acetyltransferase and glucosamine-fructose-6-phosphate aminotransferase suggests the ISO4-H5 genome does not harbour sufficient genes to produce a functional pseudomurein. The S-layer of *Aneurinibacillus thermoaerophilus* DSM 10155 is found to be composed of crosslinking rhamnose and glycerol-manno-heptose (Wugeditsch *et al.* 1999). ISO4-H5 is predicted to encode a dTDP-L-rhamnose biosynthesis (AR505_0552 – 0554) gene cluster and glycerol-manno-heptose biosynthesis genes (AR505_1754, AR505_1770 – 1772), which suggests ISO4-H5 may be capable of producing an S-layer-like extracellular structure.

The predicted extracellular proteins of ISO4-H5 were analysed to gain insight into the interaction of ISO4-H5 with the rumen environment. Proteins require a signal sequence in order for a cell to transport the protein to the cell envelope or to export the protein outside the cell (Eichler and Moll 2001). ISO4-H5 is predicted to use the signal recognition particle (SRP) and Sec61 translocation complex for translocation of gene products (Zwieb and Bhuiyan 2010). There are 101 ISO4-H5 genes predicted to encode an N-terminal signal peptide (Figure 3.4.), 35 proteins are predicted to have two or more transmembrane domain, and an additional 12

proteins are predicted to be anchored through a C-terminal transmembrane domain. These genes include 10 ABC transporter family proteins, four transporters involved in drug resistance, a mechanosensitive ion channel (AR505_0163) and a Na^+/H^+ antiporter (AR505_0385). There are 39 proteins predicted to be anchored through an N-terminal transmembrane domain. This includes four ABC transporter family proteins, two of which are classified as regulatory genes; a carbon starvation protein, CstA (PF02554) (Schultz and Matin 1991), anchored with 12 transmembrane domains and a secreted sugar fermentation stimulation protein, SfsA (PF03749) (Kawamukai *et al.* 1991). The CstA may sense the environment and regulate ISO4-H5 carbon metabolism. It is unclear how a DNA-binding SfsA would function extracellularly. There are 45 adhesin-like proteins predicted to be surface-located proteins. In the *Methanobrevibacter smithii* genome, adhesin-like proteins were associated with host mucosal glycosaminoglycans (Samuel *et al.* 2007). Also, physical co-localization of *Methanobrevibacter ruminantium* $M1^T$ with *Butyrivibrio proteoclasticus* has been previously correlated with up-regulation of adhesin-like proteins (Leahy *et al.* 2010; Ng *et al.* 2016). Domain repeats within adhesins are often involved in binding (Cabanes *et al.* 2002). ISO4-H5 encodes 19 adhesin-like proteins predicted to harbour domain repeats (Table 3.10.). One adhesin-like protein is predicted to be involved in binding to a dockerin of a cellulosome (PF00963) (Shimon *et al.* 1997). Cellulosomes are cell surface protrusions produced in some cellulose hydrolysing microbes (Lamed *et al.* 1987), which is also present in the rumen (Oyeleke and Okusanmi 2008), including the H_2 producing *Ruminococcus flavefaciens* (Kirby *et al.* 1997; Ding *et al.* 2001; Jindou *et al.* 2008). The binding of ISO4-H5 to cellulosome could assist in acquiring H_2 required for methanogenesis. A particular type of high molecular weight adhesins have been identified in *Haemophilus influenza* with a role in the attachment to human epithelial cells (Barenkamp 1996; Barenkamp and St Geme 1996). ISO4-H5 contains seven adhesin-like proteins larger than 200kDa and it is possible these genes could facilitate interaction of ISO4-H5 with its ovine host or with other rumen microorganisms. ISO4-H5 was originally enriched from the ovine rumen with the bacterium *Succinivibrio dextrinosolvens*, and it has been demonstrated recently that there is a strong association between the order Methanomassiliicoccales and *Succinivibrio* spp. (Henderson *et al.* 2015). Whether or not ISO4-H5 uses its adhesins to associate with particular rumen surfaces or microbes remains unclear.

ISO4-H5 is predicted to utilise ammonium as a nitrogen source via either GdhA or GlnA, the ammonium may be acquired either externally by an ammonium transporter or internally for methylamine methanogenesis. The ammonium transporter may also export ammonium when

cellular ammonium is in excess. Glutamate and glutamine are central to nitrogen metabolism as well as representing the protein building blocks for the biosynthesis of other amino acids. ISO4-H5 utilises GdhA for interconversion between 2-oxoglutarate and glutamate with ammonia and $NAD(P)^+$, as well as type-III GlnA for ATP driven interconversion of glutamate to glutamine (Helling 1998) (Table 3.11.). The predicted nitrogenase and nitroreductase genes along with molybdenum cofactor biosynthesis genes suggests ISO4-H5 may be capable of nitrogen fixation (Leigh 2000). However nitrogen fixation ability cannot be concluded based on the genes alone, as it has been shown in *Methanocaldococcus jannaschii* that these genes do not necessarily confer nitrogen fixation activity (Staples *et al.* 2007).

Nicotinamide adenine dinucleotide (NAD) is an electron carrier involved in redox reactions (Pollak *et al.* 2007). It is an important coenzyme in methanogens, as demonstrated by *Mbb. ruminantium* $M1^T$, NADP are used to relay reducing potential from ethanol to F_{420} (Leahy *et al.* 2010). ISO4-H5 does not possess the full pathway for *de novo* NAD biosynthesis, as the the *nadB* and *nadD* genes are absent. The L-aspartate dehydrogenase AspDH has been reported to be able to replace NadB (Yang *et al.* 2003), the L-aspartate dehydrogenase gene is predicted in the ISO4-H5, which may function similary in ISO4-H5. ISO4-H5 is predicted to carry the full set of genes required to salvage nicotinamide and nicotinic acid for NAD biosynthesis, so it is plausible that NAD can be synthesised in ISO4-H5.

3.4. Conclusions

The genome of ISO4-H5 is similar to those of other members of the order Methanomassiliicoccales, Mx1201, B10 and BRNA1. ISO4-H5 encodes the key genes and pathways required for H_2-dependent methylotrophic methanogenesis by reduction of methyl substrates, without the ability to oxidize methyl substrates to CO_2. The wide range of methyl substrates predicted to be used by ISO4-H5 suggests it is more metabolically versatile than other methylotrophic methanogens within the rumen. Members of Methanomassiliicoccales co-exist in the rumen with *Methanosphaera* spp. (Janssen and Kirs 2008; Seedorf *et al.* 2015) and share similar substrate requirements. Methanomassiliicoccales are likely able to outcompete *Methanosphaera* in the rumen at low H_2 concentrations, due to the lower thresholds conferred by the low ATP gain. However, they are probably disadvantaged when substrate concentrations are high and the low ATP yield limits their ability to proliferate quickly. The variability of fermentation rates in the rumen associated with periods of feeding

or fasting is therefore expected to give both groups of methylotrophic methanogens opportunities to grow.

ISO4-H5 appears to be reliant on the Hdr, Mvh and Fpo-like complexes for electron bifurcation, membrane potential generation and energy conservation, as has been described in other members of Methanomassiliicoccales. However, ISO4-H5 is incapable of producing CoM, which suggests that ISO4-H5 has adapted to the rumen environment, where CoM produced by other methanogens would be able to supplement ISO4-H5. ISO4-H5 lacks the genes encoding cofactor F_{420} synthesis, rendering it non-fluorescent under illumination at 420 nm. This trait has also been reported amongst other members of Methanomassiliicoccales, and is likely one of the key characteristics of this particular order of methanogens. However, a culture of B10 has been reported to fluoresce (Arbing *et al.* 2012; Dridi *et al.* 2012; Borrel *et al.* 2013) and this may be consistent with B10 belonging to the deepest branching group within Methanomassiliicoccales (Seedorf *et al.* 2014)**.**

The use of pyrrolysine in proteins carrying out various cellular functions suggests it is important for ISO4-H5. While pyrrolysine is important in methylamine utilisation by all members of Methanomassiliicoccales sequenced thus far, pyrrolysine appears to play a role in methanol use by ISO4-H5, as the methanol:methyltransferase corrinoid protein, MtaC1, is predicted to contain a pyrrolysine in its full length protein. The use of pyrrolysine and the possession of a Fpo-like complex by ISO4-H5 adds further weight to the hypothesis that the order Methanomassiliicoccales is evolutionary closer to the order Methanosarcinales, which supports findings from a previous phylogenetic study (Janssen and Kirs 2008). By analyzing the genome of ISO4-H5, our knowledge of the order Methanomassiliicoccales has been expanded, and together with the genomes of other members of the Methanomassiliicoccales, this will be an important resource for the development of CH_4 abatement technologies in ruminants.

Chapter 4

Genomic insight into members of the Methanomassiliicoccales isolated from the rumen as revealed by comparative genome analysis

4.1. Introduction

Members of the order Methanomassiliicoccales are methylotrophic methanogens first detected in the rumen (Yanagita *et al.* 2000) and subsequently in a variety of other anaerobic environments, including landfill (Huang *et al.* 2002), marine hydrothermal vents (Takai and Horikoshi 1999), lake sediments (Nusslein *et al.* 2001), rice paddies (Conrad *et al.* 2006), humus-feeding larva (Egert *et al.* 2003), termite and millipede hindguts (Shinzato *et al.* 1999; Paul *et al.* 2012), the reticulo-rumen compartment of ruminant animals (Yanagita *et al.* 2000; Tajima *et al.* 2001), and in the human gut and subgingival plague (Dridi *et al.* 2012; Horz *et al.* 2012). Genomes of three members of Methanomassiliiccales isolated from human faeces have been sequenced, including *"Candidatus* Methanomassiliicoccus intestinalis Mx1-Issoire" (Mx1), "*Candidatus* Methanomethylophilus alvus Mx1201" (Mx1201) and *Methanomassiliicoccus luminyensis* B10 (B10) (Borrel *et al.* 2012; Dridi *et al.* 2012; Borrel *et al.* 2013). In addition, the genome sequence of "*Candidatus* Methanoplasma termitum MpT1" (MpT1) enriched from the termite gut has been reported (Lang *et al.* 2015). Currently, four genome sequences are publically available for bovine rumen isolates, they are Thermoplasmatales archaeon BRNA1 (BRNA1) (NCBI Refeence Sequence NC_020892.1), "*Candidatus* Methylophilus" sp. 1R26 (1R26) (Noel *et al.* 2016), Methanomassiliicoccales archaeon RumEn M1 and RumEn M2 (Sollinger *et al.* 2016). Comparative genome analysis of these reported genome sequences has highlighted the lack of genes encoding the hydrogenotrophic methanogenesis pathway, but confirmed the presence of genes required for methylotrophic methanogenesis via direct reduction of methyl-substrates (Borrel *et al.* 2014). A recent comparative study which included MpT1 from the termite gut has confirmed the utilisation of pyrrolysine and the lack of an rRNA operon structure, this has expanded our understanding of energy metabolism by predicting that the Fpo-complex is responsible for membrane potential generation in the order Methanomassiliicoccales (Lang *et al.* 2015). Currently, the Methanomassiliicoccales can be considered as two clades, with one that harbors more gastrointestinal (GIT) isolates and one that harbors more environmental isolates; but neither clades are exclusive to GIT or environmental isolates (Sollinger *et al.* 2016).

The methanogenic archaeon ISO4-H5, isolated from the ovine rumen, has been sequenced as part of this thesis (see Chapter 3). Furthermore, two strains of the Methanomassiliicoccales order, ISO4-G1 and ISO4-G11, also isolated from the ovine rumen (Jeyanathan 2010) have been sequenced as part of the NZAGRC and PGGRC programmes of research and made available for comparative analysis in this thesis study (Kelly *et al.* 2016). Combined with the genome sequence of Thermoplasmataleas archaeon BRNA1, Methanomassiliicoccales archaeon RumEn M1, RumEn M2 and "*Candidatus* Methylophilus" sp. 1R26 (Noel *et al.* 2016; Sollinger *et al.* 2016), this allows for a detailed investigation into the genomes of members of the rumen Methanomassiliicoccales. In this chapter, I have compared the genome sequences of the rumen Methanomassiliicoccales with those from other environments. I have focused my analysis on gene conservation to gain insight into the metabolism and physiology of these organisms, and have identified specific sets of genes which are important for adaptation to the rumen environment.

4.2. Results

4.2.1. Phylogenetic relationship of members of the Methanomassiliicoccales isolated from gut environments

A FGD tree (Figure 4.1A.) was created to assess the similarity between organisms based on the full complement of ORFs within each genome. Based on this analysis, ISO4-H5 was most closely related to Mx1201, then ISO4-G11 and BRNA1. ISO4-H5 is more distantly related to ISO4-G1 and MpT1 strains. Mx1 and B10 fell within a completely different branch, similar to the 16*S* rRNA gene based tree (Figure 4.1B.). The environmental isolates and gastrointestinal isolates tend to cluster on different branches of the 16*S* rRNA gene-based tree.

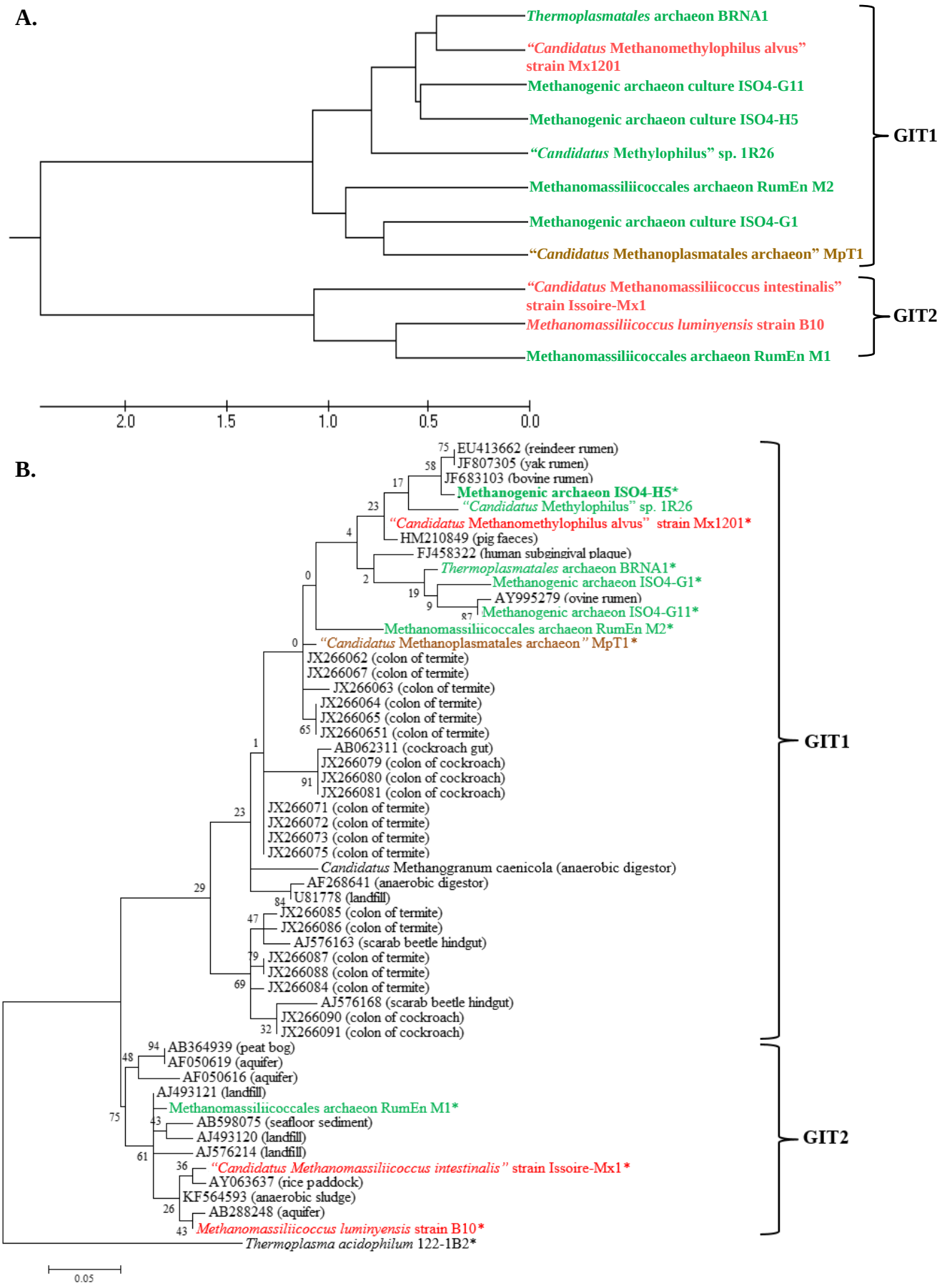

A.
Thermoplasmatales archaeon BRNA1
"Candidatus Methanomethylophilus alvus" strain Mx1201
Methanogenic archaeon culture ISO4-G11
Methanogenic archaeon culture ISO4-H5
"Candidatus Methylophilus" sp. 1R26
Methanomassiliicoccales archaeon RumEn M2
Methanogenic archaeon culture ISO4-G1
"Candidatus Methanoplasmatales archaeon" MpT1
"Candidatus Methanomassiliicoccus intestinalis" strain Issoire-Mx1
Methanomassiliicoccus luminyensis strain B10
Methanomassiliicoccales archaeon RumEn M1
GIT1
GIT2
2.0
1.5
1.0
0.5
0.0
B.
EU413662 (reindeer rumen)
JF807305 (yak rumen)
JF683103 (bovine rumen)
Methanogenic archaeon ISO4-H5*
"Candidatus Methylophilus" sp. 1R26
"Candidatus Methanomethylophilus alvus" strain Mx1201*
HM210849 (pig faeces)
FJ458322 (human subgingival plaque)
Thermoplasmatales archaeon BRNA1*
Methanogenic archaeon ISO4-G1*
AY995279 (ovine rumen)
Methanogenic archaeon ISO4-G11*
Methanomassiliicoccales archaeon RumEn M2*
"Candidatus Methanoplasmatales archaeon" MpT1*
JX266062 (colon of termite)
JX266067 (colon of termite)
JX266063 (colon of termite)
JX266064 (colon of termite)
JX266065 (colon of termite)
JX2660651 (colon of termite)
AB062311 (cockroach gut)
JX266079 (colon of cockroach)
JX266080 (colon of cockroach)
JX266081 (colon of cockroach)
JX266071 (colon of termite)
JX266072 (colon of termite)
JX266073 (colon of termite)
JX266075 (colon of termite)
Candidatus Methanogranum caenicola (anaerobic digestor)
AF268641 (anaerobic digestor)
U81778 (landfill)
JX266085 (colon of termite)
JX266086 (colon of termite)
AJ576163 (scarab beetle hindgut)
JX266087 (colon of termite)
JX266088 (colon of termite)
JX266084 (colon of termite)
AJ576168 (scarab beetle hindgut)
JX266090 (colon of cockroach)
JX266091 (colon of cockroach)
AB364939 (peat bog)
AF050619 (aquifer)
AF050616 (aquifer)
AJ493121 (landfill)
Methanomassiliicoccales archaeon RumEn M1*
AB598075 (seafloor sediment)
AJ493120 (landfill)
AJ576214 (landfill)
"Candidatus Methanomassiliicoccus intestinalis" strain Issoire-Mx1*
AY063637 (rice paddock)
KF564593 (anaerobic sludge)
AB288248 (aquifer)
Methanomassiliicoccus luminyensis strain B10*
Thermoplasma acidophilum 122-1B2*
GIT1
GIT2
0.05

Figure 4.1 Phylogenic relationship of members of the order Methanomassiliicoccales. **A.** Functional Genome Distribution tree of 11 members of the Methanomassiliicoccales with complete or draft genome sequences. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the tree. **B**. 16*S* rRNA gene-based phylogenetic tree showing the relationships of methanogenic archaeon ISO4-H5 (shown in bold print) relative to other type and non-type strains within the order Methanomassiliicoccales. * indicates strain with available genome sequences. The phylogeny was inferred from 16*S* rRNA gene nucleotide sequences (1,474 bp internal region) aligned using the Maximum Likelihood method based on the Kimura 2-parameter model. Bar: 0.05 substitutions per nucleotide position. The GenBank/nucleotide accession numbers of environmental sequences are displayed with the source habitat given in brackets. GIT1: Gastrointestinal cluster 1, GIT2: Gastrointestinal cluster 2. Strains whose genomes have been sequenced are marked with asterisk. Strains in red text are enriched or isolated from human faeces, strains in green text are enriched or isolated from rumen, and strain in brown text are enriched from termite gut. The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the MCL approach. All positions containing gaps or missing data were eliminated, giving a total of 455 positions in the final dataset. The 16S rRNA gene sequence from Thermoplasma acidophilum 122-1B2 was used as an outgroup.

4.2.2. Genome characteristics

The general genome characteristics of members of the Methanomassiliicoccales are listed in Table 4.1. All of the genomes are predicted to be single, circular chromosomes, with no evidence to support the presence of plasmids.

Ribosomal, non-coding and transfer RNAs. ISO4-H5 contains two sets of rRNA genes with an additional 5*S* rRNA gene, as described in Chapter 3 (Table 3.3.). Eight of the Methanomassiliicoccales genomes, including ISO4-G1 and ISO4-G11 have a single set of rRNA genes with an additional 5S rRNA gene, 1R26 has two additional 5*S* rRNA genes, each of the genes are scattered throughout the genome. MpT1 does not have an additional 5*S* rRNA gene (Table 4.1.). The Methanomassiliicoccales genomes analysed are predicted to carry seven ncRNAs (Table 4.2.). Two ncRNAs are conserved: an archaeal RNase P and an archaeal SRP (Table 4.2.). Both RNase P and SRP are ribozymes comprised of ncRNA and protein subunits. Two protein subunits of SRP: SRP54 and SRP19 are conserved (Table A.4.17.). The RNase P protein components are conserved across eight genomes with protein components 1 and 4, whereas 1R26, Mx1201 and B10 genomes are predicted to only have either protein component 1 or 4. Group II introns have been predicted in the Mx1201, ISO4-H5 and ISO4-G11 genomes. The genomes of ISO4-H5 (AR505_1776) and ISO4-G11 (ISO4G11_0442) encode a group II intron maturase. Two predicted group II introns in Mx1201 are located adjacent to an ATP-dependent RNA helicase (MMALV_11790), the other overlaps with a retron-type RNA-directed DNA polymerase (MMALV_06720), which is possibly encoded by the group II intron. The second predicted group II intron in the ISO4-H5 genome overlaps with a hypothetical protein (AR505_0861) without a corresponding maturase. There is only one snoRNA predicted in the ISO4-G1 genome, sR41.

Table 4.1. General genome properties of sequenced members of Methanomassiliicoccales

Property	**Methanogenic archaeon ISO4-H5**	**Methanogenic archaeon ISO4-G1**	**Methanogenic archaeon ISO4-G11**	***Thermoplasmatales* archaeon BRNA1**	**Methanomassiliicoccales archaeon RumEn M1**	**Methanomassiliicoccales archaeon RumEn M2**	***"Candidatus* Methylophilus" sp. 1R26**
Source	Ovine rumen	Ovine rumen	Ovine rumen	Bovine rumen	Bovine rumen	Bovine rumen	Bovine rumen
Culture status	Isolated	Enriched	Enriched	Enriched	Enriched	Enriched	Enriched
Project status	Complete	Complete	Draft (52 contigs)	Complete	Draft (182 contigs)	Draft (18 contigs)	Draft (50 contigs)
Genome size (bp)	1,937,882	1,593,504	1,901,999*	1,461,105	2,121,026*	1,280,797*	1,723,106*
G+C content (%)	54.0	55.5	56.9	58.3	62.1	54.6	60.4
Number of ORFs	1826	1527	1758	1577	1875	1175	1,465
Coding area (%)	90	92	88	91	68	73.3	65.9
rRNA (5*S*, 16*S*, 23*S*)	3, 2, 2	2,1,1	2,1,1	2, 1, 1	2, 1, 1	2, 1, 1	3, 1, 1
tRNAs (with introns)	47 (3)	47 (3)	47(3)*	47 (3)	46(8)	45(9)	49(3)
Non-coding RNA	8	3	5*	4	3*	2*	2*
IS	41	7	29*	7	Nd	Nd	Nd
Prophage	0	0	1	0	0	0	0
CRISPR regions (spacers)	1 (32)	0	0	1 (6)	1 (6)	2 (3, 5)	1 (13)
Adhesin-like proteins	44	18	26*	17			
NCBI accession	CP014214	CP013703	N/A	CP002916.1	LJKK00000000	LJKL00000000	LOPS00000000
Publication	Li *et al.*, 2016	Kelly *et al.*, 2016			Sollinger *et al.*, 2016	Sollinger *et al.*, 2016	Noel *et al.*, 2016

Property	***"Candidatus Methanomassiliicoccus* intestinalis Mx1-Issoire"**	***"Candidatus* Methanomethylophilus alvus Mx1201"**	***Methanomassiliicoccus luminyensis* B10**	***"Candidatus* Methanoplasma termitum MpT1"**
Source	Human faeces	Human faeces	Human faeces	Termite gut
Culture status	Enriched	Enriched	Isolated	Enriched
Project status	Complete	Complete	Draft (26 contigs)	Complete
Genome size (bp)	1,931,651	1,666,795	2,620,233*	1,488,669
G+C content (%)	41.3	55.6	60.5	49.2
Number of ORFs	1876	1651	2669	1,415
Coding area (%)	88	89	86	91
rRNA (5*S*, 16*S*, 23*S*)	2, 1, 1	2, 1, 1	2, 1, 1	1,1,1
tRNAs (with introns)	46 (7)	47 (4)	47 (8)*	46 (4)
Non-coding RNA	3	7	2*	3
IS	16	19	Nd	19
Prophage	0	0	0	0
CRISPR regions (spacers)	1 (110)	2 (47, 12)	1 (113)	1 (53)
Adhesin-like proteins	Nd	Nd	Nd	Nd
NCBI accession	CP005934.1	CP004049.1	CAJE01000001 - 26	CP010070.1
Publication	Borrel *et al.*, 2013	Borrel *et al.*, 2012	Gorlas *et al.*, 2012	Lang *et al.*, 2015

*among currently available draft sequences

Table 4.2. Non-coding RNAs in Methanomassiliicoccales genomes

	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Archaeal RNaseP (protein components)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1, 4)	+ (1)	+ (4)	+ (1, 4)	+ (4)	+ (1, 4)
Archaeal Signal recognition particle	+	+	+	+	+	+	+	+	+	+	+
Small nucleolar RNA sR41	-	+	-	-	-	-	-	-	-	-	-
Group II catalytic intron	+	-	+	-	-	-	-	+	-	-	-
Group II catalytic intron D1-D4-1	-	-	-	-	-	-	-	+	-	-	-
Group II catalytic intron D1-D4-3	+	-	-	-	-	-	-	+	-	-	-
Group II catalytic intron D1-D4-7	+	-	+	-	-	-	-	-	-	-	-
TPP riboswitch	-	-	-	-	+	-	-	-	-	-	-

+ Indicate the presence of the ncRNA, - indicates absence of the ncRNA. (Number) indicates the number of protein subunits predicted to be associated with the ncRNA.

Ten Methanomassiliicoccales genomes analysed are predicted to contain a tRNA gene for all 21 amino acids (Table 4.3.), the amber suppressor $tRNA^{Pyl}$ is not predicted in the RumEn M2 genome. Seven genomes are predicted to contain 47 tRNAs, four genomes are predicted to contain 45 to 49 tRNAs, this is due to the variation in copy numbers of $tRNA^{Leu}$, $tRNA^{Lys}$, $tRNA^{Pro}$ and $tRNA^{Thr}$ in the Mx1, RumEn M1, RumEn M2 and 1R26 genome. The numbers of tRNAs corresponding to each amino acid is highly conserved between the Methanomassiliicoccales genomes, with some tRNAs in high numbers and some tRNA genes in low numbers, as shown in Table 4.3. Introns have been found consistently within $tRNA^{Trp}$ and $tRNA^{Tyr}$.

Table 4.3. Predicted tRNAs of Methanomassiliicoccales

tRNA corresponding Amino acid	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Alanine	3	3	3	3	3	3	3	3	3	3	3
Arginine	5	5	5	5	5**	5**	5	5	6**	4*	5
Asparagine	1	1	1	1	1*	1*	1	1	1*	1*	1
Aspartate	1	1	1	1	1	1	1	1	1	2	1
Cysteine	1	1	1	1	1*	1*	1	1	1*	1*	1
Glutamine	2	2	2	2	2	2	2	2	2	2	2
Glutamate	2	2	2	2	2	2	2	2	2	3*	2
Glycine	3	3	3	3	3	3	3	3	3	3	3
Histidine	1	1	1	1	1	1	1	1	1	1	1
Isoleucine	1	1	1	1	1	1	1	1	1	2	1
Leucine	5	5	5	5	5	5	6	5*	4	4	5*
Lysine	2	2	2	2	1	1	2	2	1	1	2
Methionine	3*	3*	3*	3*	3**	3**	3*	3*	3	2	3*
Phenylalanine	1	1	1	1	1	1	1	1	1	1	1
Proline	3	3	3	3	3	3	2	3	3	3	3
Pyrrolysine	1	1	1	1	1	0	1	1	1	1	1
Serine	4	4	4	4	4*	4*	4	4	4*	4*	4
Threonine	3	3	3	3	3	3	4	3	3	3	3
Tryptophan	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
Tyrosine	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
Valine	3	3	3	3	3	3	3	3	3	3	3
Pseudo	1	0	0	0	0	0	1	0	0	1*	0
Total	47	47	47	47	46	45	49	47	46	47	47
Intron-containing tRNAs	3	3	3	3	8	9	3	4	7	8	4

*intron containing tRNAs. ** two tRNAs that contain one intron each

Codon and amino acid usage. The codon usage for each genome is summarised in Table A.4.1, and visualized in Figure 4.2. The overall codon usage pattern is similar between ISO4-G1, Mx1201, ISO4-H5, ISO4-G11, BRNA1, 1R26, RumEn M1, RumEn M2 and B10 (Figure 4.2.), while Mx1 and MpT1 displayed different patterns. Mx1 codon usage is likely a result of its low G+C content of 41.3% as reflected in the codon usage for the amino acids phenylalanine, asparagine, lysine, glutamate, tyrosine, alanine and cysteine, where Mx1 preferentially uses codons with a lower G or C content (Table A.4.1.). RumEn M1 has the highest G+C content (62.1%) of the genomes analysed, and GGC and GGG are preferred over GGA and GGU to encode glycine, GUG and GUC are preferred over GUU and GUA to encode valine (Table A.4.1.). No correlation of G+C content and codon preference were observed in MpT1 (Table A.4.1.).

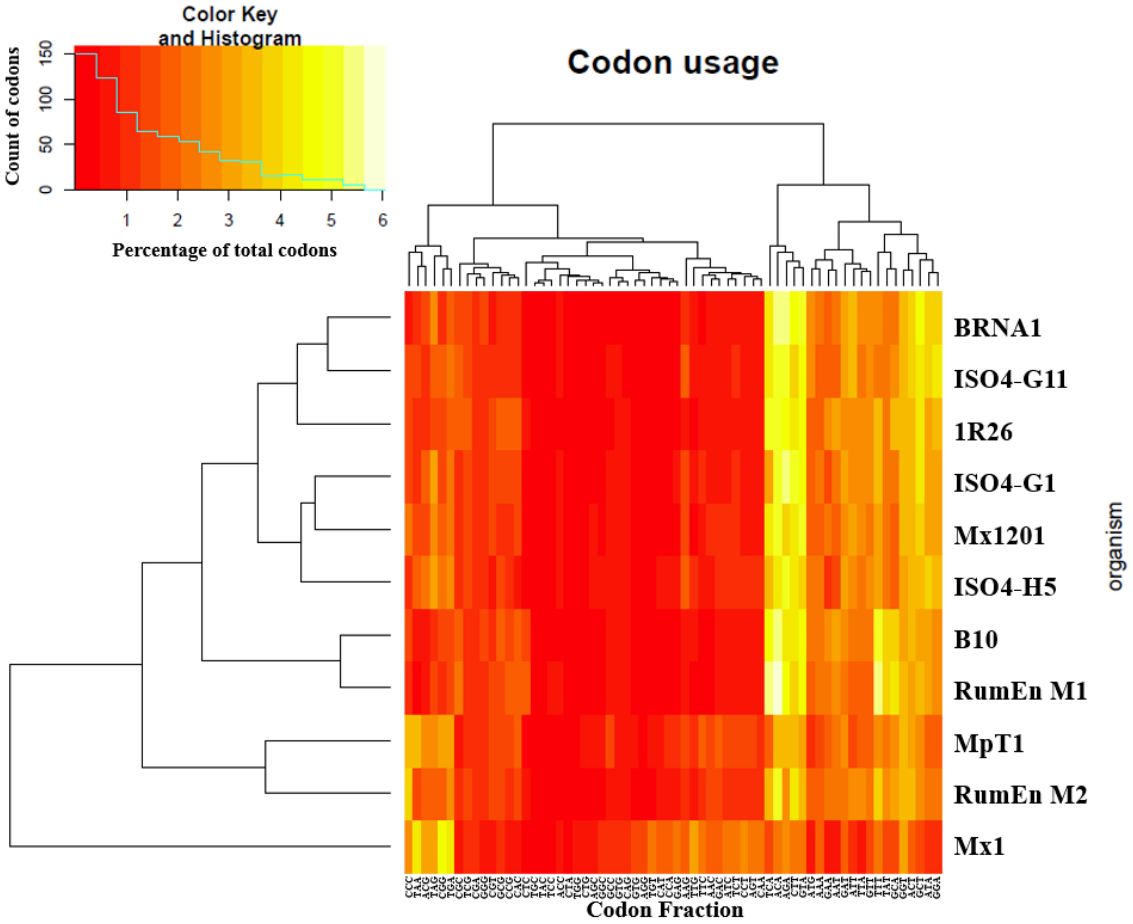


Figure 4.2. Codon usage heatmap of the Methanomassiliicoccales genomes analysed. The color key on the top left hand corner, indicates codons with the lowest percentage codons in red, and most abundant codons are displayed in pale yellow, while the histogram displays the count of codons in 0.37% intervals across genomes.

Translational start and stop codons appear to differ among the genomes investigated. B10 has the lowest usage of methionine as a translational initiator, and the highest usage of valine as a translational initiator, whereas BRNA1 has the highest usage of leucine as a translational initiator (Table A.4.1.). All of the genomes analysed have a high preference for using the opal codon as a stop codon, with the exception of Mx1 genome which uses the opal codon and ochre codon at similar levels (Table A.4.1.). The amber codon is associated with pyrrolysine incorporation in all genomes examined.

Despite the differences in codon usage, the genomes investigated are rather conserved in their amino acid usage pattern (Figure 4.3.). All 11 genomes analysed have a high percentage of branched-chain amino acids, glycine, alanine and serine, and uses 4% to 7% of charged amino acids glutamate, lysine, arginine and aspartate. The least used amino acid is tryptophan for all the genomes analysed (Figure 4.3.). The use of cysteine and histidine is also quite low in all 11

genomes. These three low frequency amino acids only have one single tRNA associated with each of them.

CRISPR elements. The CRISPR elements of 11 Methanomassiliicoccales genomes were predicted and summarised in Table 4.1, the CRISPR associated genes were summarised in Table A.4.2. Based on the CRISPR associated genes, the ISO4-H5 genome may possess a type I-E CRISPR element, eight other genomes were predicted with *cas3* gene homologue, which suggests they may also be type I CRISPR element, however, the subtypes cannot be determined due to lack of other associated gene homologues.

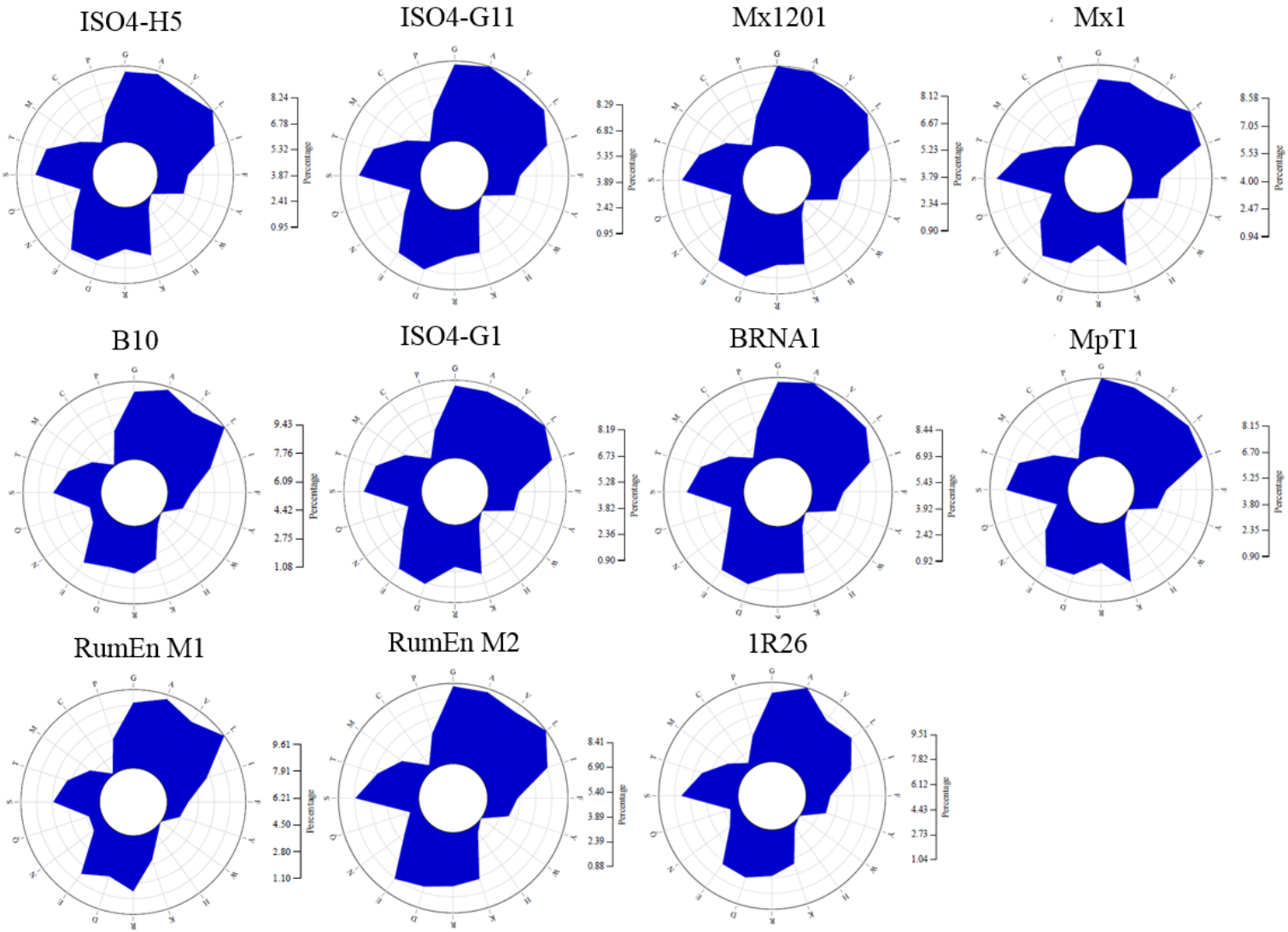


Figure 4.3. Amino acid usage map of the Methanomassiliicoccales genomes analysed. Pyrrolysine is not included in this figure. The amino acid usage is displayed as percentages, scales are different for every genome and are displayed to the right of each diagram.

4.2.3. Pyrrolysine usage

All of the genes required for pyrrolysine biosynthesis and incorporation are found among ten of the Methanomassiliicoccales genomes and the structure of the pyrrolysine operon is conserved (Figure 4.4.), only the incomplete RumEn M2 genome lacks the pyrrolysine biosynthesis genes. The B10 genome contains two pyrrolysine operons located in two different contigs. The pyrrolysine operon is located adjacent to *mtmB, mtbB, mttB* and other genes involved in methylamine utilisation, including corrinoid proteins associated with each methyltransferase *mtmC, mtbC, mttC*, methylamine permeases *mtbP* and *mttP* and methyltransferase corrinoid activation protein, *ramA* (Figure 4.4.). The genes containing in-frame amber codons were examined to predict pyrrolysine incorporation in ISO4-H5, ISO4-G11, ISO4-G1, BRNA1, RumEn M1, RumEn M2, 1R26 and MpT1, and were classified into three categories: Class 1 genes which have amber codon read-through and predicted incorporation of pyrrolysine; Class 2 genes that were predicted to use the amber codon as a stop codon; Class 3 genes with uncertain amber codon usage due to lack of homologous genes. The amber codon usage and putative Pyl-containing proteins in the Methanomassiliicoccales genomes analysed are listed in Table 4.4 (also see Tables A.4.1, A.4.3 – A.4.9) and are displayed in Figure 4.4.

The Methanomassiliicoccales genomes fall into two different patterns of pyrrolysine utilisation. The genomes of B10, Mx1, RumEn M1, RumEn M2 and MpT1 are predicted to contain very few proteins incorporating pyrrolysine outside of MtmB, MtbB and MttB, whereas ISO4-H5, ISO4-G1, ISO4-G11, Mx1201, 1R26 and BRNA1 are predicted to incorporate pyrrolysine at a relatively higher frequency (Table 4.4., Figure 4.5.). The genomes of B10 and Mx1 both have a high number of genes containing an in-frame amber codon, however other than the MtmB/MtbB/MttB genes, only a putative Fe-S binding protein (AGY50215), is predicted to incorporate pyrrolysine, with the remainder predicted to use the amber codon as a translational stop. The remaining Methanomassiliicoccales genomes have a relatively high number of genes predicted to use pyrrolysine and a low number of genes predicted to contain an in-frame amber codon.

Table 4.4. Amber codon utilisation in Methanomassiliicoccales

Genomes	CDS Class 1 (%)*	CDS Class 2 (%)	CDS Class 3 (%)	Total
ISO4-H5	25 (1.4)	10 (0.6)	11 (0.6)	46 (2.5)
ISO4-G1	28 (1.9)	6 (0.4)	9 (0.6)	43 (2.9)
ISO4-G11	20 (1.1)	5 (0.3)	9 (0.5)	34 (2.0)
BRNA1	25 (1.7)	6 (0.4)	1 (0.1)	32 (2.1)
RumEn M1	2 (0.1)	252 (11.1)	41 (1.8)	295 (13.0)
RumEn M2	1 (0.1)	90 (6.6)	22 (1.6)	113 (8.2)
1R26	10 (0.6)	37 (2.1)	27 (1.5)	74 (4.1)
Mx1201	20 (1.2)	26 (1.6)	Nd	46 (2.8)
Mx1	5 (0.3)	92 (5.0)	Nd	97 (5.3)
B10	7 (0.3)	293 (11.0)	Nd	300 (11.3)
MpT1	4 (0.3)	13 (0.9)	8 (0.6)	25 (1.8)

*Number of coding DNA sequence (CDS) after read-through is merged, including pseudogenes. Nd: not determined in this study.

MpT1 has a very low number of genes containing in-frame amber codons. Only two proteins were encoded by these genes, other than the MtmB/MtbB/MttB are predicted to incorporate pyrrolysine (Table A.4.5.). Additionally, the amber-codon containing genes are either hypothetical proteins, mobile elements or genes with a very short distance between the amber codon and the next stop codon.

Figure 4.4. Gene organisation of the pyrrolysine operon in Methanomassiliicoccales genomes and their co-localisation with methylamine utilisation genes. The pyrrolysine operon is shaded in yellow and includes: Amber suppressor $tRNA^{CUA}$ (*pylT*). Pyrrolysine-tRNA ligase (*pylS*). Methylornithine synthase, *pylB*; (2*R*,3*R*)-3-methylornithyl-N^6-lysine synthase (*pylC*) and pyrrolysine synthase (*PylD*). Methylamine utilisation genes include: Monomethylamine:corrinoid methyltransferase (*mtmB*), dimethylamine:corrinoid methyltransferase (*mtbB*), trimethylamine:corrinoid methyltransferase (*mttB*), monomethylamine methyltransferase corrinoid protein (*mtmC*), dimethylamine methyltransferase corrinoid protein (*mtbC*), trimethylamine methyltransferase corrinoid protein (*mttC*), dimethylamine permease (*mtbP*), trimethylamine permease (*mttP*), methyltransferase corrinoid activation protein (*ramA*). Genes with dashed border were pseudogenes. Gene sizes are not drawn to scale.

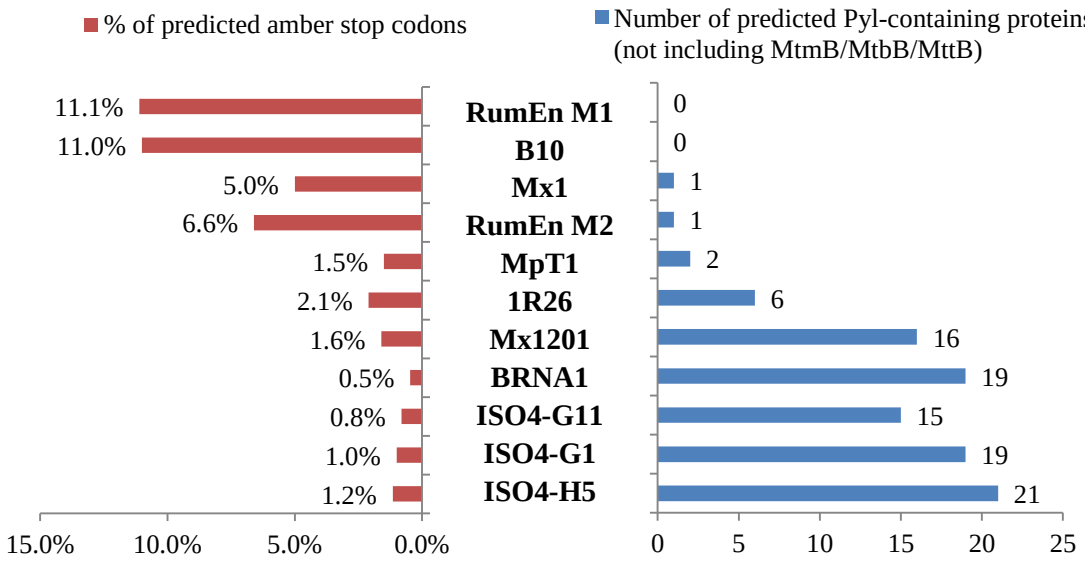


Figure 4.5. Comparison of predicted pyrrolysine containing proteins (excluding MtmB/MtbB/MttB) to the percentage of predicted amber stop codons.

4.2.4. Secretome

The predicted secretome (surface associated and/or released from the cell) proteins of the 11 genomes are summarised in Table 4.5 and Table A.4.6. The genomes analysed are predicted to devote between 2.5% and 5.6% of their ORFs to the export of extracellular and surface-associated proteins. A large proportion of the secretomes examined are predicted to have ORFs with motifs or domains for attachment to the cell surface. A range of between 16 to 75 genes per genome are predicted to be membrane anchored by one to two transmembrane helices or by an N-terminal lipobox. Around four to 25 genes per genome are predicted to be either unattached (i.e. secreted) or associated with the cell wall using other (unknown) mechanisms. The extracellular proteins were analysed for the presence of repeat domains and periplasmic domains (the periplasmic domain belongs to the substrate binding subunit of ABC transporters) and the results are listed in Table A.4.10.

Table 4.5. Predicted secretome of Methanomassiliicoccales

Secretome	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Integral membrane protein	13	12	15	15	18	8	13	13	15	25	9
Lipobox	6	2	8	3	9	2	13	2	4	0	1
Two TMH Membrane anchor	22	14	17	12	12	6	10	10	31	23	29
C-terminal membrane anchor	11	6	7	4	3	1	3	6	3	13	2
N-terminal membrane anchor	35	27	36	21	24	7	28	10	35	39	17
Secreted	15	8	9	5	25	10	15	13	9	24	4
Total (% of total ORFs)	102 (5.6%)	69 (4.5%)	92 (5.2%)	60 (3.8%)	91 (4.0%)	34 (2.5%)	82 (4.6%)	54 (3.2%)	97 (5.2%)	124 (4.6%)	62 (4.4%)

Prediction E value cutoff 1 x 10^{-05}.

4.2.5. Synteny

ISO4-H5 shares a high degree of synteny and displays an X-shaped alignment with BRNA1 and Mx1201 (Figure 4.6A., 4.6E.). A comparison of the nucleotide identity across all sequenced genomes using ISO4-H5 as the reference (Figures 4.7., A.4.3F.) revealed two regions of 100% identity to the genome of ISO4-G11. Region 1 encompasses 30.3 kb and Region 2 encompasses 26.3 kbp in the ISO4-H5 genome, and the genes in both regions are summarised in Table 4.6.

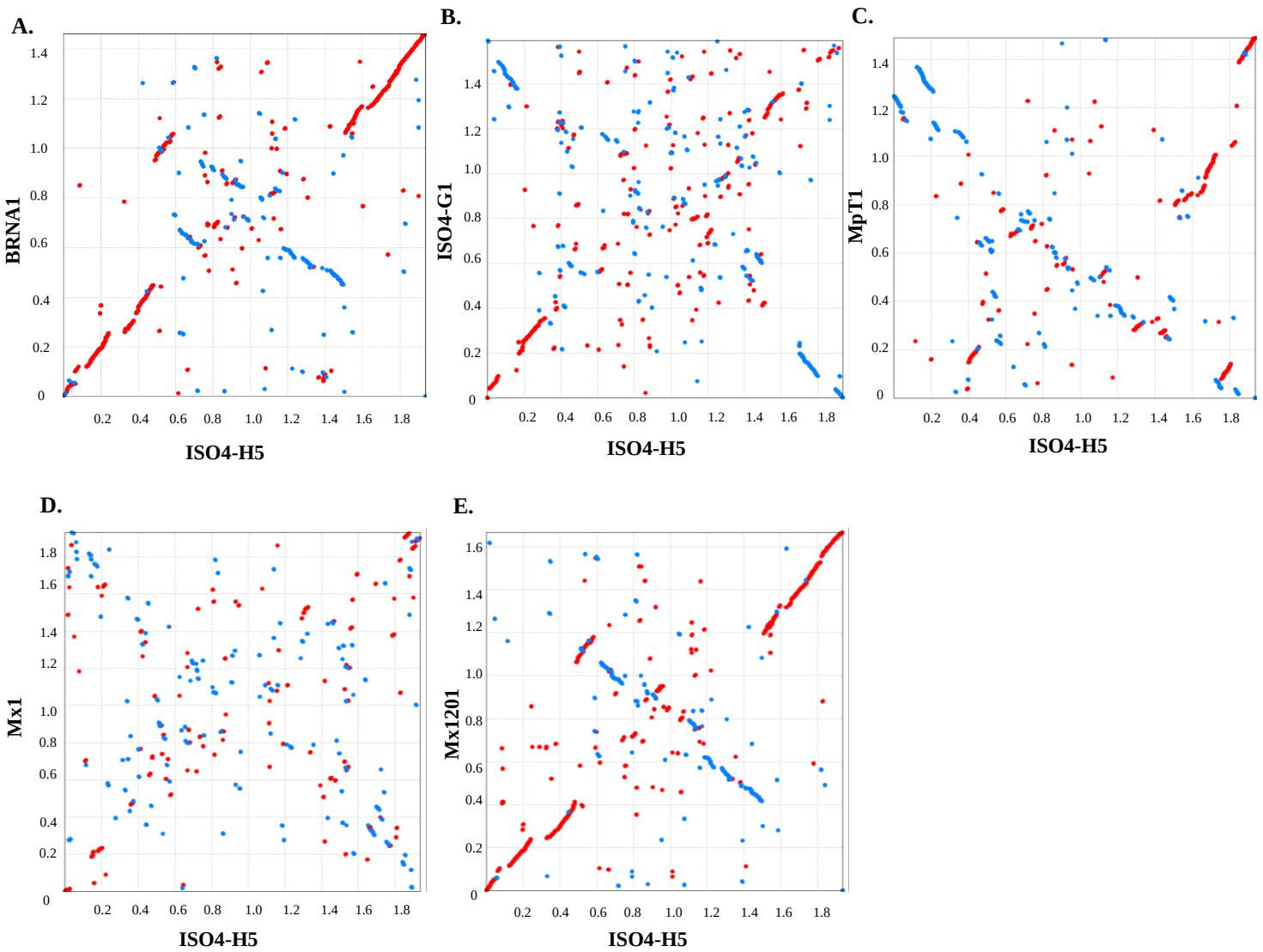


Figure 4.6. Gene synteny plots for completed genomes of Methanomassiliicoccales. PROmer alignments of the ISO4-H5 genome against genomes of other Methanomassiliicocaccales are shown. The alignments were plotted using MUMmer (Delcher *et al.* 2003) with forward matches shown in red and reverse matches in blue. The units displayed on both axes are in Mb.

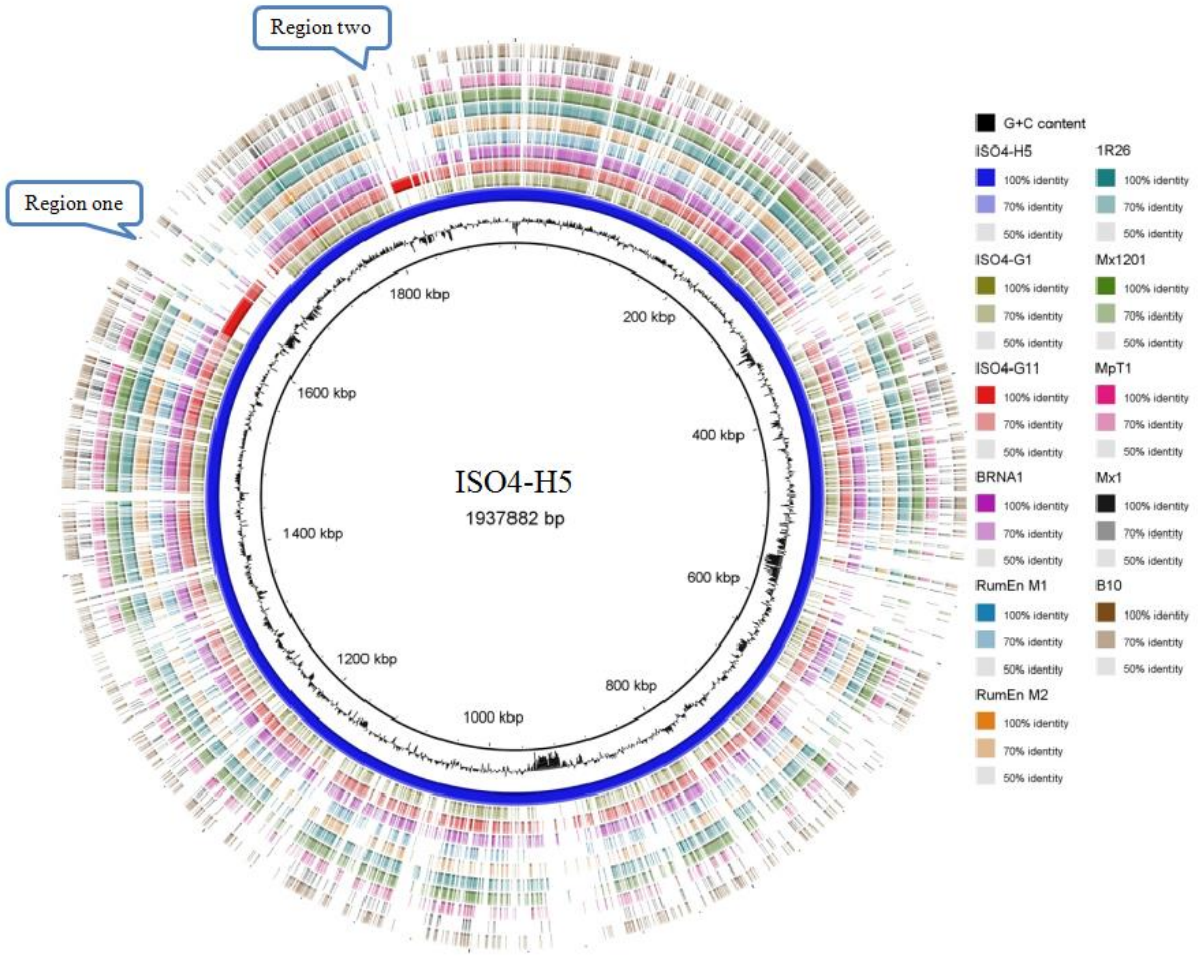


Figure 4.7. Genome map comparison of members of the order Methanomassiliicoccales. Methanomassiliicoccales genomes are plotted using ISO4-H5 as a reference. The colors in each genome ring accord with the identity illustrated in the legend. Circles are ordered 1 (innermost) to 10 (outermost). Circle 1: DNA base pairs of ISO4-H5. Circle 2: G+C content of ISO4-H5 genome. Circle 3: ISO4-H5 genome nucleotide sequence identity to itself. Circles 4 to 10: ISO4-H5 genome identity to ISO4-G1, ISO4-G11, BRNA1, RumEn M1, RumEn M2, 1R26, Mx1201, MpT1, Mx1, and B10, genomes respectively.

Table 4.6. Gene content of Regions 1 and 2 in ISO4-H5 and ISO4-G11 genomes

ISO4-H5	ISO4-G11	Predicted gene product
Region 1		
AR505_1521	ISO4G11_1250	Adhesin-like protein
AR505_1522	ISO4G11_1251	Hypothetical transmembrane protein
AR505_1523	ISO4G11_1252	ABC transporter substrate binding protein
AR505_1524	ISO4G11_1253	Adhesin-like protein
AR505_1525	ISO4G11_1254	ABC transporter permease protein
AR505_1526	ISO4G11_1255	ABC transporter ATP-binding protein
AR505_1527	ISO4G11_1256	Hypothetical protein
AR505_1528	ISO4G11_1257	Hypothetical protein
AR505_1529	ISO4G11_1258	Hypothetical protein
AR505_1530	ISO4G11_1259	Adhesin-like protein
AR505_1531	ISO4G11_1260	ATP-dependent DNA helicase
AR505_1532	ISO4G11_1261	Hypothetical transmembrane protein
AR505_1533	ISO4G11_1262	Adhesin-like protein
AR505_1534	ISO4G11_1263	Adhesin-like protein
AR505_1535	ISO4G11_1264	Hypothetical protein
AR505_1536	ISO4G11_1266	Hypothetical protein
AR505_1537	ISO4G11_1267	Hypothetical protein
Region 2		
AR505_1705	ISO4G11_1645	ABC transporter ATP-binding protein
AR505_1706	ISO4G11_1644	ABC transporter permease protein
AR505_1707	ISO4G11_1643	Adhesin-like protein
AR505_1708	ISO4G11_1642	ABC transporter substrate binding protein
AR505_1709	ISO4G11_1641	Hypothetical transmembrane protein
AR505_1710	ISO4G11_1640	Hypothetical transmembrane protein
AR505_1711	ISO4G11_1639	Adhesin-like protein
AR505_1712	ISO4G11_1638	Adhesin-like protein
AR505_1713	ISO4G11_1637	Adhesin-like protein
AR505_1714	ISO4G11_1636	Hypothetical protein
AR505_1715	ISO4G11_1635	Hypothetical protein
AR505_1716	ISO4G11_1634	Hypothetical protein
AR505_1717	ISO4G11_1633	Hypothetical protein
AR505_1718	ISO4G11_1632	Hypothetical protein

In order to identify similar regions between other members of Methanomassiliicoccales, each genome was used as reference genome and the nucleotide similarity was displayed (Figure A.4.3.). A less conserved, smaller genomic island of approximately 5 kb was detected between BRNA1 and ISO4-G11, however, the alignment ended prematurely as it was located at the end of a contig in the ISO4-G11 draft genome (Figure A.4.4.). The two conserved genes found in this region are a Type I site-specific deoxyribonuclease and a hypothetical protein.

4.2.6. Comparative analysis of gene families

The term core-genome was used to describe the genes or gene families present in all completed genomes analysed while the order level pan-genome was used to define the full complement of genes or gene families present in all sequenced genomes. When only completed genomes are analysed, the core genome consists of 415 gene families (Figure 4.10A.). When all genomes are analysed, 229 gene families are conserved between all genomes analysed (Figure 4.10C.) and the order level pan-genome consists of 8,767 gene families (Figure 4.8.). The inclusion of draft genomes into analysis revealed 78 gene families conserved between members of GIT1, and 226 gene families conserved between members of GIT2 (Figure A.4.5.). During this

analysis it was found that the incomplete genomes may confound the gene conservation analysis due to low level completion and poor quality sequence, the RumEn M1, RumEn M2 and 1R26 genomes are only 96.6%, 94.2% and 93% complete respectively, with 396, 194 and 315 predicted pseudogenes respectively, including some functionally important and conserved genes. For example the *mrtA* gene was marked as a pseudogene in the RumEn M1 genome, the *mrtG* gene was entirely absent from the 1R26 annotation due to a lack of a stop codon. To prevent false negatives introduced by the incomplete genomes, only the completed genomes were analysed to identify the core genome and environmental conserved gene families.

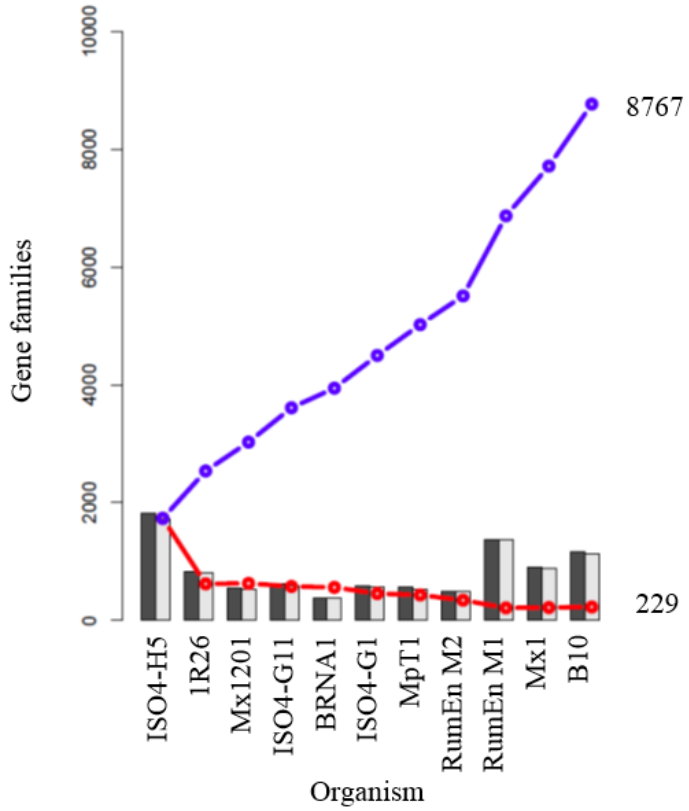


Figure 4.8. Order level pan-genome and core-genome plot of Methanomassiliicoccales genomes. The blue line indicates the cumulative order level pan-genome, and the red line indicates the core-genome. Dark grey bars indicate the number of new genes with the addition of each genome across the x-axis; the light grey bar indicates new gene families.

The pattern of gene family conservation (as illustrated in Figure 4.9) shows that six members of the Methanomassiliicoccales are more similar to each other, while B10, RumEn M1 and Mx1 are more similar to each other. The analysis shows that ISO4-H5 has the highest percentage of paralogs. The highest predicted gene family conservation is between ISO4-G11 and BRNA1 at 44%.

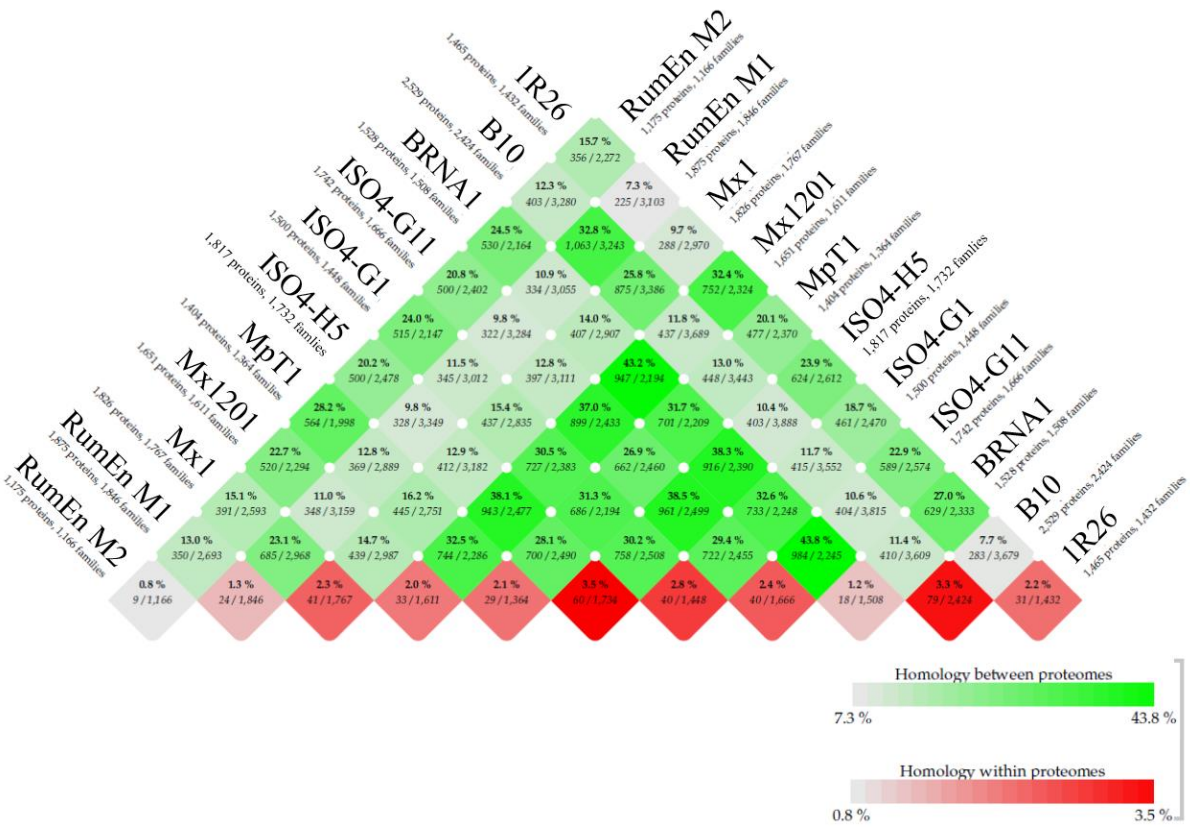
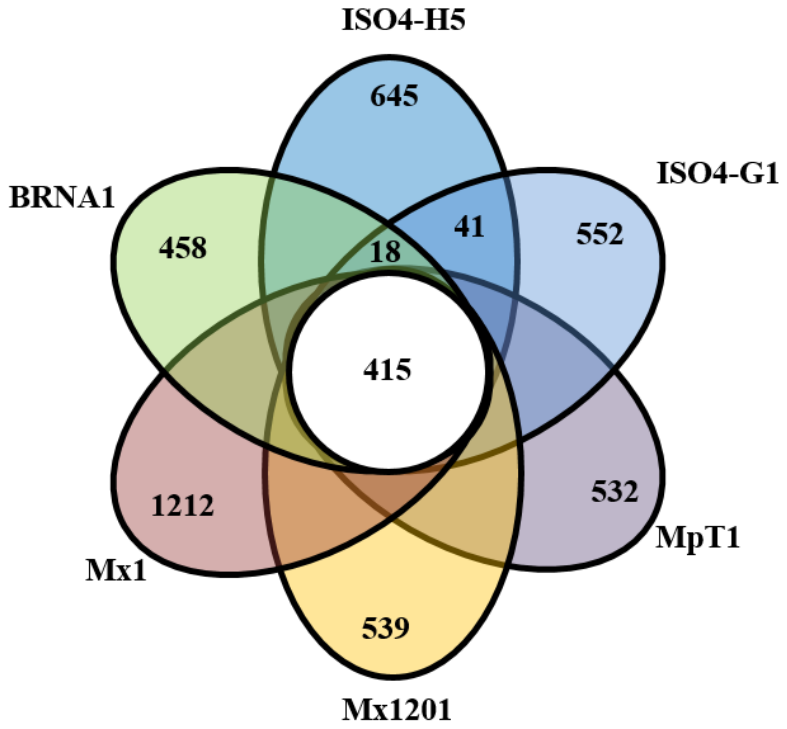


Figure 4.9. BLASTP matrix illustrating the number of conserved protein families between Methanomassiliicoccales genomes. Conservation is defined as 50% coverage and 50% identity. The color intensities are based on the relative percentage of conserved gene families; green depicts conserved protein families between genomes, red depicts protein homology within genome.

The number of gene families in the core-genome corresponds with the phylogenetic relationship between the organisms, as does the number of gene families specific to each genome. The number of gene families specific to RumEn M1, Mx1 and B10 are high, with 40%, 43% and 46.9% of their total gene families being specific to their genomes, respectively (Figure 4.10B.).

A.



B.

Genome	Gene families
ISO4-H5	1,732
ISO4-G1	1,448
ISO4-G11	1,666
1R26	1,432
BRNA1	1,508
Mx1201	1,611
MpT1	1,364
RumEn M1	1,846
RumEn M2	1,166
Mx1	1,767
B10	2,424

C.

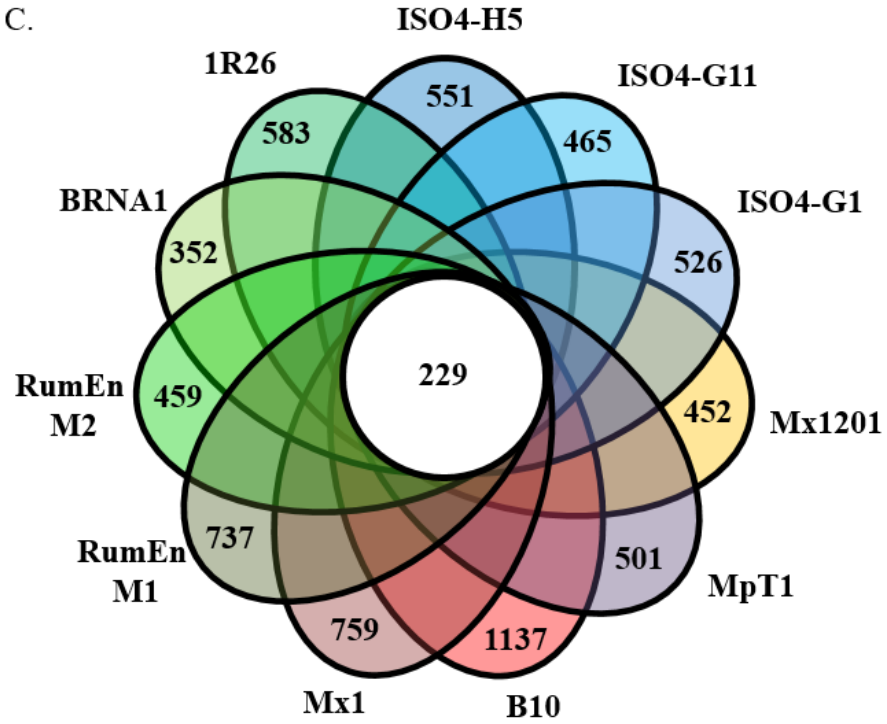


Figure 4.10. Conserved and novel gene families among the 11 Methanomassiliicoccales genomes. **A**. Venn diagram indicating the number of conserved gene families between completed genomes based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria. **B**. Table listing the total number of gene families in each of the Methanomassiliicoccales genomes. **C**. Venn diagram indicating the number of conserved gene families between all genomes analysed based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria.

Analysis of the Methanomassiliicoccales core genome. The number of gene families in the core genome is summarised in Figure 4.11 and Table 4.7. A full list of conserved gene families is presented in Appendix (Table A.4.11.). The core genome includes an operon of type III ribulose bisphosphate carboxylase (*rbcL*), ribose-1,5-bisphosphate isomerase (*e2b2*) and AMP phosphorylase (*deoA*) genes.

Table 4.7. ISO4-H5 conserved gene families and novel gene families classified by COG functional categories

COG	**Code**	**Gene families**				
		Total	**Core**	**Rumen**	**Ovine**	**Unique**
Translation	J	144	113	0	0	3
RNA processing and modification	A	1	0	0	0	0
Transcription	K	68	21	3	0	13
Replication, recombination and repair	L	123	17	1	0	38
Chromatin structure and dynamics	B	1	0	0	0	0
Cell cycle control, mitosis and meiosis	D	9	2	0	0	2
Nuclear structure	Y	0	0	0	0	0
Defense mechanisms	V	17	3	0	0	2
Signal transduction mechanisms	T	20	3	1	0	9
Cell wall/membrane biogenesis	M	27	2	0	0	11
Cell motility	N	2	0	0	0	2
Cytoskeleton	Z	0	0	0	0	0
Extracellular structures	W	0	0	0	0	0
Intracellular trafficking and secretion	U	13	6	0	0	0
Post-translational modification, protein turnover, chaperones	O	51	20	0	0	12
Energy production and conversion	C	89	34	1	4	4
Carbohydrate transport and metabolism	G	40	15	2	0	4
Amino acid transport and metabolism	E	99	34	0	2	7
Nucleotide transport and metabolism	F	46	29	1	2	1
Coenzyme transport and metabolism	H	104	31	0	4	8
Lipid transport and metabolism	I	18	9	1	1	6
Inorganic ion transport and metabolism	P	89	2	1	7	14
Secondary metabolites biosynthesis, transport and catabolism	Q	14	0	0	0	2
General function prediction only	R	241	37	1	4	73
Function unknown	S	126	24	0	3	24
Unclassified COG	Unclassified COG	91	6	2	3	40
Not in COG	-	393	3	4	11	276

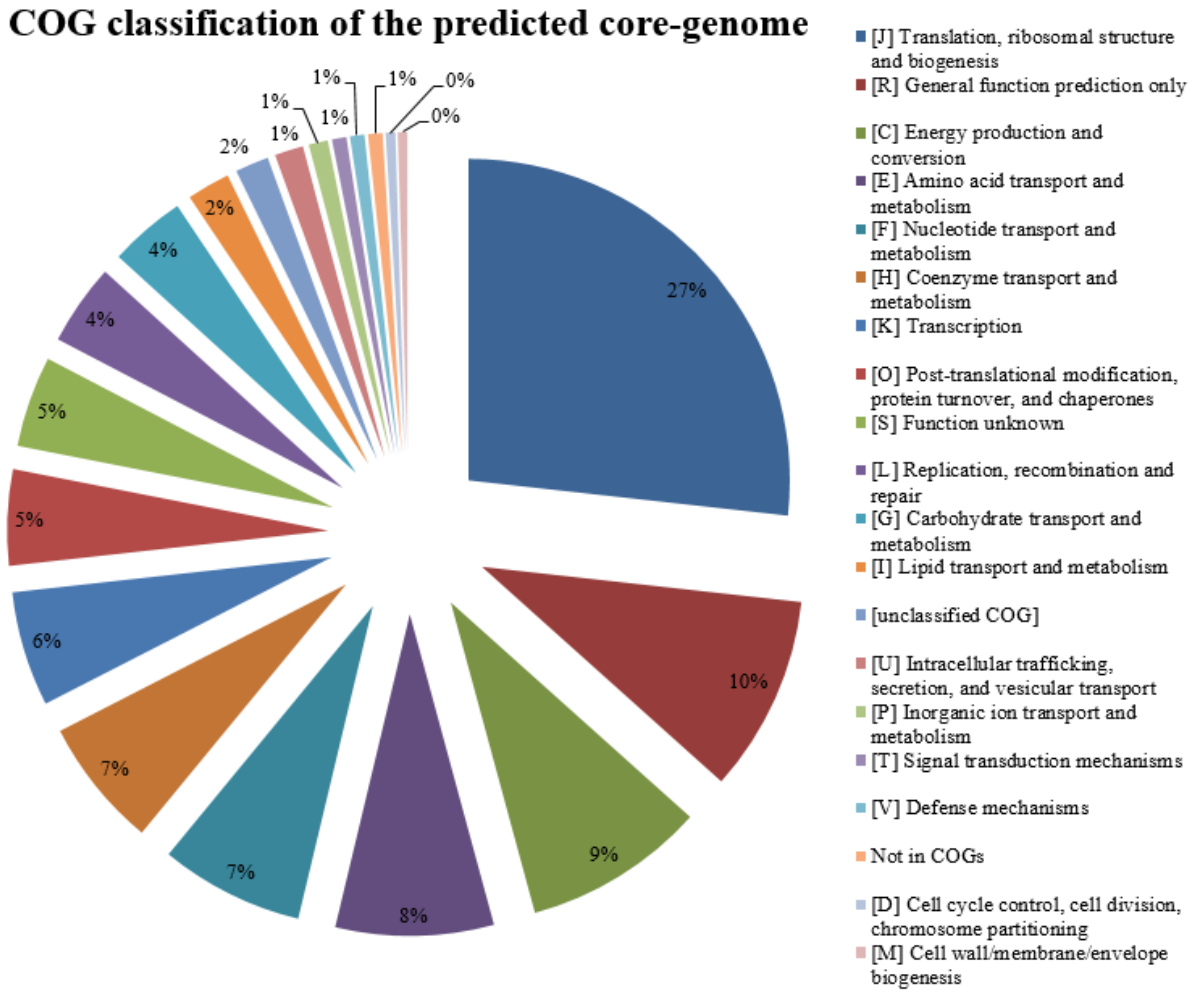


Figure 4.11. Chart displaying the COG classification of predicted core-genome for members of Methanomassiliicoccales. COG categories that contain 0 gene family amongst the core-genome are not displayed.

Adaptation to the rumen environment. 18 gene families are uniquely conserved between all four genomes of Methanomassiliicoccales sourced from the rumen (Table 4.8.) while 41 gene families are conserved between the ISO4-H5 and ISO4-G1 genomes, both of which have been sequenced from the ovine rumen (Table 4.9.).

Table 4.8. Gene families conserved only in rumen Methanomassiliicoccales by functional category

Locus_tag*	Predicted gene product	COG category
AR505_0479	hypothetical protein	Not in COGs
AR505_0645	carbon starvation protein CstA	[T]
AR505_0681	4Fe-4S Fdx iron-sulfur binding domain-containing protein	[C]
AR505_0718	TfoX N-terminal domain protein	[K]
AR505_0801	HTH domain-containing protein	[K]
AR505_0806	MFS transporter	[P]
AR505_0831	transcriptional regulator MarR family	[K]
AR505_0847	hypothetical protein	Not in COGs
AR505_0921	hypothetical protein	[unclassified COG]
AR505_0960	Arylsulfotransferase AssT	[unclassified COG]
AR505_0995	xanthine/uracil permease family protein	[R]
AR505_1037	DNA polymerase IV	[L]
AR505_1083	hypothetical protein	Not in COGs
AR505_1104	phosphoenolpyruvate synthase PpsA2	[G]
AR505_1105	hypothetical protein	Not in COGs
AR505_1106	dolichol kinase (7 TMHs)	[I]
AR505_1118	MTA/SAH nucleosidase MtnN	[F]
AR505_1272	transporter MIP family	[G]

*only one gene from ISO4-H5 for each gene family is represented

Table 4.9. Gene families conserved only in ovine rumen Methanomassiliicoccales by functional category

Locus_tag*	Predicted gene product	COG category
AR505_0100	thiamine monophosphate synthase ThiE1	[H]
AR505_0103	thiamine biosynthesis protein ThiF1	[H]
AR505_0104	thiamine biosynthesis protein ThiS	[H]
AR505_0112	ABC transporter substrate-binding protein	[P]
AR505_0314	hypothetical protein	Not in COGs
AR505_0320	hypothetical protein	Not in COGs
AR505_0322	hypothetical protein	Not in COGs
AR505_0323	hypothetical protein	Not in COGs
AR505_0346	hypothetical protein	Not in COGs
AR505_0597	anaerobic cobalt chelatase CbiK	[H]
AR505_0629	transmembrane protein	Not in COGs
AR505_0678	Citrate synthase GltA	[C]
AR505_0707	hypothetical protein	Not in COGs
AR505_0720	hypothetical protein	[S]
AR505_0725	4Fe-4S Fdx iron-sulfur binding domain-containing protein	[C]
AR505_0726	ApbE family protein	[S]
AR505_0812	Na/Pi-cotransporter II-like protein	[P]
AR505_0822	Nitroreductase family protein	[C]
AR505_0835	transporter Na^+/H^+ antiporter family	[P]
AR505_0875	hypothetical protein	Not in COGs
AR505_0933	Cardiolipin synthase	[I]
AR505_0941	MFS transporter	[P]
AR505_0978	flavodoxin-like protein	[C]
AR505_0983	ATPase	[R]
AR505_1039	NADPH-dependent FMN reductase	[R]
AR505_1058	hypothetical protein	Not in COGs
AR505_1081	small multidrug resistance protein	[P]
AR505_1082	small multidrug resistance protein	[P]
AR505_1088	hypothetical protein	Not in COGs
AR505_1100	K^+-dependent Na^+/Ca^+ exchanger	[P]
AR505_1102	thymidylate kinase Tmk1	[F]
AR505_1114	hypothetical protein	[unclassified COG]
AR505_1115	DEAD/DEAH box helicase domain-containing protein	[R]
AR505_1178	radical SAM domain containing protein	[R]
AR505_1327	monomethylamine methyltransferase MtmB1	[unclassified COG]
AR505_1363	hypothetical protein	[unclassified COG]
AR505_1464	L-2,4-diaminobutyrate decarboxylase	[E]
AR505_1465	2,4-diaminobutyrate 4-transaminase	[E]
AR505_1470	transmembrane protein	Not in COGs
AR505_1740	Fic family protein	[S]
AR505_1743	5'-nucleotidase SurE	[F]

*only one gene from each gene family is represented

Novel gene families in genomes of ovine rumen strains. The analysis of gene family conservation identified 551 unique gene families in ISO4-H5 (Table A.4.12.), 526 in ISO4-G1 (Table A.4.13.) and 465 in ISO4-G11 (Table A.4.14.). Among the 551 gene families novel to ISO4-H5, 343 are predicted to be hypothetical proteins. ISO4-H5 contains a high number of insertion elements (Table 3.6.), consisting of nine gene families out of the 40 gene families categorised under the replication, recombination and repair COG function. ISO4-H5 carries two copies of the bifunctional dihydrofolate synthetase *folC* gene involved in tetrahydrofolate production. One copy (AR505_0729) is considered novel under the criteria used in this study. ISO4-H5 carries two copies of the cell division protein *ftsZ* gene involved in cell replication, one (AR505_1792) is considered novel under the criteria used in this study. Three copies of signal peptidase I have been found in ISO4-H5 genome, and one (AR505_1640) is considered to be novel under the criteria of this study.

Among the 465 gene families novel to ISO4-G11, 292 are predicted to be hypothetical proteins, and 73 gene faimiles are poorly characterized. There are 130 gene families assigned to COGs groups with known functions, including 20 gene families involved in ion transport and a series of genes between ISO4G11_0238 and ISO4G11_0275 on a single contig, including genes with a toxin and antitoxin annotation.

ISO4-G1 has 526 unique gene families, 261 are predicted to be hypothetical proteins, 82 are poorly characterized, and 196 are assigned to COGs groups with known function. These include 34 gene families involved in ion transport, of which 19 gene families were predicted to be involved in iron transport. A NRPS is unique to ISO4-G1 (ISO4G1_0984), and is predicted to encode a 5,216 aa protein. This NRPS contains two in-frame amber codons, and is predicted to incorporate pyrrolysine. Adjacent to the NRPS gene is an ABC transporter operon (ISO4G1_0989, ISO4G1_0990 and ISO4G1_0991), including an ABC transporter permease (ISO4G1_0990) that might be involved in export of the NRPS product. There are three genes encoding a chaperone near the NRPS gene (ISO4G1_0975, ISO4G1_0978, ISO4G1_0979) potentially involved in formation of the NRPS product. Domain analysis of the ISO4-G1 NRPS shows it contains four modules, each containing an adenylation domain, a condensation domain and ending with a peptidyl carrier protein domain (Figure 4.12.). The first module includes an additional adenylation domain and the NRPS ends with an additional condensation domain. The substrate specificity for adenylation domains were predicted from a database of adenylation domains with known substrate, and the predicted ISO4-G1 NRPS

adenylation substrate include ornithine, 2,4-diamino-butyric acid, alanine, tyrosine and pipecolic acid**.**

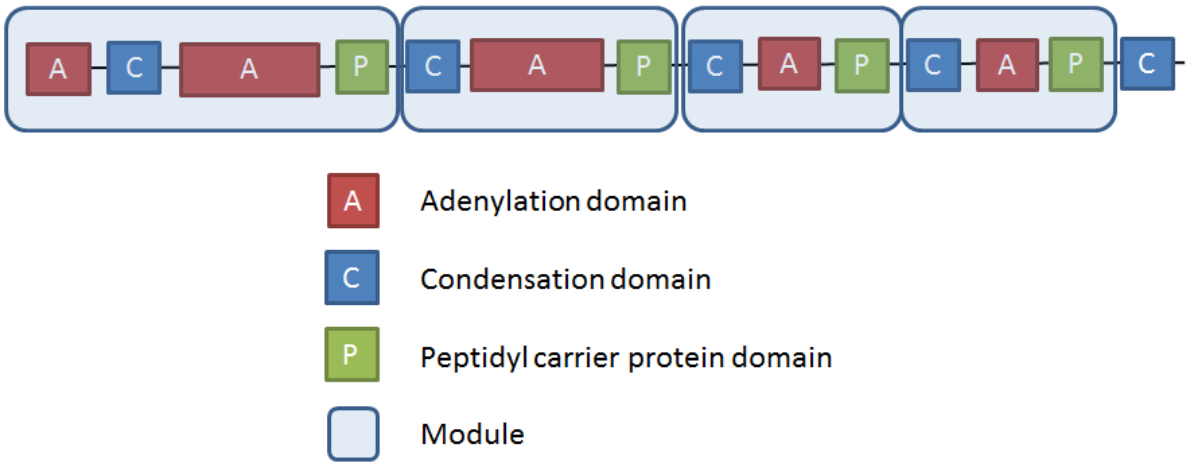


Figure 4.12. NRPS of ISO4-G1. The diagram is not drawn to scale.

4.2.7. Comparative analysis of metabolic profiles predicted from genomic sequences

Comparative analysis of the ORFeomes of the 11 Methanomassiliicoccales genomes suggests that the central metabolism and the methanogenesis pathway of these strains are largely similar, with the exception of RumEn M1 genome, which lacks methyl-substrate utilising genes.

Methanogenesis. All of the genes predicted to be involved in methanogenesis and energy generation in the genomes analysed are summarised in Table A.4.15. All of the genomes possess a methyl-CoM reductase (*mrt*), a heterodisulfide reductase (*hdr*), and a methyl viologen-dependent hydrogenase (*mvh*) and lack the genes encoding the hydrogenotrophic section of the methanogenesis pathway (Figure 4.13.). An Fpo-like complex is predicted to be the key cation transporter linking methanogenesis to the generation of membrane potential, which allows the ATP synthase complex to produce energy driven by a Na^+ or H^+ gradient (Table A.4.15.).

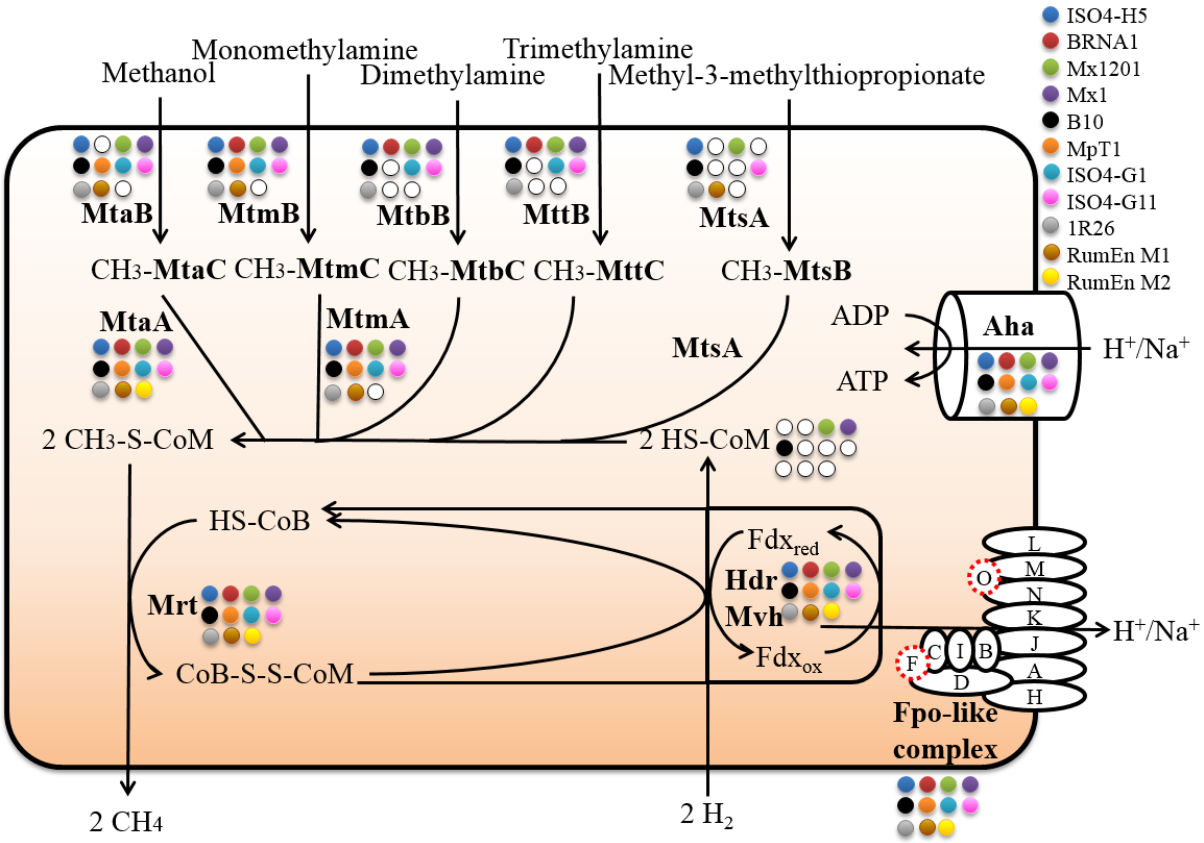


Figure 4.13. Methanogenesis pathway from methyl compounds by Methanomassiliicoccales. Key genes are shown in bold. The presence or absence of each gene is highlighted by colored circles. A white circle indicates absence of that gene in the corresponding genome. A red dotted ring indicates the absence of a certain protein subunit within the enzyme complex. Methanol (MeOH), monomethylamine (MMA), dimethylamine (DMA), trimethylamine (TMA), methyl-3-methylthiopropionate (M3MTP), methanol corrinoid protein (MtaC), methyl: CoM methyltransferase (MtaA), methanol:corrinoid methyltransferase (MtaB), trimethylamine:corrinoid methyltransferase (MttB), trimethylamine corrinoid protein (MttC), bifunctional methylthiol:corrinoid methyltransferase (MtsA), methylthiol corrinoid protein (MtsB), monomethylamine:corrinoid methyltransferase (MtmB), monomethylamine corrinoid protein (MtmC), dimethylamine:corrinoid methyltransferase (MtbB), dimethylamine corrinoid protein (MtbC), methyl-CoM reductase (Mrt), methylviologen hydrogenase (Mvh), heterodisulfide reductase (Hdr), ferredoxin (Fdx), A_1A_0 ATP synthase (Aha).

A phylogenetic tree inferred from the amino acid sequence of the McrA/MrtA subunit clearly places members of the order Methanomassiliicoccales in a branch of their own, away from those of Methanosarcinales, Methanobacteriales, and Methanococcales (Figure 4.14.). The *mrtG* gene was absent in the 1R26 genome (a frameshift mutation or sequencing error in the stop codon led to its lack of annotation), the *mrtG* gene was identified based on homology and is present between *mrtD* and *mrtA*. The operon organisation of the *mrt* genes of Methanomassiliicoccales has been observed to follow the order of *mrtB*, *mrtD*, *mrtG*, *mrtA* (Figure 4.14A.), followed by three to seven other genes, then resumes with *atwA*, *mrtC* downstream on the same strand. Mx1, RumEn M1 and MpT1 are exceptions to this operon organisation; the *atwA* and *mcrC* are located far apart (920 kb apart for Mx1, 14.5 kb apart for RumEn M1 and 427 kb apart for MpT1) from the remainder of the operon (Table A.4.15.). The

sequence and operon organisation indicate an MCRII/Mrt-type enzyme. In the Methanomassiliicoccales genomes investigated, there is a short conserved hypothetical protein of approximately 92 amino acids, which is located immediately downstream of *mrtA* on the same strand, followed by 2-6 genes which are usually encoded on the opposite strand. This gene is absent from other methanogens that encode *mrt* operons (Figure 4.15A.). This hypothetical protein displays a high level of conservation at the aa sequence level ranging from 31.9% aa identity between 1R26 and B10 to 72.8% aa identity between ISO4-H5 and ISO4-G11 (Figure 4.15C.). The phylogenetic relationship inferred from the aa sequence of this hypothetical protein is similar to the one that is inferred from the McrA/MrtA sequences (Figure 4.15B.), placing RumEn M1, Mx1 and B10 clearly away from other members of the Methanomassiliicoccales order. This hypothetical protein may be functionally associated with the Mrt enzyme complex, but a lack of homology to any known protein precludes assignment of its function.

A number of genes with unknown functions have been identified as genetic markers for methanogens via the TigrFam database, they are known as methanogenesis marker proteins (MMP). Six MMP genes (*mmp3, mmp6, mmp5, mmp15, mmp7* and *mmp17*) are located downstream and on the same strand as *mrtC*, whereas *mmp9, mmp10, mmp12* and *mmp14* are absent amongst the genomes analysed (Table A.4.15.).

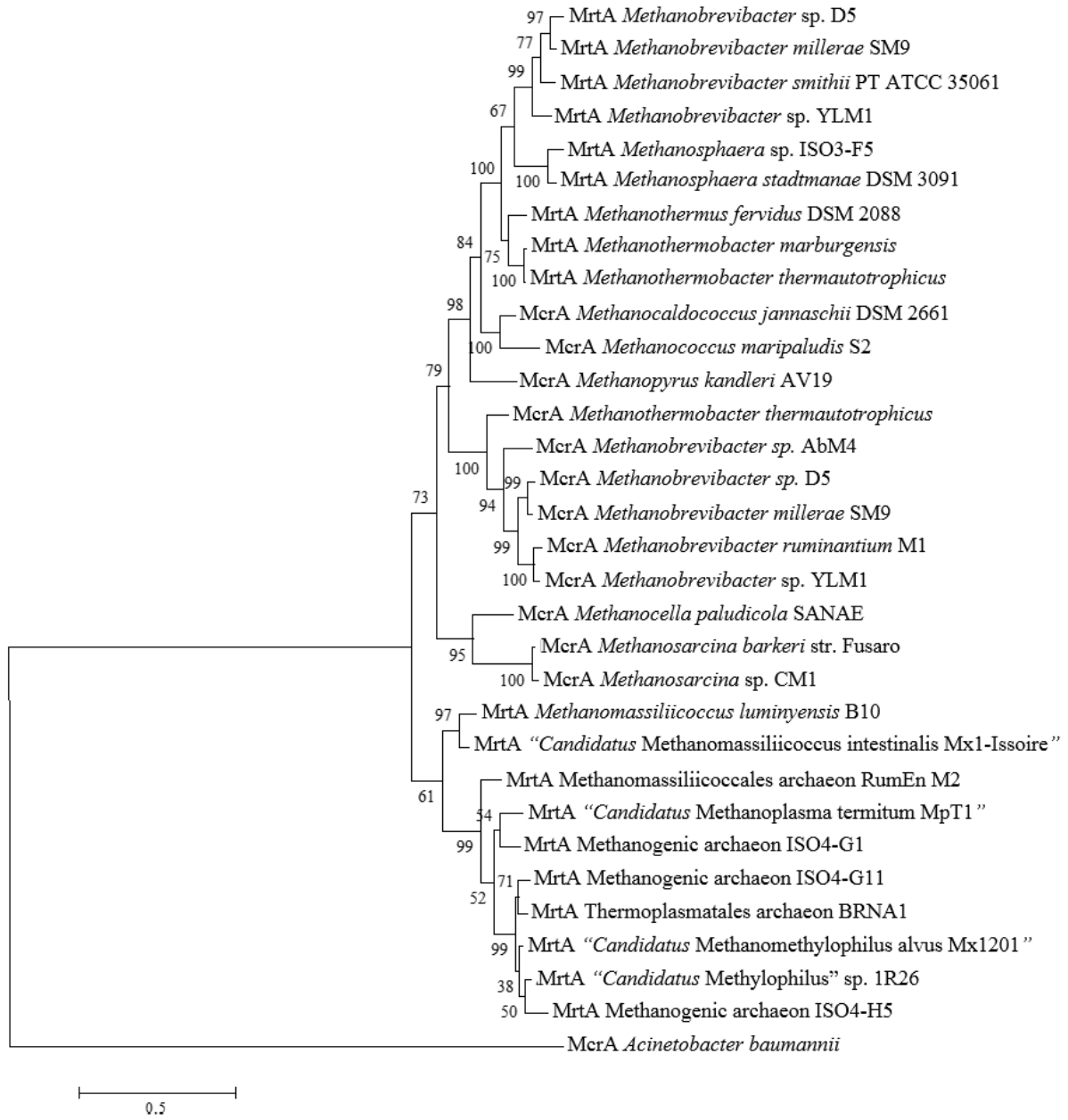


Figure 4.14. Phylogenic tree of methanogens based on McrA/MrtA amino acid sequences. The phylogenic relationships of 26 methanogens were inferred using Maximum Likelihood based on the JTT matrix-based model (Jones *et al.* 1992). Scale bar indicates the number of substitutions per site. The MrtA of RumEn M1 is a pseudogene and is precluded from this analysis. Numbers at the branch point represent boostrap value from 1000 replicates. *Acinetobacter baumannii* was used as an outgroup.

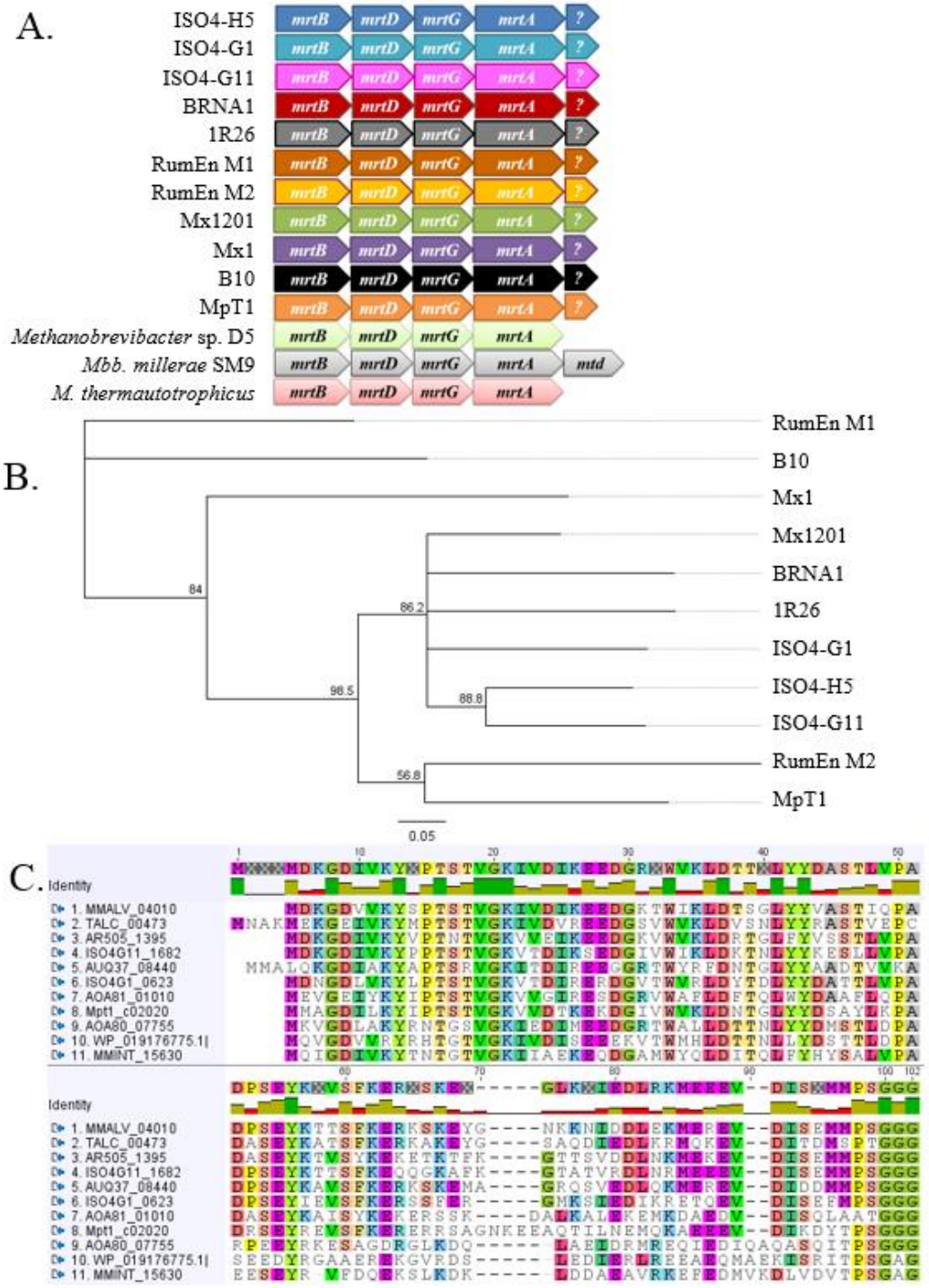


Figure 4.15. Conservation of a hypothetical protein associated with the Mrt operon in Methanomassiliicoccales. **A**. *mrt* operon organisation of members of Methanomassiliicoccales compared to *mrt* operons of Methanobacteriales; *Methanobrevibacter*. sp. D5, *Methanobrevibacter* sp. strain D5; *Methanobrevibacter millerae* SM9 and *Methanobacterium thermautotrophicus*. Methyl-CoM methyltransferase subunit X (*mrtX*, where X = A, B, D or G). Methylenetetrahydromethanopterin dehydrogenase (*mtd*). Conserved hypothetical protein downstream of *mrtA* gene (?). **B**. Phylogenic tree inferred from aa sequence of conserved hypothetical protein downstream of *mrtA* gene. Tree generated by Geneious Tree Builder using JTT distance model, built by Neighbour-Joining (Kearse *et al.* 2012),. The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) was used to infer the evolutionary history of the taxa analysed. Bar: 0.08 substitutions per amino acid position. **C**. ClustalW alignment of aa sequence of conserved hypothetical protein downstream of *mrtA* gene (Thompson *et al.* 2002), BLOSUM was used as cost matrix 10 gap open cost and 0.1 gap extend cost (Mount 2008).

Genes encoding methyl compound utilisation. RumEn M2 genome encodes no genes involved in methyl compound utilisation. Ten members of the Methanomassiliicoccales encode the methylamine:corrinoid methyltransferase MtmB and its corresponding corrinoid protein MtmC. All members of the Methanomassiliicoccales order except BRNA1 and RumEn M2 are predicted to utilise methanol, and all except MpT1 and RumEn M1 to use dimethylamine and trimethylamine (Figure 4.13., Table A.4.15.). The copy number of methylamine:corrinoid methyltransferases varies between genomes. B10 carries four copies of *mtmB*, ISO4-G1 carries five copies of trimethylamine:corrinoid methyltransferase (*mttB*). Two *mttB* genes (ISO4G1_0945 from ISO4-G1 and TALC_00306 from BRNA1) do not contain an in-frame amber codon.

In addition to methylamine methyltransferases, utilisation of methylamine requires an activation protein for corrinoid proteins and a permease to import the substrates. The methyltransferase corrinoid activation protein (*ramA*) which is required to reactivate cobalt at the active site of MtmC, MtbC and MttC, was found across all members of the Methanomassiliicoccales examined (Table A.4.15.), including a RamA homologue in RumEn M2 that shares 43.8% aa identity to experimentally verified RamA of *Methanosarcina barker*i. There were no monomethylamine permease genes found in any genome, but homologues of dimethylamine permeases were predicted in members of the Methanomassiliicoccales excluding RumEn M1, RumEn M2, MpT1 and B10. Homologues of a trimethylamine permease are similarly distributed in the Methanomassiliicoccales genomes but not in ISO4-G11. No methylamine permease genes could be found at all in the B10 and MpT1 genomes. The bifunctional methylcobalamin: CoM methyltransferase (*mtsA*) and the corresponding corrinoid protein (*mtsB*) required for methylthiol utilisation are only predicted in ISO4-H5, ISO4-G11, RumEn M1, 1R26, Mx1, Mx1201 and B10 (Table A.4.15.).

Heterodisulfide reductase and methyl-viologen hydrogenase. The cytoplasmic heterodisulfide reductase (*hdrABC*) and methyl-viologen dependent hydrogenase (*mvhAGD*) is highly conserved across all members of the Methanomassiliicoccales order. The *hdrBC* and *mvhAGD* genes are organised as operons, with *mvhD* present in multiple copies in Mx1 and B10 (Table A.4.15.). The presence of a lone *hdrD* gene of the periplasmic HdrDE is also conserved across the genomes analysed.

F_{420} methanophenazine oxidoreductase (Fpo)-like complex and hydrogenases. Members of the Methanomassiliicoccales lack the hydrogenotrophic section of the methanogenesis pathway

(Figure 1.2.) and depend on the large transmembrane Fpo-like complex (*fpoABCDHIJKLMN*). However, all genomes lack the FpoF and FpoO subunits, thought to be involved in F_{420} binding and methanophenazine electron transfer, respectively. The Fpo-like complex operon organisation is conserved across the eight genomes analysed, with a gene encoding a conserved hypothetical protein present between *fpoJ* and *fpoK* (Table A.4.15.). This particular hypothetical protein is highly conserved at the aa level, ranging from 17.3% to 59% aa identity (Figure 4.16.). In addition, the energy-conserving hydrogenase (Ech) complex is predicted to be only present in Mx1, RumEn M1 and B10 (Table A.4.15.).

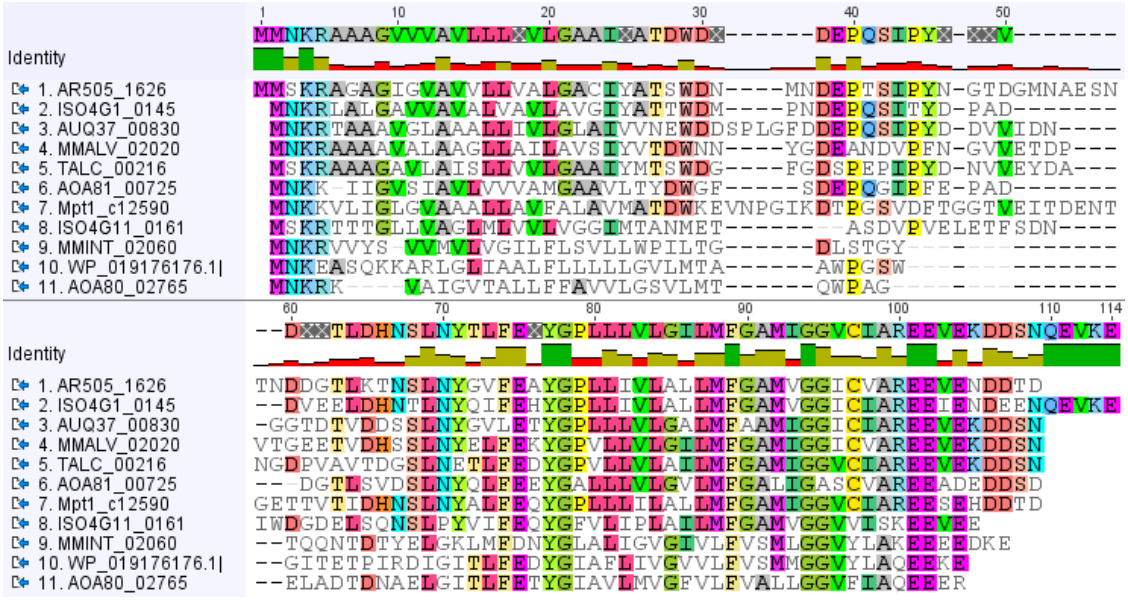


Figure 4.16. ClustalW alignment of predicted protein amino acid sequence of conserved hypothetical protein located between the *fpoJ* and *fpoK* genes. Quality evaluation used BLOSUM as the substitution matrix with 10 gap opening cost and 0.1 gap extention cost (Mount 2008).

RumEn M1, RumEn M2, B10 and MpT1 contain genes encoding several subunits of a Hyf-like complex similar to the [NiFe] hydrogenases of *Escherichia coli*, and a formate dependent hydrogenase. However, the large subunit of the Hyf-like hydrogenase is absent in RumEn M1 due to end of contig, furthermore, in RumEn M2, MpT1 and B10 lack the signature NiFe-binding motif (Figure 4.17.), therefore it is uncertain whether the Hyf-like complex is functional as a hydrogenase in these organisms.

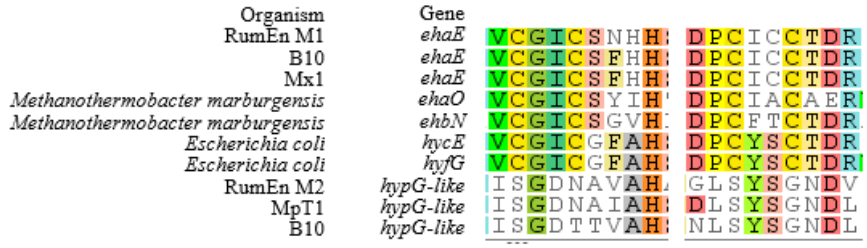


Figure 4.17. ClustalW alignment of predicted protein amino acid sequence of NiFe-binding motif in hydrogenase large subunits. Quality evaluation used BLOSUM as the substitution matrix with 10 gap opening cost and 0.1 gap extention cost (Mount 2008).

Coenzyme M biosynthesis. The capability to make CoM varies among the members of the Methanomassiliicoccales. In Mx1201, Mx1 and B10 where it is predicted to be synthesised, it appears not to be made via the typical sulfolactate synthase (*comABC*) pathway, but instead requires cysteate synthase (homologous to threonine synthase), together with a multisubstrate phosphoserine/aspartate aminotransferase (*serC)* followed by a sulfopyruvate decarboxylase (*comDE*) (Figure 4.18.).

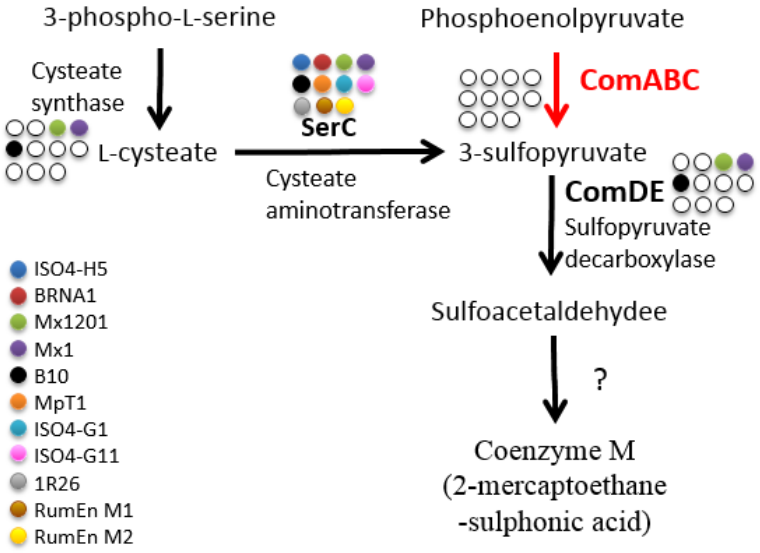


Figure 4.18. CoM biosynthesis pathway in the Methanomassiliicoccales genomes analysed. Broad substrate specificity phosphoserine transaminase (SerC), (2*R*)-phospho-3-sulfolactate synthase (ComA), 2-phosphosulfolactate phosphohydrolase (ComB), (2*R*)3-sulfolactate dehydrogenase (ComC), sulfopyruvate decarboxylase (ComDE), unidentified enzyme (?). Black arrows represent existing pathways. Red arrows represent absent pathways.

A multifunctional phosphoserine/aspartate aminotransferase is found across all members of the Methanomassiliicoccales as it is involved in a wide range of other pathways. However, *comD* and *comE* were only found in Mx1 and B10, and a fused *comDE* was also found in Mx1201. A cysteate synthase homologue has also been found in these three members of the Methanomassiliicoccales. BRNA1 was found to have a phosphonopyruvate decarboxylase

(TALC_00082) with homology to *comDE* at the 3' end, however, the presence of only a single copy of threonine synthase gene (TALC_00599) means BRNA1 is unlikely to produce CoM. None of the other members of the Methanomassiliicoccales possess the genes encoding CoM biosynthesis (Table A.4.15.).

F_{430} biosynthesis. F_{430} is used in the active site of the Mcr/Mrt complex but only one gene, *corA* encoding uroporphyrin-III C-methyltransferase is known to be involved in F_{430} biosynthesis. All of the Methanomassiliicoccales genomes encode *corA*, except ISO4-G1, ISO4-G11 and RumEn M2, it is likely they require exogenous sources of F_{430} to survive.

A_1A_0-ATP synthase. Genes encoding nine subuntis of A_1A_0-ATP synthase are present in all the genomes analysed, and are all positioned close to the putative origins of replication for each organism. The conserved operon structures consist of the genes ordered *ahaHIKECFABD* (Table A.4.15.).

H^+/Na^+ antiporters. All the genomes analysed possess H^+/Na^+ antiporter genes (Table A.4.15.).

Central carbon metabolism

Gluconeogenesis. All members of the order Methanomassiliicoccales examined encode the enzymes necessary to carry out gluconeogenesis from pyruvate to glucose-1-phosphate (Figure 4.19., Table A.4.16.). However, the enzymes differ and three variants in the gluconeogenesis pathway were found among the genomes analysed. The 2,3-bisphosphoglycerate-independent phosphoglycerate mutase gene (*apgM*) specific to archaea was seen in all genomes while ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, RumEn M1, 1R26 and Mx1201 also possess the bacterial bisphosphoglycerate-dependent *gpmA* variant. B10 was the only genome to possess the phosphofructokinase gene, *pfkB* involved in glycolysis. The MpT1 genome is the only genome lacking the phosphoglucose/phosphomannose isomerase gene, *pgi*.

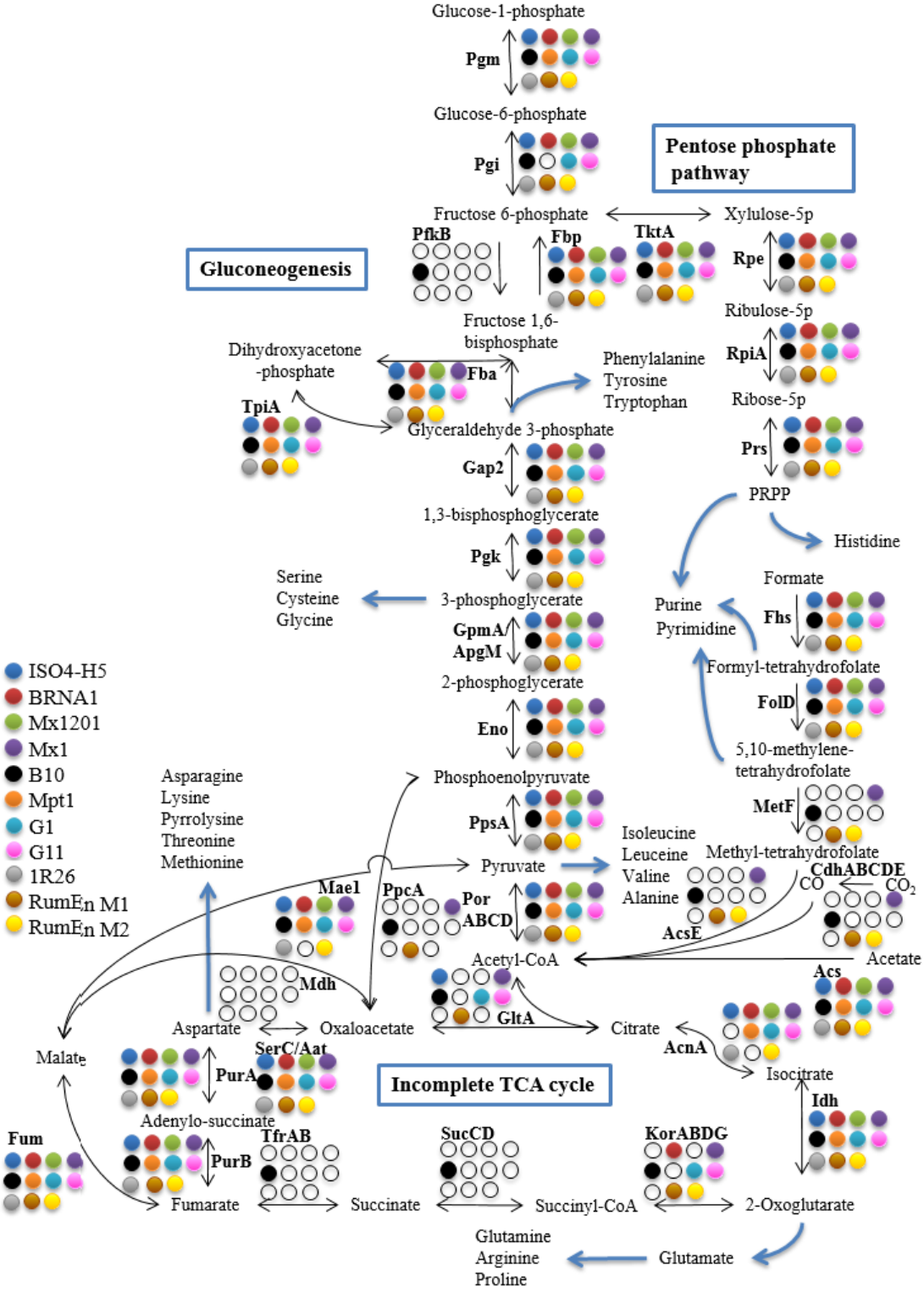

Glucose-1-phosphate
Pgm
Glucose-6-phosphate
Pgi
Pentose phosphate pathway
Fructose 6-phosphate
Xylulose-5p
Gluconeogenesis
PfkB
Fbp
TktA
Rpe
Fructose 1,6-bisphosphate
Ribulose-5p
RpiA
Dihydroxyacetone -phosphate
Fba
Phenylalanine
Tyrosine
Tryptophan
TpiA
Glyceraldehyde 3-phosphate
Ribose-5p
Prs
Gap2
PRPP
1,3-bisphosphoglycerate
Histidine
Pgk
Serine
Cysteine
Glycine
3-phosphoglycerate
Formate
Purine
Pyrimidine
Fhs
GpmA/ ApgM
Formyl-tetrahydrofolate
2-phosphoglycerate
FolD
Eno
5,10-methylene-tetrahydrofolate
ISO4-H5
BRNA1
Mx1201
Mx1
B10
Mpt1
G1
G11
1R26
RumEn M1
RumEn M2
Asparagine
Lysine
Pyrrolysine
Threonine
Methionine
Phosphoenolpyruvate
MetF
PpsA
Isoleucine
Leuceine
Valine
Alanine
Methyl-tetrahydrofolate
Pyruvate
CdhABCDE
CO
CO2
Mae1
PpcA
Por ABCD
AcsE
Acetyl-CoA
Acetate
Acs
Mdh
Aspartate
Oxaloacetate
GltA
Citrate
AcnA
Malate
SerC/Aat
PurA
Isocitrate
Incomplete TCA cycle
Idh
Adenylo-succinate
Fum
PurB
TfrAB
SucCD
KorABDG
Fumarate
Succinate
Succinyl-CoA
2-Oxoglutarate
Glutamine
Arginine
Proline
Glutamate

Figure 4.19. Central carbon metabolism pathways predicted for Methanomassiliicoccales. Key genes are shown in bold. Black arrows indicate pathways, blue arrows represents pathways leading out of central carbon metabolism. The presence or absence of each gene in a particular genome is highlighted by colored circles: a white circle indicates absence from a genome, and a coloured circle indicates the presence in the corresponding genome. Formate-tetrahydrofolate ligase (Fhs), NADP-dependent methylene tetrahydrofolate dehydrogenase (FolD), 5,10-methylenetetrahydrofolate reductase (MetF), carbon-monoxide dehydrogenase complex (CdhABCDE), methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (AcsE), acetyl CoA synthetase (Acs), phosphoglucomutase (Pgm), phosphoglucose isomerase (Pgi), fructose-1,6-bisphosphatatase (Fbp), phosphofructokinase (PfkB), fructose-bisphosphate aldolase (Fba), glyceraldehyde-3-phosphate dehydrogenase (Gap2), phosphoglycerate kinase (Pgk), phosphoglycerate mutase (ApgM), phosphopyruvate hydratase (Eno), phosphoenolpyruvate synthase (PpsA), pyruvate:Fdx oxidoreductase (PorABCD), citrate synthase (GltA), aconitate hydratase (AcnA), isocitrate dehydrogenase (Idh), 2-oxoglutarate synthase (KorABDG), succinyl-CoA synthetase (SucCD), thiol:fumarate reductase (TfrAB), fumarate hydratase (Fum), adenylosuccinate lyase (PurB), adenylosuccinate synthetase (PurA), malate dehydrogenase (Mdh), phosphoserine aminotransferase/aspartate aminotransferase (SerC/Aat), NAD-dependent malic enzyme (Mae1), phosphoenolpyruvate carboxylase (PpcA), transketolase (TktA), ribulose-phosphate 3-epimerase (Rpe), ribose-5-phosphate isomerase (RpiA), ribose-phosphate diphosphokinase (Prs).

Incomplete TCA cycle. All the genomes are predicted to have an incomplete reductive TCA cycle (Figure 4.19.), but the extent of this pathway differs between members of the Methanomassiliicoccales. The ISO4-H5, Mx1, B10, ISO4-G1 and ISO4-G11 possess a citrate synthase gene (*gltA*), which allows the incorporation of acetyl-CoA into the TCA cycle. The B10 genome lacks a recognisable aconitate hydratase gene, *acnA*, but it is the only genome with genes encoding succinate-CoA ligase (*sucCD*) and CoM/CoB dependent thiol:fumarate reductase (*tfrAB*) necessary to interconvert 2-oxoglutarate and fumarate. The ISO4-H5, Mx1201, 1R26 and MpT1 genomes lack genes encoding the 2-oxoglutarate synthase complex (*korABDG*). A gene for malate dehydrogenase (*mdh*) which interconverts 2-oxaloacetate and malate is absent in all the genomes analysed, but the combined actions of the gene products of adenylosuccinate lyase (*purB*), adenylosuccinate synthetase (*purA*) and the broad substrate specificity phosphoserine aminotransferase (*serC*) can produce oxaloacetate. Oxaloacetate can also be produced by phosphoenolpyruvate carboxylase (*ppcA*), but only the RumEn M1, B10 and Mx1 genomes have the *ppcA* gene.

Pentose phosphate pathway. The genes necessary to produce phosphoribosylpyrophosphate (PRPP, used for histidine and purine biosynthesis) from fructose-6-phosphate can be found in all the genomes (Figure 4.19., Table A.4.16.), including transketolase (*tktA*), ribulose-phosphate 3-epimerase (*rpe*), ribose-5-phosphate isomerase (*rpiA*) and ribose-phosphate pyrophosphokinase (*prs*).

Reductive acetyl-CoA pathway. All members of the order Methanomassiliicoccales analysed are predicted to produce acetyl-CoA from acetate via acetyl-CoA synthetase (*acdA*). While all members of the Methanomassiliicoccales possess a formyltetrahydrofolate synthetase gene

(*fhs*) and and a bifunctional methenyltetrahydrofolate cyclohydrolase/dehydrogenase (*folD*) to produce 5,10 methylene-tetrahydrofolate (5,10-mTHF, an important cofactor found with many enzymes) from formate, only Mx1 and B10 encode an incomplete CO dehydrogenase complex (*cdhABCDE*) granting the capabilty of *de novo* biosynthesis of acetyl-CoA from formate and CO_2 (Figure 4.19., Table A.4.16.).

Alternative carbon sources. Members of the Methanomassiliicoccales may utilise alcohol as an alternative carbon source. A type IV iron containing alcohol dehydrogenase (*adh*) is predicted in ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26 and Mx1201 (Table A.4.16.) which is homologous to a multifunctional NADP-dependent alcohol dehydrogenase (NC_001988.2) in *Clostridium*, capable of producing acetyl-CoA from ethanol. A homologue to a cytoplasmic D-lactate dehydrogenase (*ldhA*) is also predicted in BRNA1, RumEn M1, RumEn M2, 1R26 and Mx1201, which may convert lactate to pyruvate.

Carbon fixation. Homologues of ribulose 1,5-bisphosphate carboxylase (*rbcL*) and ribulose 1,5-bisphosphate isomerase (*e2b2*) have been identified among the core genome of Methanomassiliicoccales. In all eight genomes analysed, *rbcL* and *e2b2* form a conserved operon with AMP phosphorylase (*deoA*).

Amino acid biosynthesis. The main pathways of amino acid biosynthesis in the Methanomassiliicoccales are shown in Figure 4.20 and the genes encoding these pathways are presented in Table A.4.17.

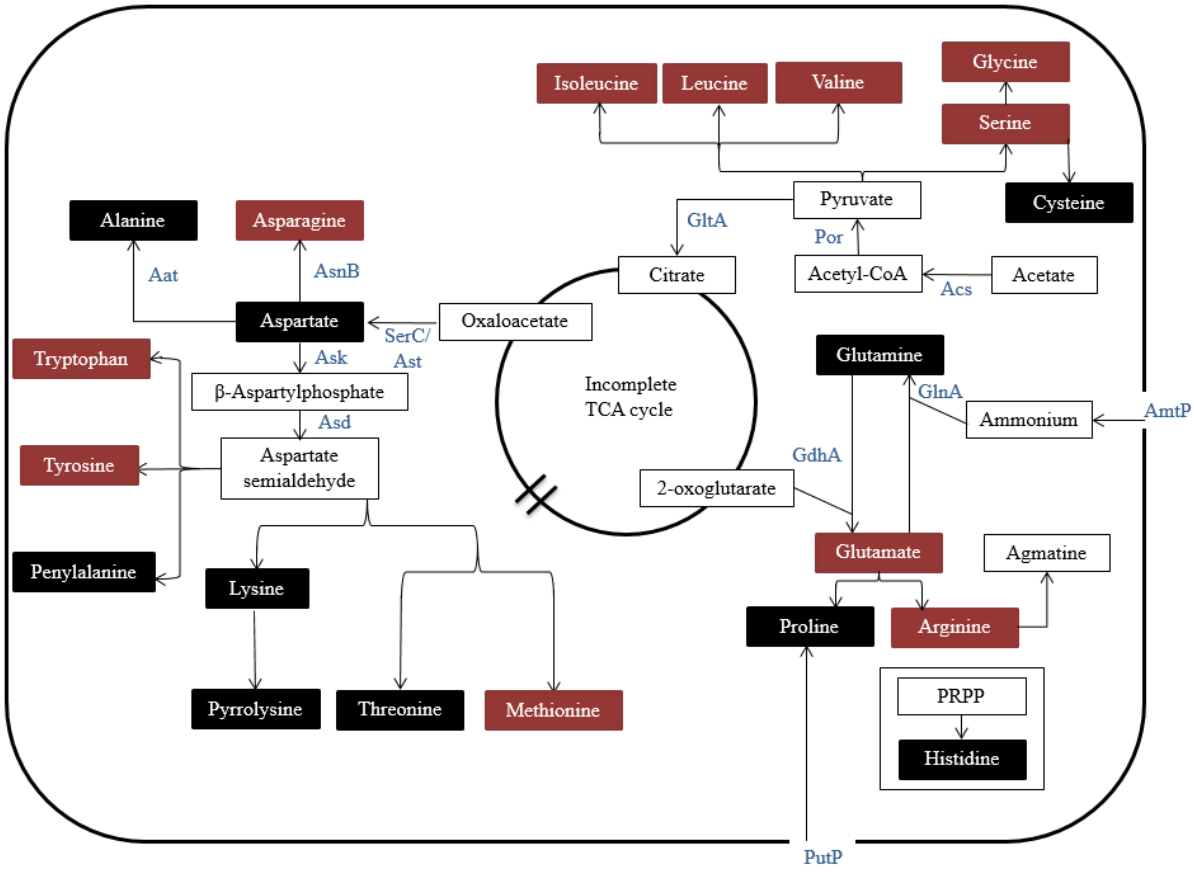


Figure 4.20. Amino acid biosynthesis of Methanomassiliicoccales. Amino acids predicted to be biosynthesized by all Methanomassiliicoccales analysed are displayed in black boxes, while amino acids predicted not to be biosynthesized by one or more Methanomassiliicoccales analysed are in red boxes. Intermediates and molecules are displayed in white boxes. Enzymes are displayed as blue text. Citrate synthase (GltA), pyruvate:Fdx oxidoreductase (Por), acetyl-CoA synthase (Acs), glutamate dehydrogenase (GdhA), glutamine synthetase (GlnA), ammonium transporter (AmtP), sodium/proline symporter (PutP), phosphoribosylpyrophosphate (PRPP), asparagine synthase (AsnB), alanine aminotransferase (Aat), broad substrate specificity phosphoserine aminotransferase (SerC), aspartate amino transferase (Ast), aspartate kinase (Ask), aspartate semialdehyde dehydrogenase (Asd).

All members of order Methanomassiliicoccales analysed encodes genes required to produce alanine, asparagine, aspartate, cysteine, glutamine, histidine, lysine, proline, pyrrolysine and threonine (Tables A.4.17., 4.10.).

Several genomes possess incomplete pathways for amino acid biosynthesis. Mx1 and MpT1 lack a glutamate dehydrogenase gene (*gdhA*) required to produce glutamate. B10 lacks the acetylglutamate kinase gene (*argB*) for arginine biosynthesis, *serAB* genes required for serine biosynthesis, shikimate dehydrogenase (*aroE*) required to produce chorismate, and prephenate

dehydrogenase (*tyrA*) required to produce tyrosine. ISO4-H5 lacks the serine hydroxymethyltransferase (*glyA*) required to convert serine to glycine. Mx1201 and Mx1 lack the broad substrate aromatic aminotransferase (*aro8*) involved in interconversion between aromatic amino acids. The isopropylmalate/citramalate isomerase *leuCD* genes involved in branched chain amino acid biosynthesis is predicted to be absent in the ISO4-G1, ISO4-G11 and BRNA1 genomes. Furthermore, ISO4-G1 lacks the acetolactate synthase (*ilvBN*) and dihydroxy-acid dehydratase (*ilvD*) gene. The genes required for tryptophan biosynthesis are entirely missing in the ISO4-G1 genome, while Mx1 encodes only tryptophan synthase beta chain (*trpB*) and anthranilate synthase (*trpD*) genes (Table A.4.17.).

Table 4.10. Amino acid biosynthesis summary

Amino acids	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Alanine	+	+	+	+	+	+	+	+	+	+	+
Arginine	+	+	+	+	+	+	+	+	+	-	+
Asparagine	+	+	+	+	-	+	+	+	+	+	+
Aspartate	+	+	+	+	+	+	+	+	+	+	+
Cysteine	+	+	+	+	+	+	+	+	+	+	+
Glutamine	+	+	+	+	+	+	+	+	+	+	+
Glutamate	+	+	+	+	+	-	+	+	-	+	-
Glycine	-	+	+	+	+	+	-	+	+	+	+
Histidine	+	+	+	+	+	+	+	+	+	+	+
Isoleucine	+	-	-	-	-	+	+	+	+	+	+
Leucine	+	-	-	-		+	+	+	+	+	+
Lysine	+	+	+	+	+	+	+	+	+	+	+
Methionine	-	-	-	-	-	-	-	+	-	-	-
Phenylalanine	+	+	+	+	+	+	+	-	-	+	+
Proline	+	+	+	+	+	+	+	+	+	+	+
Pyrrolysine	+	+	+	+	+	-	+	+	+	+	+
Serine	+	+	+	+	+	+	+	+	+	-	+
Threonine	+	+	+	+	-	+	+	+	+	+	+
Tryptophan	+	-	+	+	-	-	+	+	-	+	+
Tyrosine	+	+	+	+	+	+	+	+	+	-	+
Valine	+	-	-	-	-	+	+	+	+	+	+

+: This genome is predicted to be able to produce this amino acid. -: This genome is predicted to be unable to produce this amino acid.

Only the Mx1201 genome encodes a complete methionine biosynthesis pathway (Table A.4.17.). While the *metA* (homoserine *O*-sucinyltransferase**)** and *oah* (*O*-acetyl-L-homoserine sulfhydrylase) genes were identified in members of the Methanomassiliicoccales analysed except RumEn M1, RumEn M2 and B10, the other genes encoding the pathway are present sporadically in the genomes. All the genes necessary to carry out the *S*-adenosyl-L-methionine (SAM) cycle, apart from methionine synthase itself, have been identified in the ISO4-H5 and Mx1 genomes. All the genomes analysed possess the *speA* gene involved in agmatine biosynthesis and lack the *speE* gene involved in spermidine biosynthesis.

Nucleotide biosynthesis. All the genes required for purine and pyrimidine biosynthesis have been identified across members of the Methanomassiliicoccales analysed (Table A.4.18.;

Figure 3.12 shows a schematic for the pathway of nucleotide biosynthesis). The only exception is the *ndrJ* gene which is absent from the RumEn M1, RumEn M2 and B10 genomes.

Cell cycle. The genes necessary for cellular replication are summarised in Table A.4.19. The B10 draft genome lacks genes encoding a replication factor C large subunit, a DNA primase large subunit and a DNA gyrase subunit A (*gyrA*) as well as a cell division GTPase. All members of the Methanomassiliicoccales examined have two copies of the *orc1/cdc6* origin of replication genes.

Cofactor biosynthesis. Tetrapyrrole is the backbone of cobalamin and vitamin cofactors, including F_{430}. Members of the Methanomassiliicoccales produce tetrapyrrole by using the glutamyl-tRNA via an uroporphyrinogen-III intermediate, but are incapable of producing cytochromes. The entire set of genes involved in tetrapyrrole biosynthesis is absent in the ISO4-G1 genome, and the ISO4-G11 genome only has the porphobilinogen deaminase (*hemC*) gene, RumEn M2 genome only has the siroheme synthase (*cysG*) gene (Table A.4.20.), therefore ISO4-G1, ISO4-G11 and RumEn M2 are predicted not to produce F_{430}. Several genes required for adenylosylcobalamin biosynthesis are predicted to be absent in RumEn M1 and RumEn M2 genomes (Table A.4.20.), therefore RumEn M1 and RumEn M2 are predicted not to carry out *de novo* biosynthesis of adenylocobalamin.

All the Methanomassiliicoccales genomes analysed are predicted to possess the genes required for riboflavin biosynthesis, as well as subsequent FAD and FMN biosynthesis (Table 4.19.). Two genes are considered to be variants of the bacterial gene rather than the archaeal gene, the bifunctional diaminohydroxy phosphoribosylaminopyrimidine deaminase /5-amino-6-(5-phosphoribosylamino) uracil reductase (*ribD*) and the bacterial ATP-driven riboflavin kinase (*ribC*). The archaeal *ribC* gene product is CTP-dependent.

The genes required for *de novo* biosynthesis of NAD^+ and $NADP^+$ were only found in the B10 genome. The other genomes lack the L-aspartate oxidase gene (*nadB*) which encodes the first step of NAD^+ biosynthesis. All the genomes analysed possess the genes required for salvaging of NAD and NADP (Table A.4.20.).

Secretion. All the Methanomassiliicoccales genomes analysed are predicted to possess genes required for protein secretion (Table A.4.21.). The secretion is SRP dependent, the SRP is recognized by signal recognition receptor FtsY and cleaved by signal sequence peptidase, then exported via the Sec61 translocon.

4.3. Discussion

Eleven genomes from members of the order Methanomassiliicoccales sourced from four different gut environments (human, termite, bovine rumen, and ovine rumen) have been compared in this study, to identify the extent of gene conservation and to gain insight into the metabolism and physiology of this unusual group of methanogens. The genomes analysed all contain circular chromosomes with no extrachromosomal elements. All genomes are complete with the exception of B10, RumEn M1 (96.6% completion), RumEn M2 (94.1% completion) (Sollinger *et al.* 2016), 1R26 (92.98% complete) (Noel *et al.* 2016) and G11 which are currently in draft form. Amongst the sequenced members of the Methanomassiliicoccales, the genome size ranges between 1.46 million base pairs (Mbp) (BRNA1) and 2.62 Mbp (B10), and the % G+C range is between 41.3% (Mx1) and 62.1% (RumEn M1). A closer examination of those strains isolated from the rumen reveals a genome size of between 1.46 Mbp (BRNA1) and 1.94 Mbp (ISO4-H5), and a %G+C range of between 54.0% (ISO4-H5) and 58.3% (BRNA1). Members of the Methanomassiliicoccales order were originally associated with the Thermoplasmatales order (Tajima *et al.* 2001), whose genomes are known to lack an operon organisation of their rRNAs (Tu and Zillig 1982). Separation of the 16*S*, 23*S* and 5*S* rRNA genes has been reported in genomes of the Methanomassiliicoccales isolated from human and termite gut environments (Borrel *et al.* 2014; Lang *et al.* 2015). Analysis of the rRNA gene structure in the genomes isolated from the rumen environment confirms this separation of the rRNA genes, and appears to be a genomic trait shared by all members of the Methanomassiliicoccales.

The phylogenetic relationship of members of the Methanomassiliicoccales can be inferred from the 16*S* rRNA gene sequence (Seedorf *et al.* 2014). The relationship inferred from the 16*S* rRNA gene suggests that the organisms B10 and Mx1, both isolated from human faeces, and RumEn M1 isolated from bovine rumen are phylogenetically distant to the other sequenced members of the Methanomassiliicoccales. These three organisms group more closely with the environmental isolates of Methanomassiliicoccales (Figure 4.1B.), as reported previously (Sollinger *et al.* 2016). A FGD analysis reveals a similar profile (Figure 4.1A.) whereby the RumEn M1, B10 and Mx1 genomes group together in a gastrointestinal-tract cluster (GIT2) away from the ISO4-H5, BRNA1, 1R26, RumEn M2, ISO4-G11, Mx1201 and ISO4-G1 genomes in GIT1. The conservation in genome structure and gene order between different genomes is known as synteny (Passarge *et al.* 1999). When two genomes have a high degree

of synteny, a linear genome alignment can be observed across the entire length of the genome, and when an X-alignment is observed, as reported for *Mycobacterium leprae* and *Mycobacterium tuberculosis*, this indicates that matching sequences occur at the same distance from the origin but not necessarily on the same side of the origin (Eisen *et al.* 2000). This is often seen in synteny studies of moderately diverged genomes (Eisen *et al.* 2000). Using the ISO4-H5 genome as a reference, it was shown that a high degree of synteny is shared with the BRNA1 and Mx1201 genomes. An X-alignment was observed in these analyses and the identified syntenic breakpoints included a TPR repeat-containing protein (AR505_0536) and a region downstream of a an ATP-dependent DNA helicase (AR505_1369) and upstream of a ribosomal protein S27e (AR505_1370). This large-scale inversion encompases 955 kb of the genome.

The codon usage of the genomes was compared and found to be similar beween all members of the Methanomassiliicoccales, with one exception being the Mx1 genome. Mx1 was found to have a low %G+C (41.3%) content, and this was reflected in the codon usage of the amino acids, phenylalanine, asparagine, lysine, glutamate, tyrosine, alanine and cysteine, where Mx1 preferentially uses codons with less G or C content (Table A.4.1.). Despite this observed difference in codon usage, all of the Methanomassiliicoccales genomes maintain a conserved amino acid usage pattern. An analysis of the genome nucleotide similarity using the ISO4-H5 genome as a reference, highlighted the presence of a 25,429 bp region with 99.96% sequence identity and a 22,621 bp region with 99.26% sequence identity to the draft genome sequence of ISO4-G11. Eight adhesin-like protein genes, six ABC transporter-related genes, an ATP-dependent DNA helicase gene and 15 hypothetical protein genes are predicted within these two regions. This high level of sequence identity was not observed between other Methanomassiliicoccales genomes analysed. The ISO4-G11 genome was assembled from an enrichment culture and is currently in draft form, as such, the presense of such regions will need to be investigated further in the ISO4-G11 assembly.

All of the genomes analysed are predicted to encode a full set of tRNAs corresponding to 21 amino acids, except the incomplete RumEn M2 that lacks $tRNA^{Pyl}$. The tRNAs which contain introns differ between genomes. Intron-containing tRNAs are found in all three kingdoms of life (Yoshihisa 2014). The RumEn M1, RumEn M2, Mx1 and B10 genomes are predicted to possess a high number of intron containing tRNAs (eight, nine, seven and eight respectively), with conserved intron-containing tRNAs ($tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Cys}$ and $tRNA^{Ser}$), while the other strains possess a low number of intron containing tRNAs (three or four) with conserved

intron containing $tRNA^{Met}$. All strains contain a conserved set of two intron-containing tRNAs ($tRNA^{Trp}$ and $tRNA^{Tyr}$). While Mx1201 is closely related to the genomes of ovine rumen origin based on 16*S* rRNA gene inferred phylogeny, the MpT1 genome is more distantly related. The identical intron-containing tRNAs between MpT1 and Mx1201 suggests that both phylogeny and environment could contribute to the distribution of intron-containing tRNAs. RumEn M2 is clustered in GIT1 based on 16*S* rRNA gene inferred phylogeny, the distribution of intron-containing tRNAs is highly similar to RumEn M1, B10 and Mx1 of GIT2. This suggests RumEn M2 may belong to a currently under-represented clade of Methanomassiliicoccales and differ to other members of GIT1.

All except one genome of the Methanomassiliicoccales examined are predicted to encode and utilise the 22nd amino acid, pyrrolysine. The incomplete RumEn M2 genome is predicted to lack the pyrrolysine biosynthesis genes and all methyltransferases that incorporates pyrrolysine. The biosynthetic machinery of pyrrolysine is organized as a highly conserved operon (Gaston *et al*. 2011b). This operon structure is conserved among the genomes analysed, along with the surrounding methylamine utilisation genes (Figure 4.4.). This co-localisation is important because pyrrolysine is thought to be required for the catalytic activity of MtmB, MtbB and MttB (Krzycki 2004). The genes containing in-frame amber codon(s) have been reported for the genomes of B10, Mx1 and Mx1201 (Borrel *et al.* 2014). In this study, the in-frame amber-containing genes of the 11 genomes were examined and then classified into three categories: Class 1 genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2 genes that utilised the amber codon as a stop codon; Class 3 genes with uncertain amber codon usage due to lack of homologous genes. From this investigation it is clear there is a distinct difference in the predicted pyrrolysine incorporation amongst the genomes analysed (Figure 4.5.). The genomes of RumEn M1, RumEn M2, B10 and Mx1 encode a high number of genes (13%, 8.2%, 11.3% and 5.3% respectively) with an in-frame amber codon (Table 4.4.) and a small number of Class 1 genes predicted to incorporate pyrrolysine. This is in comparison to the other genomes which have a lower number of genes (1.8% to 4.1%) predicted to have an in-frame amber codon (Table 4.4.), but a higher number of Class 1 genes predicted to incorporate pyrrolysine (Figure 4.5.). Based on this analysis, two separate genomic profiles of pyrrolysine incorporation may exist within the gut environment and correspond to the two GIT clusters identified based on FGD analysis, with the exception of RumEn M2. The RumEn M2 genome contained no pyrrolysine biosynthesis genes, however, its genome is only 94.2% complete, it is likely the pyrrolysine biosynthesis genes are absent

from the currently sequenced genome, but may be present once the genome is completed and circularised. This assumption is supported by two factors in addition to the predicted usage of amber codon. i) The absence of any methyl-substrate utilising methyltransferase suggests a vital part of the genome is missing, the current genome indicates it is incapable of carrying out all known methanogenesis pathways. The methylamine methyltransferase genes are typically closely located with the pyrrolysine biosynthesis genes, therefore both the methylamine utilising genes and the pyrrolysine biosynthesis genes may present itself once the genome is completed. ii) A RamA homologue was predicted in the RumEn M2 genome, which is required to activate the corrionid protein utilised by the methylamine methyltransferases, therefore this study analysed RumEn M2 genome in assumption that the full genome would contain a pyrrolysine biosynthesis operon.

For the Class 2 genes predicted to use the amber codon as a stop codon (10, 6, 5, 6, 37, 26 and 13 genes for ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26, Mx1201 and MpT1 respectively), the distance between the amber codon and the next opal/ochre stop codon is usually short, suggesting the additional 3' amino acids are unlikely to affect the function of the gene product (Tables 3.6., A.4.2., A.4.3., A.4.4., A.4.5., A.4.8.). These results indicate that even if every predicted gene with an in-frame amber codon incorporates pyrrolysine due to a lack of regulation, the genomes of ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26 and Mx1201 are unlikely to suffer any adverse consequences, as the Class 2 genes in these genomes are predicted to be non-essential. For example, in the ISO4-H5 genome, nine IS elements and one adenylate kinase gene are predicted to utilise the amber codon as a stop codon. The absence of any pyrrolysine regulation in the ISO4-H5 organism would allow production of an extended protein past the amber codon, which may be misfolded and dysfunctional. The production of dysfunctional IS elements and an adenylate kinase may cost the ISO4-H5 organism some energy, but a second copy of an adenylate kinase gene in the ISO4-H5 genome is likely to keep the metabolism of ISO4-H5 functionally intact. For organisms such as B10 and Mx1, it has been proposed that they are able to distinguish the genes that require pyrrolysine to function, and as such, selectively incorporate the pyrrolysine, whereas all other genes with an in-frame amber codon are regulated to use the amber codon as a stop codon (Borrel *et al.* 2014). This analysis suggests that some members of the Methanomassiliicoccales may employ a mechanism of regulation of amber-codon suppression, whereas others do not, but rely on constitutive expression of the pyrrolysine biosynthesis operon.

This difference in regulation is likely related to the phylogenetic divergence between members of the Methanomassiliicoccales as inferred from FGD analysis with RumEn M1, B10 and Mx1 clustered separately from the remaining genomes (Figure 4.1A.). The availability of methylamine substrates may also be linked to this difference in regulation. For example, the Mx1201 isolate is proposed to have evolved in an environment where methylamines are not limiting, and alongside methylamine producing microorganisms (Smith and Macfarlane 1996; Borrel *et al.* 2014). Colonic anaerobes of the genera *Clostridium, Bifidobacterium* and *Bacteriodes* have been shown to utilise choline and glycine betaine, producing methylamines and propylamines in the presence of fermentable carbohydrates (Allison and Macfarlane 1989; Smith and Macfarlane 1996). The high availability of methylamines may stimulate the constant expression of the pyrrolysine operon as Mx1201 evolved, and led to the low usage of the amber codon as a stop codon (Borrel *et al.* 2014). Similarily, methylamines are readily available in the rumen, as choline from plant sap (Maizel *et al.* 1956), cell membrane (Neill *et al.*, 1978) and betaine produced from chloroplasts (Rhodes and Hanson 1993) and are metabolised in the rumen (Neill *et al.* 1978; Mitchell *et al.* 1979). The availability of methylamines may have contributed to the routine use of pyrrolysine in rumen members of the Methanomassiliicoccales. The Mx1201 genome is known to contain essential genes requiring amber codon read-through to fully translate proteins (Borrel *et al.* 2014), which indicated that the constitutive expression of the pyrrolysine operon and pyrrolysine is essential for growth in Mx1201 (Borrel *et al.* 2014). Similar genes have been found in ISO4-H5, ISO4-G1, ISO4-G11 and BRNA1, such as a bifunctional phosphoglucose/phosphomannose isomerase (AR505_0560), aconitate hydratase (ISO4G1_0572), serine hydroxymethyltransferase (ISO4G11_0237) and carbamoylphosphate synthase (TALC_00554). All of these genes require pyrrolysine incorporation, which suggests the pyrrolysine operon may be constitutively expressed and essential for growth in these members of Methanomassiliicoccales.

Although the MpT1 and 1R26 genomes display a genomic profile of pyrrolysine incorporation akin to what is observed for ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, Mx1201, it does not possess genes encoding essential functions, which require amber codon read-through to be fully translated. The MpT1 and 1R26 genes with an in-frame amber codon are either hypothetical proteins, mobile elements or genes with a very short distance between the amber codon and the next stop codon, which suggests that MpT1 and 1R26 may not suffer adverse consequences without pyrrolysine usage regulation. The pyrrolysine usage profile of the MpT1 and 1R26 genomes places them between the group that regulates pyrrolysine usage, (RumEn M1, B10,

Mx1) and the group that minimizes in-frame amber codons (Mx1201, BRNA1, ISO4-H5, ISO4-G1 and ISO4-G11), which correlates with the phylogenetic tree inferred from 16*S* rRNA genes (Figures 4.1., 4.5.).

A variety of pyrrolysine regulation mechanisms have been observed previously, including substrate regulated pyrrolysine operon expression in *Acetohalobium arabaticum* and *Methanosarcina acetivorans* (Mahapatra *et al.* 2006; Prat *et al.* 2012), and use of different release factors in *M. barkeri* and *M. acetivorans* (Alkalaeva *et al.* 2009). The presence of a pyrrolysine IS within the *mtmB* gene has also been shown to regulate pyrrolysine usage in *M. barkeri* at the transcription level (Longstaff *et al.* 2007). *Acetohalobium arabaticum* has been observed to regulate pyrrolysine incorporation by presence of a substrate (Prat *et al.* 2012). It is unclear at this time whether regulation methods such as these operate within members of the Methanomassiliicoccales analysed.

Extracellular and surface-associated proteins are likely to play an important role in many essential interactions and adaptations of members of the Methanomassiliicoccales to their environment. In the Methanomassiliicoccales genomes analysed, an average of 4.3% of their ORFeomes are devoted to the process of protein export. ISO4-H5 is predicted to have the highest proportion of exported gene products, 5.6% of its ORFeome, while RumEn M2 is predicted to have the lowest proportion of exported gene products, 2.5% of its ORFeome. Proteins predicted to contain a signal peptide without a transmembrane helice or lipobox are thought to be secreted into the environment. The number of proteins predicted to be secreted into the environment range between 4 and 25 per genome for the genomes analysed. Extracellular proteins with domain repeats are often involved in binding to the cell (Cabanes *et al.* 2002), whereas those that contain a transglutaminase domain may be able to mediate adhesion to a binding partner (Ng *et al.* 2016). All genomes examined contain genes with predicted repeat domains (17, 6, 19, 9, 4, 4, 10, 8, 4, 27 and 10 genes for ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, RumEn M1, RumEn M2, 1R26, MpT1, Mx1201, Mx1 and B10 respectively), while genes containing a transglutaminase-like domain have been predicted in eight of the genomes, ISO4-G1, ISO4-G11, BRNA1, RumEn M1, 1R26, MpT1, Mx1201 and MpT1. These genes suggest that members of the Methanomassiliicoccales are capable of binding to other microorganisms or their hosts, similar to how *Mbb. ruminantium* $M1^T$ binds to protozoa (Ng *et al.* 2016). Amongst the secretome, there are two transporters with Pfam domains conserved but not limited to six rumen members of the Methanomassiliicoccales. These domains are the PF00924 mechanosensitive ion channel and PF01032 FecCD transport

family. The mechanosensitive ion channels are pressure sensitive ion channels involved in osmoregulation according to environmental conditions (Kloda and Martinac 2001), which may be an important adaptation to the rumen environment. Such transporters may help members of the Methanomassiliicoccales, who lack a cell wall structure, to cope with the osmotic pressure of the outside environment. The FecCD transport family is involved in iron transport (Staudenmaier *et al.* 1989). Methanogens depend on the electron transport chain to conserve energy from methanogenesis, Fe-S centres are present in Hdr, Fdxs involved in methanogenesis and other enzymes of various metabolic functions, therefore an efficient iron uptake system may be beneficial to the members of the order Methanomassiliicoccales. Iron has been found to act as an electron donor for methanogenesis in *Methanococcus maripaludis* KA1 (Uchiyama *et al.* 2010), however *M. maripaludis* KA1 utilises the hydrogenotrophic methanogenesis pathway that is absent in members of Methanomassiliicoccales, and therefore it is unlikely for Methanomassiliicoccales to utilise iron as an electron donor. The attempt to verify secretome predictions by a proteomic approach in three *Sulfolobus* spp. has revealed 64 proteins amongst the supernatant and cell surface subproteome, with cell surface proteins dominating the secreted proteins, furthermore, the majority of secreted proteins originated from the cell surface (Ellen *et al.* 2010). This suggests genuine protein secretion is a rare event in archaea, and experimental validation is required to assess predicted secreted proteins and their biological role.

Rumen methanogens are known to utilise hydrogenotrophic, acetoclastic and methylotrophic methanogenesis pathways to produce energy (Thauer 1998). The methylotrophic methanogens within the rumen environment include *Methanosarcina* spp., *Methanosphaera* spp. and members of the Methanomassiliicoccales. The *Methanosarcina* spp. are versatile methanogens containing cytochromes (Kühn *et al.* 1983) and methanophenazine (Abken *et al.* 1998), and rely on both a membrane bound and cytoplasmic heterodisulfide reductase (Buan and Metcalf 2010) in addition to a proton pump for energy conservation (Baumer *et al.* 2000). They are capable of carrying out hydrogenotrophic methanogenesis from H_2 and CO_2, acetoclastic methanogenesis by disproportionating acetate (Fournier and Gogarten 2008), and methyltrophic methanogenesis by disproportionating methanol (Muller *et al.* 1986), where the reducing potential is derived from oxidising one mole of methanol to reduce three moles of methanol to CH_4 (Thauer *et al.* 2008). *Methanosphaera* spp. are methylotrophic methanogens that possess the pathway of hydrogenotrophic methanogenesis, yet rely on methanol and H_2 for CH_4 production (Fricke *et al.* 2006). *Methanosphaera stadtmanae* has been shown to be

incapable of molybdopterin biosynthesis, which is required for an active formylmethanofuran dehydrogenase in hydrogenotrophic methanogenesis, making it unable to carry out hydrogenotrophic methanogenesis or to carry out disproportionation of methanol (Fricke *et al.* 2006). Members of the Methanomassiliicoccales are specialised in direct reduction of methylated compounds for methylotrophic methanogenesis, and are unable to disproportionate the methylated substrates like members of the *Methanosarcina* (Welander and Metcalf 2005) (Figure 4.13.). Members of the Methanomassiliicoccales distinguish themselves from *Methanosphaera* spp. by the absence of the F_{420} (E_0' = -360 millivolts (mV)), and the use of Fdx (-500 mV) as an electron carrier, which is more exergonic (Welte and Deppenmeier 2014; Lang *et al.* 2015), furthermore, the Hdr/Mvh are not predicted to be coupled to an energy converting-hydrogenase complex (Thauer *et al.* 2008), but rather are coupled to a F_{420}-dehydrogenase Fpo-like complex. As such, these variations in the methlytrophic methanogenesis pathway has allowed *Methanosarcina* spp. (0.02% abundance in the rumen environment), members of the Methanomassiliicoccales (10.4%), and *Methanosphaera* spp. (13.8%) (Seedorf *et al.* 2015) to co-exist and to find their particular niche within the rumen environment. Interestingly, the Tibetan yak was found to have 80.9% of its rumen methanogen population as members of the Methanomassiliicoccales (Huang *et al.* 2012), which could possibly be contributed by diet. This particular study sampled yaks that were fed on *Kobresia* pasture (Huang *et al.* 2012), a dominant genera of plants in the alpine meadow. *Kobresia tibetica* Maximowicz was found to contain high level of condensed tannin (4.79% dry matter) (Niu 2014). Tannin has been proposed to be inhibitory to methanogenesis (Goel and Makkar 2012), potentially through inhibition of H_2-producing protozoa (Makkar *et al.* 1995). In *Bos indicus*, treatment with condensed tannin increases the Methanomassiliicoccales population by 21.9% (Tan *et al.* 2011). Due to the electron bifurcation of the heterodisulfide reductase and the difference between reducing potential of F_{420} and Fdx (Buckel and Thauer 2013), members of ths Methanomassiliicoccales are predicted to have a lower H_2 threshold in comparison to *Methanosphaera* spp. and *Methanosarcinales* spp., which may explain the increased abundance of members of the Methanomassiliicoccales in response to a tannin-rich diet.

The predicted substrates of methyltrophic methanogenesis in Methanomassiliicoccales include methanol, mono-, di-, tri-, methylamines and dimethyl-sulfide (Borrel *et al.* 2014). Previously, methanol was considered to be a methyl substrate common between members of the Methanomassiliicoccales (Lang *et al.* 2015), however, in the analysis reported here, the comparison to BRNA1 ruled out methanol as a conserved substrate, as the methanol:corrinoid

methyltransferase *mtaB* and the corresponding *mtaC* necessary for methanol utilisation are not present in the genome of BRNA1. Only the genes involved in methylamine utilisation were conserved across all of the Methanomassiliicoccales genomes analysed, except RumEn M2 where the incomplete genome lacks known methyl-substrate utilising methyltransferase. All of the ovine rumen strains analysed and two bovine rumen strains are predicted to utilise mono-, di- and tri-methylamine as a substrate, and presumably this reflects the availability of methylamines from plant derived choline and betaine within the rumen (Neill *et al.* 1978; Mitchell *et al.* 1979). Choline is utilised by the rumen protozoan, *Entodinium caudatum*, which incorporates it into phosphatidylcholine, but does not use it for trimethylamine production (Bygrave and Dawson 1976). A choline-utilising gene cluster has been identified in the genera *Desulfovibrio, Clostridia, Streptococcus, Klebsiella*, and *Proteus* (Craciun and Balskus 2012), all of which can be found in the rumen (Huisingh *et al.* 1974; Widyastuti *et al.* 1992; Ghali *et al.* 2004; Zadoks *et al.* 2011). The genera *Streptococcus, Klebsiella* and *Proteus* have also been found inside rumen protozoa, *Entodinium caudatum* (White 1969). Betaine and choline have been shown to be utilised by *Clostridium sporogenes – Methanosarcina barkeri* co-cultures (Hippe *et al.* 1979; Naumann *et al.* 1983). Some sulphate utilising marine methanogens, such as *Methanococci* spp., can utilise betaine and choline (L'Haridon *et al.* 2014; Watkins *et al.* 2014). There is no genomic evidence to suggest that members of the Methanomassiliicoccales can use betaine and choline directly as substrates to produce methylamines. Therefore, a close association between betaine and choline degrading bacteria and members of the Methanomassiliicoccales may be advantageous. A member of the Methanomassiliicoccales has been found in co-culture with the anaerobic rumen fungus *Piromyces* sp. (Jin *et al.* 2014). The ovine rumen strains ISO4-H5, ISO4-G1 and ISO4-G11 were enriched with a strain of *Succinivibrio dextrinosolvens* (Jeyanathan 2010) however, previous studies have shown that this species does not utilise betaine (Gomez-Alarcon *et al.* 1982). Nevertheless, a strong association was found between *Succinivibrionaceae* and *Methanomassiliicoccaceae* in a global census of the rumen microbiome (Henderson *et al.* 2015). This association may be explained, at least partially, by the release of methanol by *Succinivibrio* spp. from the degradation of pectin (Dehority 1969), which may be utilised subsequently by members of the Methanomassiliicoccales as a substrate for methanogenesis (Henderson *et al.* 2015).

Methyl-CoM reductase is a core enzyme of methanogenesis, and the holoenzyme is composed of six subunits as an $\alpha_2\beta_2\gamma_2$ heterohexamer (Ellefson and Wolfe 1981). The genes that code for the isozymes Mcr and Mrt display different operon structures within a genome, *mcrBDCGA*

and *mrtBDGA* respectively (Friedrich 2005). Based on this observation, all members of the Methanomassiliicoccales genomes analysed encode the Mrt isozyme. In organisms which encode both Mcr and Mrt, the Mrt is predominant during early log phase of growth when H_2 and CO_2 are sufficient, whereas the Mcr predominates when growth is limited by available H_2 and CO_2 which also correlates with Mrt having a higher specific activity than Mcr (Rospert *et al.* 1990; Bonacker *et al.* 1993; Pihl *et al.* 1994; Reeve *et al.* 1997). The Mcr/Mrt are highly conserved among methanogens and the sequence of the α subunit is often used to infer phylogenetic relationships between methanogens (Luton *et al.* 2002; Friedrich 2005; Chaudhary *et al.* 2011). The phylogenetic analysis of McrA (Figure 4.13.) suggests the Methanomassiliicoccales Mrt has diverged significantly from the Mrt and Mcr identified from *Methanobrevibacter* and *Methanosphaera* species. A recent metagenome/metatranscriptome study has classified the Mcr/Mrt operons of ovine rumen methanogens into three groups, and represents but is not limted to Methanobacteriales Mcr, Methanobacteriales Mrt and Methanomassiliicoccales Mrt respectively, of which the increased transcription of *mcr/mrt*A gene in Methanobacteriales *mcr* group and the Methanomassiliicoccales *mrt* group has been observed in the high CH_4 emitting sheep (Shi *et al.* 2014). In this thesis, a conserved protein has been identified in the *mrt* operon of all 11 genomes, located immediately downstream of the *mrtA* gene. The gene is 91 to 96 bp in size, with conservation in amino acid identity ranging from 31.9% between 1R26 and B10 to 72.8% between ISO4-H5 and ISO4-G11, suggesting an involvement in the Mrt system. This gene will require structural studies to elucidate whether it is a new subunit of the Mrt complex or a protein somehow associated with Mrt (Figure 4.15.).

At the central active site of Mrt is a highly reduced tetrapyrrole, F_{430} (Ermler *et al.* 1997). Members of the Methanomassiliicoccales produce tetrapyrrole by using the glutamyl-tRNA via uroporphyrinogen-III intermediate by enzymes encoded by the *hemABCDL* and *corA* genes (Phillips *et al.* 2003; Lang *et al.* 2015). These genes are highly conserved between eight of the Methanomassiliicoccales genomes analysed, while the ISO4-G11 draft genome only possesses a *hemC* homologue, and all of the genes involved in tetrapyrrole biosynthesis are absent from the ISO4-G1 and RumEn M2 genomes. In *Pseudomonas fluorescens* and *Escherichia coli*, a mutation in the *hemA* gene renders these organisms auxotrophic to the tetrapyrrole biosynthesis intermediate aminolevulinate (Avissar and Beale 1989; Baysse *et al.* 2001). The absence of *hemA* suggests that ISO4-G1, ISO4-G11 and RumEn M2 are incapable of producing tetrapyrrole, and therefore are likely to require an external source of F_{430} or tetrapyrrole intermediates to survive.

CoM is an important co-factor required for the methanogenesis pathway as all methyl groups are transferred to CoM before being reduced to CH_4 (Thauer *et al.* 2008). The *comABCDE* genes are required for CoM biosynthesis (White 1985; Graupner *et al.* 2000; Graupner *et al.* 2000). All the Methanomassiliicoccales genomes sourced from the rumen and the MpT1 genome are predicted to lack *comABCDE*, which suggests they are incapable of CoM biosynthesis, and require an external supply of CoM to survive. The inability to produce CoM has also been observed in *Methanobrevibacter ruminantium* $M1^T$ (Leahy *et al.* 2010), which is predicted to be able to survive by uptake of external CoM from the rumen environment. Rumen fluid contains 10^7 to 10^9 archaea cells per mL (Hungate 1966), and the rumen archaeal community is composed almost exclusively of methanogens (Janssen and Kirs 2008). It has been shown that *Mbb. ruminantium* $M1^T$ requires rumen fluid for growth, and that the supplementation of 3 ng/mL of CoM in the absence of rumen fluid achieves half the maximal growth, indicating that rumen fluid can act as a source of CoM (Taylor *et al.* 1974). A further study has observed the uptake of CoM in *Mbb. ruminantium* $M1^T$ using labelled $H^{35}S$-CoM (Balch and Wolfe 1979). The availability of CoM in the rumen environment is likely to allow the seven rumen strains of Methanomassiliicoccales lacking CoM biosynthesis functions to survive within the rumen. Similarly, within the termite gut, CoM is available for uptake by organisms such as MpT1 (Lang *et al.* 2015). The differences between the environment in the termite gut (Brune 1998) and the rumen (Hungate 1966) suggests the loss of CoM biosynthesis genes occurred independently. Loss of function can confer adaptive advantage in *E. coli* (Hottes *et al.* 2013). Pairwise competition experiments in *E. coli* and *Acinetobacter baylyi* has shown, in a nutrient-rich environment, evolution favors the loss of biosynthetic genes and the uptake of metabolites that can circumvent the loss of biosynthetic functions (D'Souza *et al.* 2014). Similarly, the loss of the CoM biosynthesis pathway may also confer an adaptive advantage to members of the Methanomassliicoccales, and enhance their survival in the rumen and termite gastrointestinal environments.

Members of the Methanomassiliicoccales isolated from human faeces have the ability to produce CoM based on their genomic sequences. Methanogens account for approximately 10% of anaerobes in the human colon (Eckburg *et al.* 2005), of which 80% to 100% are *Methanobrevibacter smithii* (Matarazzo *et al.* 2012), which is capable of CoM biosynthesis (Samuel *et al.* 2007). Using the same rationale proposed by Lang *et al.* (2015) for MpT1, conceivably CoM can be expected in the human gut, and indeed CoM is listed in the Human Metabolome Databse as being present (http://www.hmdb.ca/ metabocard for Mesna (CoM)

HMDB03745). However, the three genomes of Methanomassiliicoccales isolated from the human gut have not lost their genes encoding CoM production. Perhaps the difference between the hosts' digestive systems could help explain the difference in CoM biosynthesis capability. Although the members of the Methanomassiliicoccales analysed in this study were isolated from human faeces, it is conceivable that the human colon also harbours members of the Methanomassiliicoccales. Indeed, members of Methanomassiliicoccales (EF628038, EF369488) has been detected in the human colon by *mcrA* gene based diversity analysis (Scanlan *et al.* 2008), although the *mcrA* genes were previously recognized as members of *Methanosarcinales*, subsequent phylogeny studies has verified them to be members of Methanomassiliicoccales (Paul *et al.* 2012; Borrel *et al.* 2013). The retention time of digesta is similar between ovine rumen and the human large bowel (Poppi *et al.* 1981; Van Soest 1984), however, the motion of digesta differs between the human colon and rumen (Hertz and Newton 1913; Phillipson 1939). The sheep rumen has a very regular cycle of contraction as well as regurgitation for rumination (Phillipson 1939), whereas the human colon is completely inactive most of the day with three to four rapid movements lasting only a few seconds (Hertz and Newton 1913). The movement of rumen digesta allows even distribution of CoM throughout the rumen, whereas the inactivity of the human colon may make environmental CoM difficult to diffuse and therefore difficult to access. However, CoM is a very small cofactor of 141 g/mol (Taylor and Wolfe 1974), which suggests the diffusion is only limited to some degree, and there may be a more fitting explanation. The cross-feeding of cofactors has been observed between *Treponema primitia*, *Serratia grimesii*, *Lactococcus lactis* and *Treponema azotonutricium* in the termite gut, where *S. grimessi* and *L. lactis* supplies folate to *T. primitia* (Graber and Breznak 2005), and *T. primitia* supplies biotin, pyridoxal phosphate and CoA to *T. primitia* (Rosenthal *et al.* 2011). Similar cross-feeding may occur between the CoM auxotrophic MpT1 and CoM producing methanogens in the termite gut. The methanogen within human gastrointestinal environment is predominantly *Mbb. smithii* (Matarazzo *et al.* 2012), and it tends to be highly transcriptionally active compared to other species (Franzosa *et al.* 2014). Although *Mbb. smithii* produces CoM, members of Methanomassiliicoccales B10, Mx1 and Mx1201 may not be able to consistently and reliably acquire CoM from *Mbb. smithii*, as biosynthetic pathway for small metabolites tend to be underexpressed (Franzosa *et al.* 2014), possibly CoM biosynthesis pathway is also underexpressed and there is no excess CoM to facilitate cross-feeding, which may explain why the human strains retained the genetic capability for CoM biosynthesis.

The purpose of methanogenesis is to generate membrane potential and produce energy in the form of ATP (Khakh and Burnstock 2009). Methanogens are strictly chemiosmotic, and only use the archaeal A_1A_0-ATP synthase, which consists of nine different subunits (Muller *et al.* 1999). The conservation of A_1A_0-ATP synthase genes between methanogens has been exploited as a potential drug target for CH_4 mitigation (Aung *et al.* 2015). The A_1A_0-ATP synthase of *Mbb. ruminantium* $M1^T$ uses a Na^+ gradient for energy generation, but a H^+ gradient has been found to be used when it is more favourable (McMillan *et al.* 2011). A conserved PET sequence motif in the rotor subunits is associated with H^+ transport in methanogens (Muller *et al.* 1999). A PET motif has been identified in the *ahaK* genes at amino acid 62 – 64 in all the Methanomassiliicoccales genomes analysed, therefore a H^+ gradient may drive energy generation in members of the Methanomassiliicoccales. This is consistent with the H^+-translocating Fpo-like complex conserved across the Methanomassiliicoccales strains, as well as the lack of the Na^+ gradient-generating Mtr of the hydrogenotrophic methanogenesis pathway (Lienard *et al.* 1996). However, the presence of Na^+/H^+ antiporter and sodium:solute sympoters such as sodium/proline symporter PutP and sodium bile acid symporter suggests Na^+ gradient may have a role in transportation of solute in Methanomassiliicoccales.

Members of the Methanosarcinales rely on the Fpo complex and $F_{420}H_2$ for energy conservation (Baumer *et al.* 2000) while members of the Methanomassiliicoccales depend on the Fpo-like complex for H^+ gradient generation. The FpoF subunit is involved in interacting with $F_{420}H_2$ (Welte and Deppenmeier 2011), however, all of the Methanomassiliicoccales genomes, lack the genes required to produce F_{420} as well as a gene encoding the FpoF subunit (Borrel *et al.* 2014). FpoO is proposed to be involved in mephanophenazine electron transfer (Welte and Deppenmeier 2011). Genes encoding the subunit FpoO are also absent in all the Methanomassiliicoccales genomes, and there are no genes encoding mephanophenazine biosynthesis. Members of the Methanomassiliicoccales therefore require an alternative electron carrier in the absence of the F_{420}. In *Methanosaeta* spp. which lack the *fpoF* gene, it has been proposed that Fdx can act as an alternative electon carrier (Welte and Deppenmeier 2011). Therefore, it is likely that Methanomassiliicoccales Fdxs are involved in the electron transfer between Hdr, Mvh and the Fpo-like complex. The GIT2 Methanomassiliicoccales RumEn M1, B10 and Mx1 also possess the energy converting-hydrogenase complex (*echABCDEF*) that could act as an alternative H^+ transporter, however, this may have an anaplerotic role to replenish intermediates (Lie *et al.* 2012).

The main source of reducing potential for CH_4 formation in the Methanomassiliicoccales is H_2. Ethanol has been known to provide additional reducing potential for methanogenesis (Berk and Thauer 1997; Leahy *et al.* 2010). The NADP-dependent alcohol dehydrogenase is coupled to the F_{420}-dependent NADP reductase and a F_{420} dependent methylene-H_4MPT dehydrogenase to allow ethanol to be utilised (Berk and Thauer 1997). A type IV iron containing alcohol dehydrogenase (*adh*) is predicted in the genomes of ISO4-H5, ISO4-G1, ISO4-G11, BRNA1, 1R26 and Mx1201. However, the F_{420} and the F_{420} dependent NADP reductase, and F_{420} dependent methylene-H_4MPT dehydrogenase involved in supplying additional reducing potential to methanogenesis are absent in all Methanomassiliicoccales genomes analysed, which suggests the utilisation of ethanol is not linked to methanogenesis. Ethanol may be utilised as an alternative carbon source, but further work is required to define such a role.

The genomes of BRNA1, RumEn M1, RumEn M2, 1R26 and Mx1201 are predicted to encode *ldhA* genes. The Ldh enzyme converts lactate to pyruvate (Furukawa *et al.* 2014), thereby, lacate may serve as a potential carbon source. However, lactate is unlikely to be an alternative carbon source for methanogens (Mackie and Heath 1979), as within the rumen environment, it is expected that methanogens would be unable to outcompete rumen bacteria in the fermentation of lactate. The *Desulfovibrio vulgaris* was only able to catabolise ethanol and lactate in the presence of *Methanobacterium bryantii* MoH, as the free energy change of the catabolic reaction is more negative at low H_2 partial pressure, and H_2 utilising *M. bryantii* made this reaction favourable (Bryant *et al.* 1977). Conversely, lactate has been proposed as an electron donor to the Hdr system of methanogenesis in members of the methanogenic order Bathyarchaeota (Evans *et al.* 2015). A lactate dehydrogenase-like FAD-containing dehydrogenase (*glcD*) gene is co-located with *hdrD* genes in the genome sequence of BA1, a member of Bathyarchaeota. The iron sulphur cluster binding motifs of the *hdrD* gene may allow heterodisulfide to utilise lactate as a source of electrons (Evans *et al.* 2015). A *glcD* gene homologue has been found in the Methanomassiliicoccales genomes analysed (AR505_263, ISO4G1_0340, ISO4G11_0756, TALC_01288, MMALV_13780, MMINT_03090, WP_019177724.1, Mpt1_c09060) that are not co-localised with the *hdrD* gene, experimental validation is required to confirm whether lactate can be utilised by members of Methanomassiliicoccales.

There are 17 protein-coding genes hypothesized to be marker genes for methanogenic archaea, and they have been termed methanogenesis marker proteins (MMP) (Gao and Gupta 2007). The Methanomassiliicoccales genomes analysed showed conservation of the MMPs as

expected, however, they all lack MMPs 9, 10, 12 and 14. Furthermore, BRNA1, 1R26 and Mx1201 lack MMP16, RumEn M2, 1R26 and B10 lacks the MMP2, RumEn M2 lacks the MMP13 and B10 lacks the MMP 4. It is clear from this analysis that not all 17 MMPs can be considered markers of all the methanogenic archaea as previously thought (Gao and Gupta 2007).

Based on the genome sequences of members of the Methanomassiliicoccales, central carbon metabolism is predicted to involve acetate acquisition as the primary carbon source, and use of an incomplete TCA cycle followed by complete gluconeogenesis and pentose phosphate pathways. The acetyl-CoA synthetase gene is conserved in all genomes analysed, therefore acetate is predicted as the common carbon source. The RumEn M1, RumEn M2, Mx1 and B10 genomes contain genes encoding methylene tetrahydrofolate reductase (MetVF), methyltetrahydrofolate:corrinoid methyltransferase (AcsE), CO dehydrogenase (AcsA) and acetyl-CoA decarbonylase (CdhDE), which suggests Mx1 and B10 may be able to produce acetyl-CoA from formate and CO_2. MpT1 may not be able to convert CO_2 to CO, as the primary function of AcsB is in electron transfer (Morton *et al.* 1991), and AcsB homologues can easily be confused with similar electron transfer enzymes. As formate is utilised as an electron donor for hydrogenotrophic methanogens in the rumen (Liu and Whitman 2008), the rumen members of the Methanomassiliicoccales may have evolved to not utilise formate and thus avoid competition with hydrogenotrophic methanogens.

Methanogens are known to have an incomplete TCA cycle (Ekiel *et al.* 1985; Sprott *et al.* 1993), and the Methanomassiliicoccales appear no different as they are predicted to have various genes absent from the TCA pathway (Figure 4.19., Table A.4.16.). Despite this, the TCA cycle is still required for gluconeogenesis and nitrogen uptake, in addition to providing intermediates for amino acid biosynthesis. The gluconeogenesis pathway is highly conserved amongst the members of Methanomassliicoccales, only MpT1 appears to be missing the *pgi* gene encoding the phosphoglucose/phosphomannose isomerase, MpT1 is predicted to utilise an alternative, but unknown, enzyme instead of Pgi (Lang *et al.* 2015). The *pgi* gene exists as a pseudogene in the RumEn M2 and 1R26 genomes due to an interupted coding sequence, as both genomes are incomplete, this may be caused by sequencing error or a frameshift. The analyses in this thesis also predicts that members of the Methanomassiliicoccales carry out autotrophic carbon fixation, i.e. the conversion of CO_2 to organic carbon (Berg *et al.* 2010). Ribulose 1,5-bisphosphate carboxylase (RubisCO, RbcL) is the key enzyme in CO_2 fixation and functional type III RubisCO enzymes have been discovered in *Methanococcus jannaschii*

and *Pyrococcus kodakaraensis* (Ezaki *et al.* 1999; Watson *et al.* 1999). The archaeal type III RubisCO has been found to function alongside AMP phosphorylase (DeoA) and ribose-1,5-bisphosphate isomerase (E2b2) in AMP metabolism (Sato *et al.* 2007). The RubisCO operon *rbcL*, *e2b2*, *deoA* is found to be conserved across all the Methanomassiliicoccales genomes, and suggests they are capable of fixing CO_2 while recycling AMP. In addition to incorporating carbon fixation to purine recycling, another enzyme in the purine biosynthesis pathway contributes to conservation of cell energy. The genomes of the Methanomassiliicoccales analysed are predicted to use phosphoribosylaminoimidazole carboxylase (Ade2) instead of N^5-carboxyaminoimidazole ribonucleotide synthetase (PurK) to produce NCAIR (N^5-carboxyaminoimidazole ribonucleotide) (Borrel *et al.* 2014). This reduces the ATP required for purine biosynthesis and perserves energy. However, the decrease in energy expenditure may not necessarily correlate with an increase in biomass of the Methanomassiliicoccales, and it remains unknown whether the additional energy would be utilised on growth, maintenance or, be dissipated (Russell and Cook 1995).

An analysis of the amino acid biosynthesis pathways in genomes of the Methanomassiliicoccales revealed that several may not be able to make certain amino acids (Tables 4.10., A.4.17.). The Methanomassiliicoccales represent a new order of methanogens and its genes are expected to be different to other methanogens. Therefore it is possible that in instances where a single gene is missing from a pathway, a new gene of unknown function may fulfill the role of the missing gene. In instances where an entire biosynthetic pathway is absent, such as the tryptophan biosynthesis pathway predicted in the genomes of ISO4-G1 and RumEn M2, then it is likely that such organisms require an external source of tryptophan to survive. The Mx1, RumEn M1, RumEn M2 and ISO4-G1 genomes are missing the genes required for a complete tryptophan biosynthetic pathway. The genes required to produce tryptophan are also absent in two different genomes of the Methanomassiliicoccales. It has been established that particular traits often evolve repeatedly when the populations are exposed to similar ecological conditions (Schluter *et al.* 2004), which suggests the loss of tryptophan biosynthesis may have evolved in parallel in ISO4-G1, RumEn M1, RumEn M2 of rumen and Mx1 of human gut due to the similarity of the gastrointestinal environment. This is possibly due to the high cost of tryptophan biosynthesis, which requires 78 moles of ATP per mole of tryptophan (Bender 2012). Tryptophan is a scarce but important amino acid in many organisms, as tryptophan confers a strong structural influence by its affinity to aromatic and branched-chain amino acids, repelling negatively charged amino acids as well as stabilising β–hairpin

structures by Trp-Trp binding (Santiveri and Jimenez 2010). Bovine rumen fluid contains 36 µg/mL of tryptophan (Candlish *et al.* 1970). Tryptophan is energy intensive to make, so if cells can uptake tryptophan from the environment, the availability of external tryptophan permitted the removal of unrequired pathway, which conserves energy (Zamenhof and Eichhorn 1967). However, energy is not the only contributing factor to the loss of genes in evolution (Dykhuizen 1978), further investigation is required to clarify the evolutionary driving force that removed the tryptophan biosynthesis pathway. The auxotrophic Methanomassiliicoccales would likely import the deficient amino acids from the environment via an amino acid permease. An amino acid permease gene is predicted in all genomes analysed except RumEn M1, RumEn M2 and MpT1 (Table A.4.17.). All the genomes analysed are also predicted to possess a sodium/proline symporter *putP* gene. All genomes are predicted to be capable of proline biosynthesis.

Comparison between Methanomassiliicoccales genomes can reveal conserved genes important for the organisms' growth and survival as well as those genes that might be important for niche adaptation. This study included six completed genomes and five incomplete genomes of Methanomassiliicoccales, only the completed genomes were analysed to identify the core genome and environmental conserved gene families. The core genome of the six completed Methanomassiliicoccales genomes examined consists of 415 gene families, while the order level pan-genome encompasses 8,767 gene families across 11 genomes. The Methanomassiliicoccales strains originated from four environments, the termite gut (1 strain), human faeces (3), and the bovine (4) and ovine rumen (3). In this study, the analysis has been focused on genomes of rumen origin. There are 18 gene families conserved between the three completed genomes of rumen strains, and 41 gene families conserved between the two completed genomes of ovine rumen strains, which might shed light on the genes involved in adaptation to the rumen environment. Each genome also contains its own unique gene families, from 352 gene families in BRNA1 to 1,137 gene families in B10. These unique gene families can offer insight into how these members of the Methanomassiliicoccales may establish their own niche in the presence of similar organisms within the same environment.

The core genome of Methanomassiliicoccales was classified according to the COG database (Table 4.7.). A total of 89 gene families are implicated in energy conservation, which includes the genes required to utilise monomethylamine for methanogenesis, but not other substrates. This suggests monomethylamine is the conserved substrate between members of Methanomassiliicoccales. In addition, the aforementioned RubisCo operon is also identified

within the core genome, which adds weight to the potential importance and functionality of this autotrophic carbon fixation operon.

There are 18 gene families identified as conserved amongst the genomes of rumen origin (Table 4.8.), including but not limited to, a DNA polymerase IV, a methylthioadenosine nucleosidase MtnN, a MIP (major intrinsic protein) transporter, a xanthine/uracil permease family protein and a dolichol kinase. DNA polymerase IV is a Y-family DNA polymerase rarely observed in archaea (Ohmori *et al.* 2001). Due to its lack of proof-reading activity, DNA polymerase IV is more error-prone than other DNA polymerases, which provides this enzyme the ability to replicate DNA across damaged bases, also known as translesion replication (Ohmori *et al.* 2001). This particular enzyme was characterised in *Sulfolobus solfataricus*, and found to be slightly more error-prone than *Taq* polymerase (Boudsocq *et al.* 2001), as it is able to bypass all lesions tested, which was thought to be a selective advantage. The DNA polymerase IV may rescue stalling of DNA replication by translesion replication in members of the rumen Methanomassiliicoccales, allowing more efficient replication and thus cell growth. This may be an adaptation to the rumen, where microorganisms are removed continuously *via* dilution and rumen turnover.

The conservation in the cell envelope related gene, dolichol kinase, within the rumen-dwelling Methanomassiliicoccales might be associated with adaptation to their environmental. The cell envelope of Methanomassiliicoccales is likely similar to the glycocalyx envelope of Thermoplasma, with *N*-glycosylated mannose residues and glycolipids (Albers and Meyer 2011). Dolichol kinase is involved in *N*-glycan biosynthesis (Bernstein *et al.* 1989) which suggests the rumen members of the Methanomassiliicoccales have evolved a specialized modification to the cell envelope in order to adapt to the rumen environment.

There are 41 gene families conserved between ISO4-H5 and ISO4-G1 genomes of ovine rumen, including a predicted cardiolipin synthase A, whose biosynthesis is described in Chapter 3, Section 3.2.4. The methanogen cell envelope has been found to be SDS- and protease-resistant (Kandler and Konig 1978) due to pseudomurein and S-layer, however, *M. luminyensis* B10 has been found to be SDS-sensitive (Dridi *et al.* 2012), and neither electron micrographies nor genomic information suggests the presence of pseudomurein, and observations thus far do not support the presence of an S-layer ultrastructure, which makes the Methanomassiliicoccales envelope vulnerable to physical and osmotic stress. Cardiolipin has been associated with osmotic stress tolerance (Romantsov *et al.* 2009), as it has the ability to

tightly bind a transmembrane protein complex, which can assist in salinity tolerance in halophiles (Corcelli 2009). However, cardiolipin synthase of different gene families are also found in all three members of Methanomassiliicoccales from human sources, as well as other methanogens (Yoshinaga *et al.* 2012). Four transporters implicated in cation transport are also conserved between the three ovine rumen Methanomassiliicoccales, including a Na/Pi-co-transporter, a Na^+/H^+ antiporter family protein as well as two small multidrug resistance proteins that share homology to an ammonium compound efflux pump. The rumen environment has approximately 115 mM of Na^+ (Saleem *et al.* 2013) for both cattle and sheep, whereas phosphate concentrations vary between 1.85 mM to 12 mM in sheep (Wadhwa and Care 2002), and 1 mM to 4 mM in cattle (Saleem *et al.* 2013). The conservation in cation transporters and difference in cardiolipin synthase could potentially aid members of the Methanomassiliicoccales to survive in the rumen environment. The lack of a cell wall makes members of Methanomassiliicoccales vulnerable to osmotic stress, and the conservation of the cation transport genes suggests these genes are involved in maintaining the cellular osmotic balance within the rumen environment.

A large number of gene families (551) were identified as unique to ISO4-H5, including a cluster of genes predicted to be involved in exopolysaccharide production (Chapter 3, Section 3.2.4) and 30 adhesin-like proteins. Members of the Methanomassiliicoccales have been found in acid mine drainage biofilms, where they are predicted to rely heavily on exopolysaccharides for adherence (Mendez-Garcia *et al.* 2014). This particular cluster of putative ISO4-H5 exopolysaccharide producing genes and adhesin-like proteins, may play a role in adherence to solid substrates, host cells or other microorganisms within the rumen. Two adhesin-like protein encoding genes (AR505_1559, AR505_1560) are predicted to contain Listeria Bacteroides repeat domains (PF09479) and a cohesin domain (PF00963) (Table 3.10.), which may be involved in the adherence of ISO4-H5 to H_2 producing bacteria with cellulosome.

Another set of genes unique to ISO4-H5 were the *dndCD* genes (AR505_0884, AR505_0885) required for DNA phosphorothioation (Wang *et al.* 2007), the process in which DNA is modified by replacing the oxygen in the DNA back bone with sulfur, confering nuclease resistance (Wang *et al.* 2007). The other genes of the *dnd* operon (*dndABE*) were not found in ISO4-H5, however *dndA* is homologous to cysteine desulfurase, which is found in ISO4-H5 (AR505_1027), and DndB likely stabilizes secondary DNA structures during phosphorothioation, while the function of DndE remains unknown (Chen *et al.* 2010). In addition, a sulphate permease gene *sulP* (AR505_1686) is present which may encode supply

of the sulfur molecule for phosphorothioation. Therefore, it seems possible that ISO4-H5 possesses the genes encoding phosphorothioation, which may explain why all the previous attempts to verify genome assembly via PFGE resulted in a heavily degraded smear - when the voltage was applied in the gel electrophoresis, the amine group in Tris (trisaminomethane) buffer was activated, a peracid oxidant is generated that carries out nucleolytic attack at the modified DNA sites (Chen *et al.* 2010).

Other noteworthy genes unique to ISO4-H5 include a DNA alkylation repair enzyme AR505_1309 and a 6-*O*-methylguanine DNA methyltransferase *ogt* (AR505_0448) gene that repairs alkylating lessions, and a phosphoglycolate phosphatase (AR505_0604) thought to salvage the phosphoglycolate produced by the RubisCO operon during photorespiration. The presence of this gene is unexpected as ISO4-H5 is an obligate anaerobe, and the RubisCO most likely only catalyses carbon fixation that produces 3-phospho-D-glycerate, and not photorespiration.

The ISO4-G1 genome contains 526 unique gene families (Table A.4.13.) and it is noteworthy that it has a high number of predicted ferric ion transporters. In *Methanosarcina barkeri* ferric ions directly inhibited methanogenesis, as ferric ion is preferentially reduced over the methyl-substrates (Bodegom *et al.* 2004). There is no experimental evidence to suggest the high number of transporters predicted to be involved in iron transporter can be correlated to the utilisation of iron in ISO4-G1, but as Fe-S centre containing enzymes occupy important metabolic roles in methanogen, the uptake of iron may benefit ISO4-G1. Additionally, there is a large NRPS identified as unique to the ISO4-G1 genome (Kelly *et al.* 2016). Archaeal NRPSs were first described in *Methanobrevibacter ruminantium* $M1^T$ (Leahy *et al.* 2010), and subsequently discovered in other classes of *Methanomicrobia* and *Methanobacteria* (Wang *et al.* 2014) and *Mbb. millerae* SM9 (Kelly *et al.* 2016). NRPSs are responsible for biosynthesis of small peptides, and a number of peptides produced by NRPSs are currently used as antibiotics, including penicillin and bacitracin (Witting and Sussmuth 2011). NRPSs are produced by consecutive condensation of substrates, including amino acids, various fatty acids, hydroxyl acids and nonproteinogenic amino acids (Caboche *et al.* 2008). The NRPSs are composed of distinct modules; an adenylation domain that activates the amino acids, a peptidyl carrier domain that propagates the peptide chain, and a condensation domain that catalyses the condensation of the amino acids (Strieker *et al.* 2010). Keeping a gene as large as a NRPS in the genome is energetically expensive therefore it seems likely that the gene provides certain advantages for ISO4-G1 survival within the rumen. However the actual benefit of NRPS within

the rumen environment, are not currently known, but they may possibly mediate interactions between other rumen microbes.

There were 465 gene families unique to the ISO4-G11 genome (Table A.4.10.), of which 354 genes were COG classified as poorly characterized without predicted function. There were 14 genes involved in membrane biogenesis, including a cluster of five genes that are predicted to be involved in dTDP-L-rhamnose biosynthesis, six genes predicted to be involved in glycosylation of lipopolysaccharide, and two Mur ligase genes possibly involved in glycopeptide formation. Glycosylated dTDP-L-rhamnose has been found attached to the *S*-layer glycoprotein in *Haloferax volcanii* (Kaminski and Eichler 2014) and the presence of these genes in ISO4-G11 suggests it may possess a cell envelope with a glycosylated exopolysacharide. A 3,644 aa adhesin-like protein (ISO4G11_1719) is homologous to a filamentous hemagglutinin. Filamentous hemagglutinins are utilised by *Bordettela pertussis* for adherence to the host respiratory tract (Kimura *et al.* 1990), which suggests ISO4-G11 may use this adhesin for adherence to the ruminant host.

There were 352 gene familes unique to the BRNA1 genome, of which 254 gene families were COG classified as poorly characterized without predicted function. Like ISO4-G11, several unique BRNA1 genes have been predicted to be involved in cell membrane biogenesis. Amongst the unique gene families are two genes homologous to phosphoenolpyruvate phosphomutase *fom1* (TALC_00083) and phosphonopyruvate decarboxylase *fom2* (TALC_00082). These genes form part of the fosfomycin biosynthesis pathway in *Streptomyces* spp. (Kim *et al.* 2012), and catalyse the proton dependent conversion of phosphoenolpyruvate to phosphonoacetaldehyde (Kim *et al.* 2012). A 2-aminoethylphosphonate aminotransferase homologue (TALC_00081) clusters with *fom1* and *fom2* genes, which could interconvert phosphonoacetaldehyde to 2-aminoethylphosphonate (Dumora *et al.* 1983). The 2-aminoethylphosphonate has been identified as a constituent of phospholipid in the ciliate protozoa *Tetrahymena pyriformis* (Thompson 1969) and its mitochondria (Jonah and Erwin 1971). In addition, two genes were annotated as glycosyltransferases involved in cell wall biogenesis (TALC_00068 and TALC_00105). The BRNA1 gene complement suggests it is capable of producing 2-aminoethylphosphonate, but whether it can act as a constituent of its cell membrane phospholipid requires further investigation.

4.4. Conclusions

Members of the order Methanomassiliicoccales have been discovered from a wide range of environments, however only members of the Methanomassiliicoccales isolated from the gastrointestinal environments of animals have been cultured and genome sequenced thus far. The sequencing of four members of the Methanomassiliicoccales isolated from the rumen has provided valuable insights into the lifestyle and adaptation of these methanogens to the rumen. All Methanomassiliicoccales genomes of rumen origin are predicted to have lost the genetic capability to produce the essential CoM cofactor, and rely on uptake of CoM directly from the rumen environment. This streamlining of their genomes likely confers a competitive edge over CoM-producing strains due to savings in energy production and in speed of replication. Furthermore, the RumEn M2, ISO4-G1 and ISO4-G11 genomes are predicted to have lost the capability to produce the cofactors tetrapyrrole and F_{430}, and requires external uptake of these cofactors to survive. Again, if these cofactors can be easily obtained from the rumen environment, it could reduce energy expenditure by these organisms and offer them an advantage over their co-factor-producing competitors. The continual dilution and turnover of rumen contents makes growth rate and replication speed key factors in persistence within the rumen. The conservation of DNA polymerase IV in rumen Methanomassiliicoccales genomes may also reflect these selection pressure, such that genome replication accuracy is sacrificed for speed of replication.

The rumen members of the Methanomassiliicoccales are predicted to primarily utilise methylamines, which are likely derived from choline and betaine breakdown in the rumen. Comparison of the completed genomes has revealed that mono-methylamine is the only common substrate for methylotrophic methanogenesis, which was also identified amongst incomplete genomes with the exception of RumEn M2. ISO4-H5, ISO4-G11, 1R26, Mx1201 and B10 have evolved to utilise a wide variety of methyl-substrates, while ISO4-G1, BRNA1, RumEn M1, Mx1 and MpT1 utilise only a small selection of methyl-substrate. The utilisation of methylamines requires pyrrolysine incorporation in the methyltransferase enzymes, and the biosynthesis and incorporation of pyrrolysine is conserved across all genomes analysed of the members of the Methanomassilicoccales, except the incomplete RumEn M2 genome. However, it appears that members of GIT1 cluster and GIT2 likely regulate the amber-codon read through differently.

Very little is known about cell envelope biosynthesis in the Methanomassiliicoccales. They are thought to lack a rigid S-layer and pseudomurein cell wall, which suggests the cell envelope is vulnerable to osmotic stress and requires proficient ion channel to maintain cellular osmolarity. The presence of cardiolipin and dolichol kinase genes in their genomes suggest the cell envelope is similar to the glycocalyx envelope of members of the Thermoplasma. This envelope is likely flexible and potentially covered with *N*-glycosylated mannose residues, glycolipids and other exopolysaccharides that aid in binding to the host or other beneficial microorganisms. However, this would require experimental verification in ISO4-H5.

The 11 Methanomassiliicoccales representatives from four different environments revealed genomic elements that encodes central metabolism in order Methanomassiliicoccales, furthermore, conservation between the rumen members of Methanomassiliicoccales have revealed specialised DNA polymerase, cell envelope and cation transport genes that are important to the adaptation of environmental conditions, and genes predicted to be involved in exopolysaccharide production and adhesin-like proteins that are unique to ISO4-H5, these findings will require physiological and biochemical studies to verify.

Chapter 5

Phenotypic characterisation of ISO4-H5 and analysis of its growth under high and low H_2 levels

5.1. Introduction

As described in Chapter 3, methanogen cultivation studies have been undertaken in order to understand the types of Methanomassiliicoccales present in the ovine rumen (Jeyanathan 2010). The Methanomassiliicoccales enrichment cultures displayed a common coccoid morphology and lacked autofluorescence at 420 nm in these putative methanogen cells (Jeyanathan 2010). Previous attempts were made to isolate the Methanomassiliicoccales organisms from the ovine rumen enrichment cultures, including 10-fold serial dilutions, application of various antibiotics and heat treatments and although pure cultures of Methanomassiliicoccales were not obtained, the bacterial inhabitants in the enrichment cultures were reduced in both number and diversity (Jayanathan, 2010). Analysis of the bacterial partner organisms in the enrichments showed a single strain of *Succinivibrio dextrinosolvens* (NR026476, 99% 16*S* rRNA gene identity) was present (Jeyanathan 2010), suggesting there may be specific interactions between this organism and Methanomassiliicoccales (Jeyanathan 2010). Additionally, a recent global rumen census project indicates a correlation between rumen Succinivibrionaceae with some rumen Methanomassiliicoccales

Recent studies have provided some insights into Methanomassiliicoccales metabolism. Methanomassiliicoccales carries out H_2-dependent methylotrophic methanogenesis, utilizing mono-, di-, tri-methylamines, methanol and dimethyl-sulfide (Borrel *et al.* 2014). Despite the understanding gained from genomes of Methanomassiliicoccales of human and termite gut origin, it is important to study the Methanomassiliicoccales from the rumen in order to understand their function in the rumen and to assess their contribution to CH_4 production in this environment. Therefore, studies were carried out on the ISO4-H5 Methanomassiliicoccales enrichment culture with the aim of attempting to isolate the methanogenic archaeon from the enrichment, to enable its growth requirements to be determined and to explore its response to different levels of H_2 in co-culture experiments with the cellulose-degrading, H_2-producing bacterium, *Ruminococcus flavefaciens*.

5.2. Results

5.2.1. Isolation of ISO4-H5

A *S. dextrinosolvens* strain from the ISO4-H5 enrichment culture was isolated previously by colony picking from BY plates and designated as *S. dextrinosolvens* strain H5 (Cox *et al.*, pers. comm.) and was kindly supplied for this study. During growth studies of the *S. dextrinosolvens* H5 culture, pectin (1% w/v) was accidentally introduced to the ISO4-H5 enrichment culture. The culture was not discarded, but was monitored, and elevated CH_4 production relative to the control culture without added pectin, was observed. Further testing was conducted on supplementation of the ISO4-H5 enrichment cultures with glucose (10mM) and pectin (1% w/v). CH_4 production of ISO4-H5 enrichment cultures were stimulated by both glucose and pectin supplementation, however, pectin supplementation stimulated CH_4 production earlier and to a greater extent at day five (*P* =5.6E-05) than glucose (*P* =1.4E-02) (Figure 5.1.).

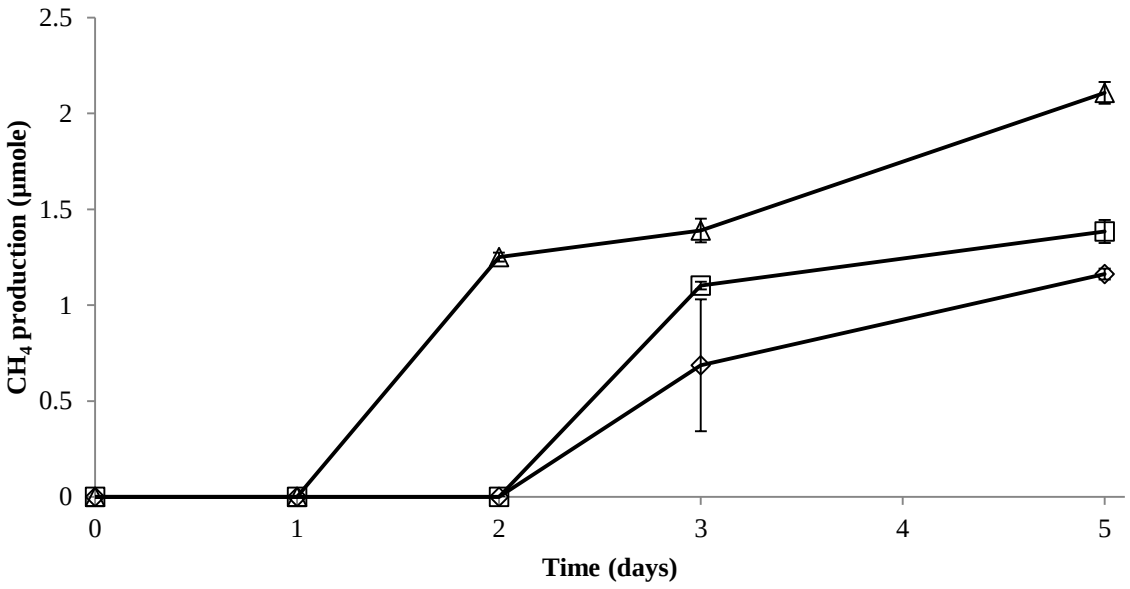


Figure 5.1. Stimulation of CH_4 production in ISO4-H5 enrichment cultures triplicates by glucose or pectin supplementation. All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. (–◊–) Control, (–□–) glucose (10 mM final concentration) or (–Δ–) apple pectin (1% w/v final concentration) was supplied to ISO4-H5 enrichment cultures, and CH_4 production per mL of headspace was measured daily. Error bars represent +/- standard errors of the means (SEMs). ISO4-H5 growth control is the standard cultivation conditions for the ISO4-H5 enrichment culture previously described (Chapter 2, section 2.2.1.)

From these observations, it was hypothesized that pectin addition indirectly stimulated CH_4 production by enhancing *S. dextrinosolvens* H5 growth, which in turn supplied a growth-limiting metabolite to the methanogenic archaeon in the ISO4-H5 enrichment culture. It was

reasoned that such a growth-limiting metabolite was likely to be soluble and present in the supernatant of *S. dextrinosolvens* cultures grown on pectin. If so, the supernatant itself would allow the growth of the methanogenic archaeon in the absence of *S. dextrinosolvens* H5 cells, therefore *S. dextrinosolvens* H5 cultures were grown for two days in BY medium containing pectin (1% w/v). The bulk of *S. dextrinosolvens* H5 cells were removed by filtering through a 0.22 µm filter to remove cells, thus filter sterilising the preparation. Although previous attempts to remove *S. dextrinosolvens* H5 with antibiotic treatments were unsuccessful, the sensitivity of *S. dextrinosolvens* and the methanogenic archaeon in the ISO4-H5 enrichment culture to antibiotics was determined. The antibiotics were added to the ISO4-H5 enrichment culture, and the culture was also supplemented with glucose (10 mM final concentration) and growth of *S. dextrinosolvens* H5 after overnight incubation was recorded. The incubation was continued for seven days after which CH_4 production from the antibiotic tested culture was measured (Table 5.1.).

Table 5.1. Antibiotic application to enrichment culture

Antibiotic	**Final concentration (µg/mL)**	***S. dextrinosolvens* H5 growth overnight**	**CH_4 production after 7 days**
Ampicillin	50	Yes	Yes
	100	Yes	Yes
	250	No	No
	500	No	No
Chloramphenicol*	50	No	No
Erythromycin*	20	Yes	Yes
Gentamycin	20	Yes	No
Kanamycin	50	Yes	Yes
Rifampicin#	1.2	Yes	Yes
Spectinomycin	100	Yes	Yes
Streptomycin	100	Yes	Yes
Tetracycline*	10	Yes	Yes
	20	Yes	Yes
	50	Yes	Yes
	100	Yes	Yes

* dissolved in ethanol. # dissolved in methanol at maximum antibiotic solubility.

Only chloramphenicol and ampicillin inhibited *S. dextrinosolvens* H5, while only gentamycin inhibited CH_4 production in the ISO4-H5 enrichment cultures. In all cases of *S. dextrinosolvens* H5 inhibition, there was no CH_4 detected after seven days of incubation, indicating that the methanogenic archaeon did not grow, possibly due to direct inhibition by the antibiotics, or possibly because ISO4-H5 did not survive in the absence of *S. dextrinosolvens* H5 growth.

The ampicillin results were particularly encouraging as archaea are insensitive to ampicillin, and suggested a concentration between 100 and 250 µg/mL might be useful for inhibiting *S. dextrinosolvens* growth. The isolation procedure used is outlined in Chapter 2 (Section 2.2.3, Figure 5.2.), but briefly it involved supplementing ISO4-H5 enrichments with cell-free, spent growth medium filtered from pectin-grown *S. dextrinosolvens* H5 cultures which is termed

Succinivibrio Spent Pectin Growth Medium Supernatant (SSPGMS). In addition to supplementing filter-sterilised SSPGMS and with ampicillin and incubating the cultures until CH_4 was detected in the headspace. The culture was then sub-cultured onto fresh medium containing the same supplements, while a parallel inoculum was introduced to medium containing glucose and incubated overnight to check for the presence of remaining *S. dextrinosolvens*. Initial attempts using 100 µg/mL ampicillin were unsuccessful and *S. dextrinosolvens* persisted in the enrichment culture. The ampicillin concentration was increased to 200 µg/mL, and the isolation procedure was repeated through three cycles of SSPGMS supplementation and ampicillin treatment before the glucose supplemented test cultures remained clear and CH_4 was detected in the culture headspace. The putative pure methanogen culture consisted of coccoid cells that usually do not fluoresce at 420 nm. The putatively pure cultures were screened with PCR amplification using primers specific for both bacterial and archaeal 16*S* rRNA genes. The bacterial 16*S* rRNA PCR failed to produce an amplification product, while the archaeal 16*S* rRNA gene PCR produced a product of the expected size. The PCR product was sequenced and identified as a methanogenic archaeon, affiliated with the order Methanomassiliicoccales, and the culture from which this amplified DNA originated was designated as strain ISO4-H5.

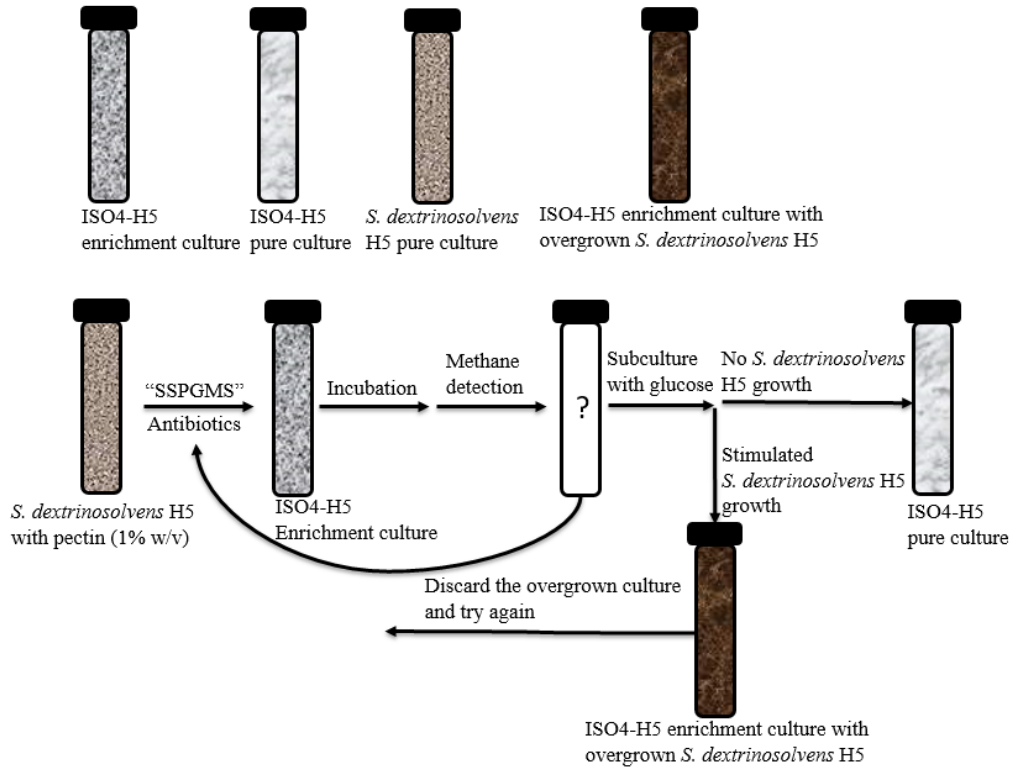


Figure 5.2. Isolation procedure for ISO4-H5. SSPGMS: *Succinivibrio* Spent Pectin Growth Medium Supernatant. Antibiotics: 200 µg/mL ampicillin. Glucose: 10 mM. Length of incubation: one week.

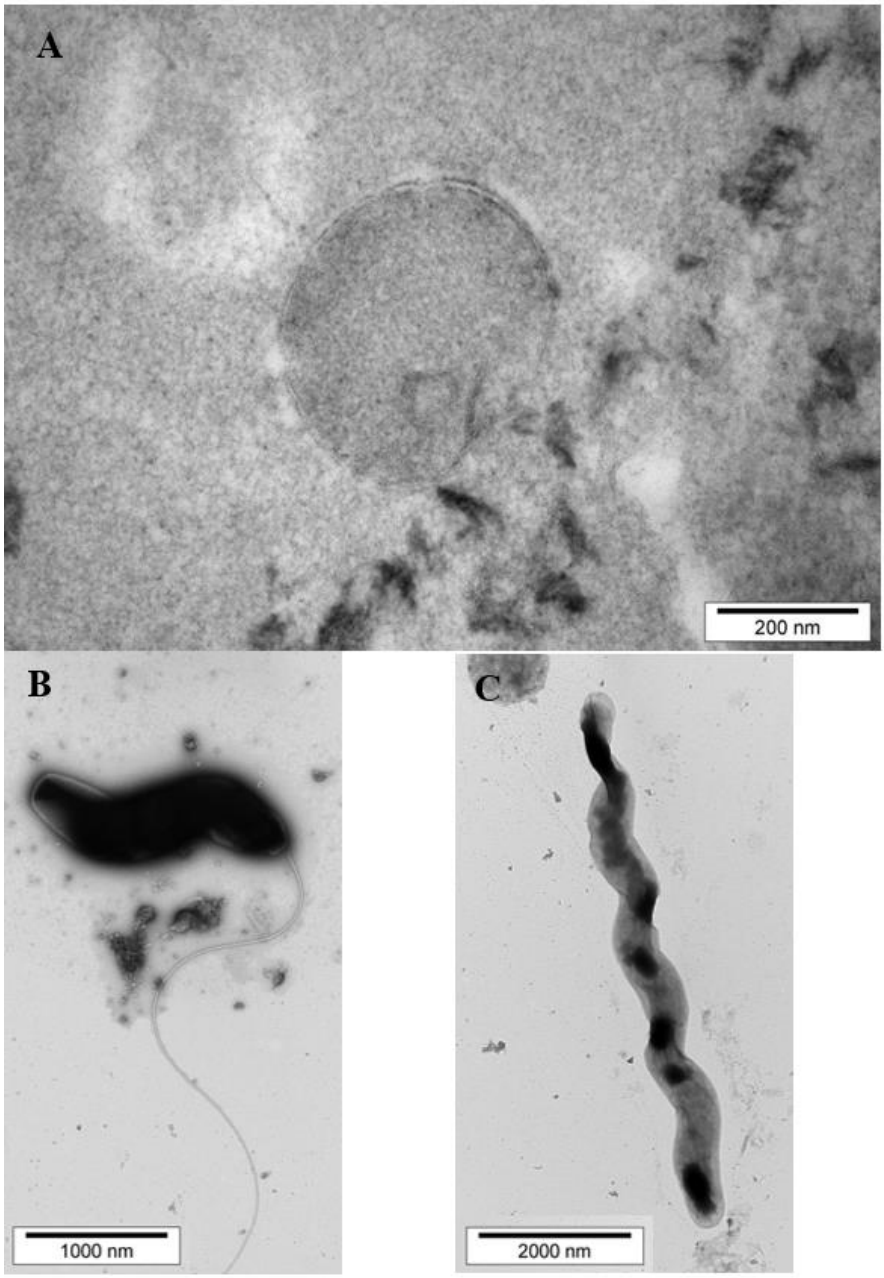


Figure 5.3. Transmission electron micrograph (TEM) of methanogenic archaeon ISO4-H5 and *S. dextrinosolvens* H5. **A.** TEM of negatively stained thin section of the methanogenic archaeon ISO4-H5. **B** and **C.** TEM of negatively stained *S. dextrinosolvens.* The samples were prepared as previously described in Chapter 2. Images were captured using a Philips CM10 Transmission Electron Microscope, using an Olympus SIS Morada camera and SIS iTEM software (Germany).

The morphologies of *S. dextrinosolvens* H5 and methanogenic archaeon ISO4-H5 were examined under transmission electron microscopy using negative staining of cultures, and ultrathin sectioning of ISO4-H5 was also carried out. The methanogenic archaeon ISO4-H5 has a thin, bi-layer cell membrane with no visible external features (Figure 5.3A.). An unidentified structure can be seen within the cell. The negatively stained *S. dextrinosolvens* H5 preparations show it possesses a single terminal flagellum (Figure 5.3B.), and some cells displayed an elongated spiral morphology (Figure 5.3C.). The size of ISO4-H5 was

approximately 0.4 µm in diameter, whereas *S. dextrinosolvens* H5 was approximately 0.5 µm in diameter and varied in length from 1.5 µm to over 7.0 µm. Observation of the ISO4-H5 enrichment culture under phase contrast microscopy did not see any evidence of specific physical interaction between the methanogenic archaeon ISO4-H5 and *S. dextrinosolvens* H5.

The growth of the ISO4-H5 pure culture was compared with that of the enrichment culture (Figure 5.4.). The enrichment culture produced significantly more CH_4 than the pure culture, and the SSPGMS stimulated CH_4 production in ISO4-H5 pure culture, although it is not necessary for the pure culture to thrive (Figure 5.4.).

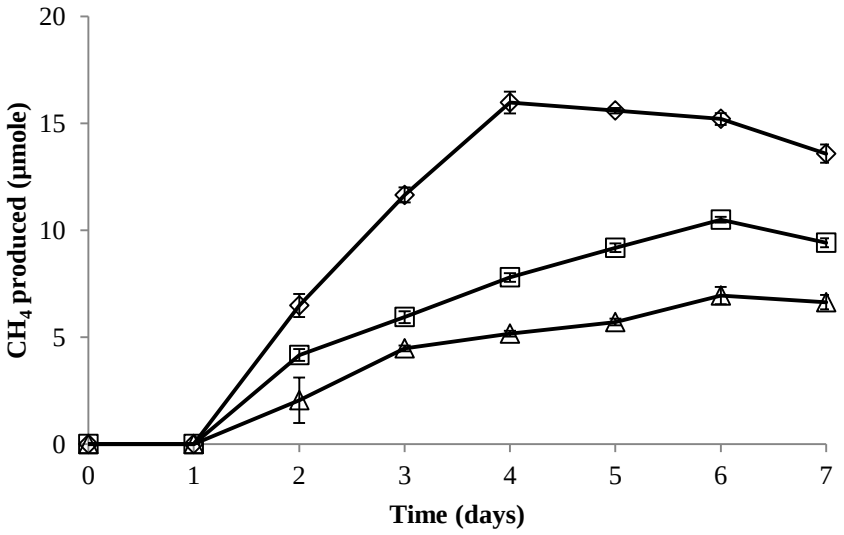


Figure 5.4. Comparison of growth between pure and enrichment cultures of ISO4-H5. All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix, (–◊–) ISO4-H5 enrichment culture, (–□–) ISO4-H5 pure culture with 10% (v/v) SSPGMS, (–Δ–) ISO4-H5 pure culture without SSPGMS. CH_4 per mL of headspace was measured and expressed as µmole per mL. Error bars represent ± SEMs.

5.2.2. Growth of *S. dextrinosolvens* H5

The growth of *S. dextrinisolvens* H5 in pectin (1%), rhamnose (10 mM), galactose (10 mM) and arabinose (10 mM) was assessed by both optical density (OD) and a decline in pH (Figure 5.5A., B.), no significant growth was observed from pectin and rhamnose.

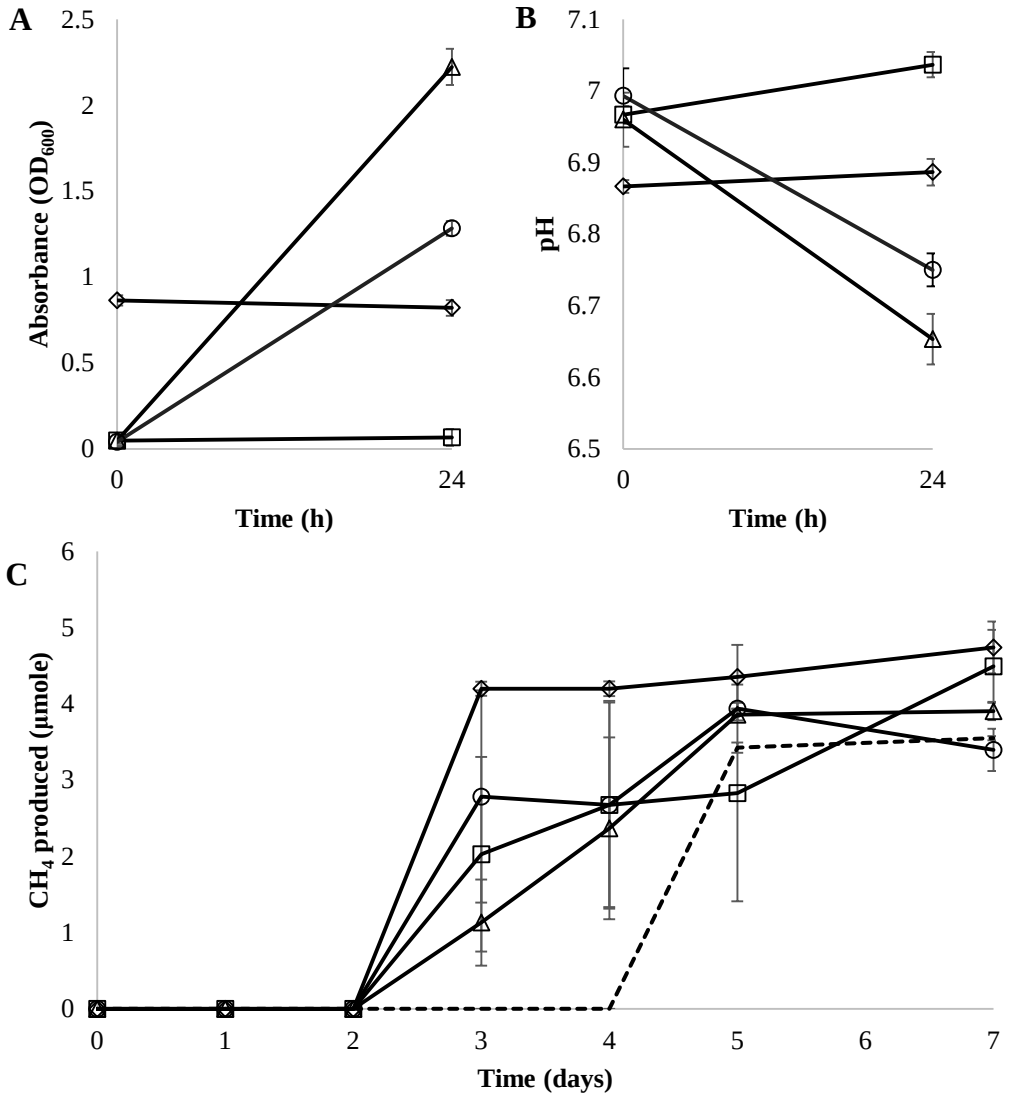
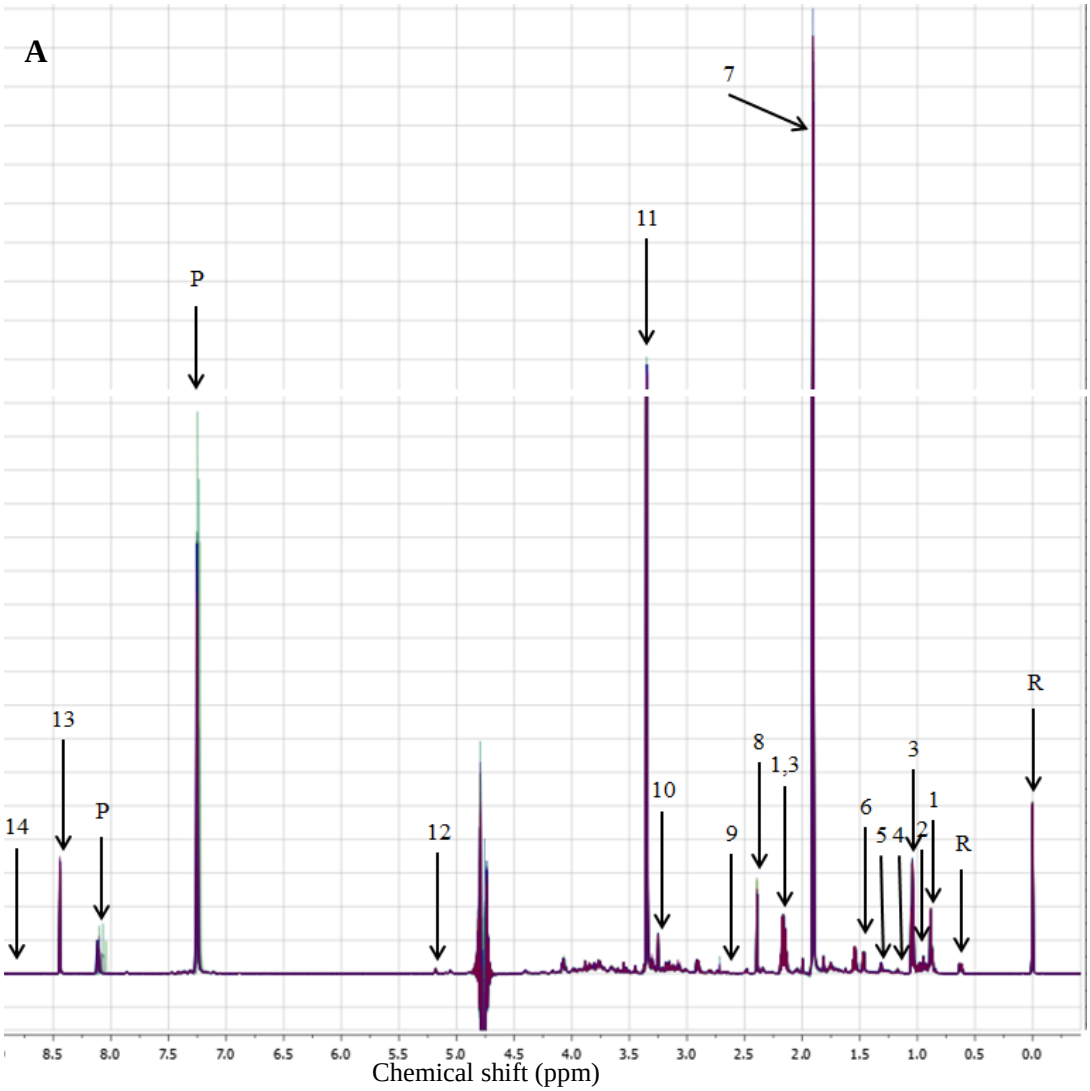


Figure 5.5. Growth of *S. dextrinosolvens* H5 on varying sources of carbohydrate and its growth medium supplementation to ISO4-H5 culture. **A.** Growth of *S. dextrinosolvens* H5 monitored by absorbance. All replicates contain 0.1× vitamin mix. Culture contain (–◊–) 1% pectin (w/v), (–□–) rhamnose (10 mM), (–Δ–) galactose (10 mM), (–○–) arabinose (10 mM). **B.** pH of *S. dextrinosolvens* H5 culture. All replicates contain 0.1× vitamin mix. Culture contain (–◊–) 1% pectin (w/v), (–□–) rhamnose (10 mM), (–Δ–) galactose (10 mM), (–○–) arabinose (10 mM). **C**. All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. *Succinivibrio* spent pectin/rhamnose/galactose/arabinose growth medium supernatant (SS(P/R/G/A)GMS). (---) ISO4-H5 pure culture with no supplementation, (–◊–) ISO4-H5 pure culture with 10% (v/v) SSPGMS, (–□–) ISO4-H5 pure culture with 10% (v/v) SSRGMS, (–Δ–) ISO4-H5 pure culture with 10% (v/v) SSGGMS, (–○–) ISO4-H5 pure culture with 10% (v/v) SSAGMS. CH_4 per mL of headspace was measured and expressed as µmole per mL. Error bars represent ± SEMs.

The growth medium supernatant of *S. dextrinisolvens* cultivated on rhamnose, galactose and arabinose was turned SSRGMS, SSGGMS and SSAGMS respectively. The stimulation of SSPGMS, SSRGMS, SSGGMS and SSAGMS on the growth of ISO4-H5 was assessed (Figure 5.5B.), all of SSPGMS, SSRGMS, SSGGMS and SSAGMS stimulated the ISO4-H5 to produce CH_4 at day three, SSPGMS and SSRGMS supplemented ISO4-H5 culture produced the highest amount of CH_4 at day seven.

5.2.3. NMR analysis of SSPGMS-supplemented ISO4-H5 culture

To identify the components of SSPGMS that stimulated the growth of ISO4-H5 and allowed its isolation, proton nuclear magnetic resonance (^{1}H NMR) was used. The purified ISO4-H5 was cultured in medium with SSPGMS supplementation, and the difference in the ^{1}H NMR spectra of the medium before and after cultivation was compared (Figure 5.6.). Uninoculated controls (medium with SSPGMS added, and incubated for the same period of time) were also analysed to rule out changes in medium composition over the duration of the incubation. The concentrations of detected metabolites estimated from the spectra are listed in Table 5.2.



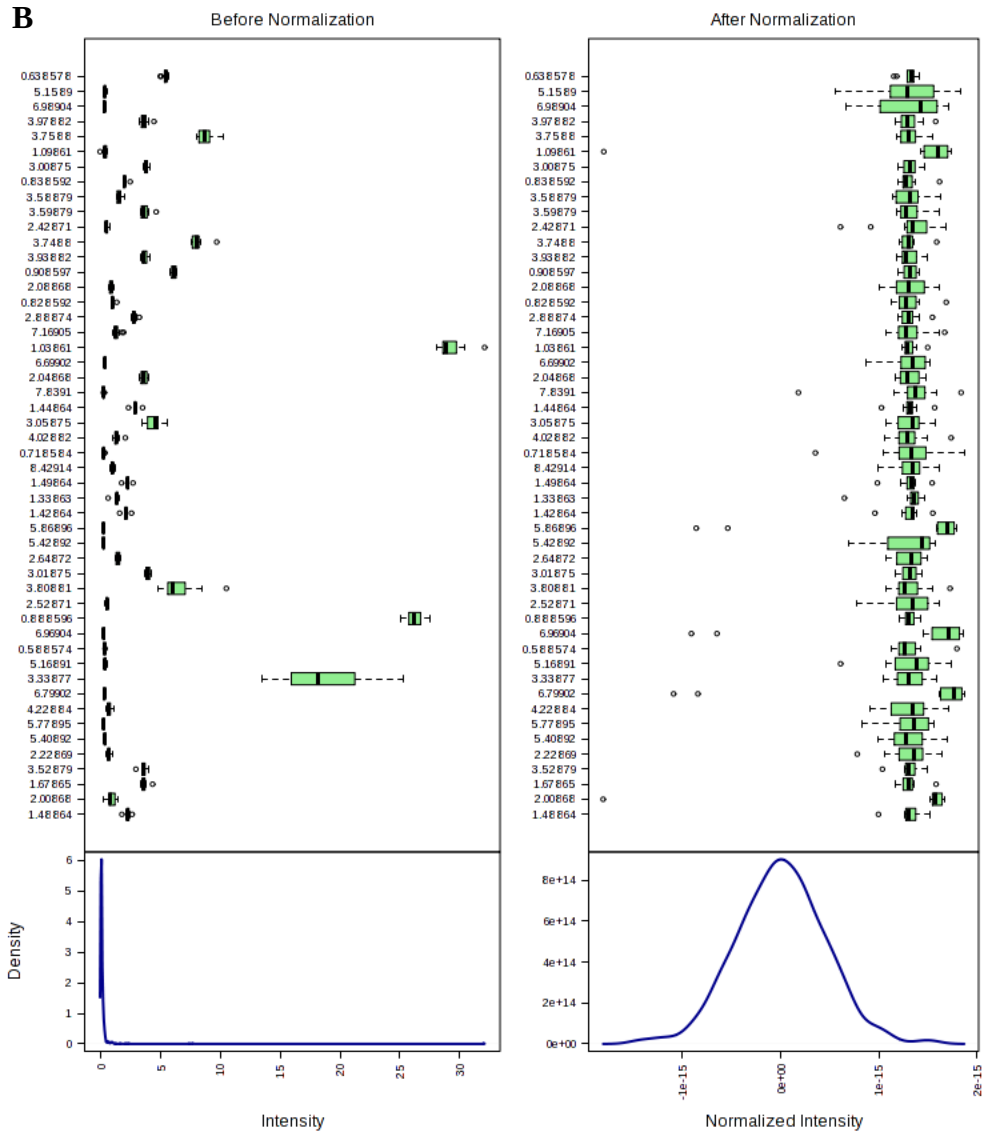


Figure 5.6. NMR analysis of medium composition before and after growth of ISO4-H5. **A.** Overlay of all spectra collected from 24 samples (6 inoculated replicates pre and post incubation, 6 uninoculated replicates pre and post incubation). R, internal standard 4,4-dimethyl-4-silapentaine-1-sulfonic acid (DSS); 1, butyrate; 2, branched-chain amino acids; 3, propionate; 4, ethanol; 5, lactate; 6, alanine; 7, acetate; 8, succinate; 9, dimethylamine; 10, betaine; 11, methanol; 12, trehalose; 13, formate; 14, nicotinamide; P, pH indicator – imidazole. **B.** Box plots and kernel density plots of binned spectra of sample replicates before and after normalisation by generalized log transformation (glog 2) and Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable). The chemical shift of spectra bins does not correlate with spectra peaks displayed in A. Intensity of resonance signals and density of spectra bins are displayed. The density plot shown in the bottom panel summarizes the distribution data from all samples. The boxplots show at most 50 features due to space limit.

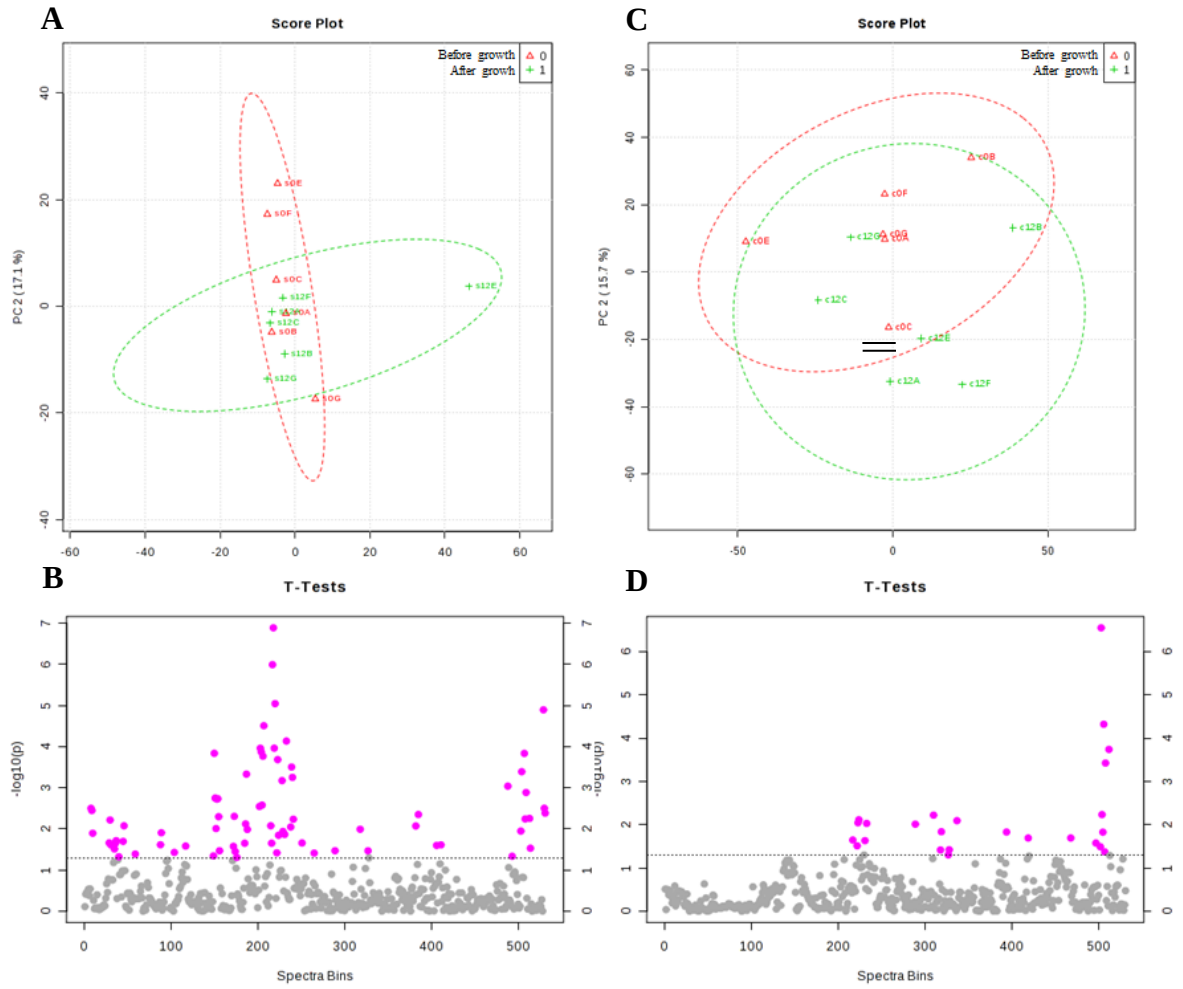


Figure 5.7. Statistical analysis of paired control and sample replicates. **A.** Score plot for principal component analysis (PC1 versus PC2) of paired sample replicates. **B.** T-test of paired sample spectra bins with *P*-value threshold of 0.05 plotted on a -$\log_{10}$ scale. The pink circles represent features above the threshold. **C.** Score plot for principal component analysis (PC1 versus PC2) of paired control replicates. **D.** T-test of paired control spectra bins with *P*-value threshold of 0.05 plotted on a -$\log_{10}$ scale. The pink circles represent features above the threshold.

Table 5.2. Metabolite concentrations estimated by ^{1}H NMR spectroscopy

Compounds	Before growth		After growth	
	mM*	SD	mM*	SD
Butyrate	1.190	0.035	1.123	0.116
2-hydroxyvalerate	0.134	0.009	0.139	0.007
Leucine	0.245	0.014	0.239	0.020
Isoleucine	0.170	0.019	0.172	0.011
Valine	0.192	0.007	0.193	0.022
Propionate	2.440	0.048	2.396	0.125
Ethanol	0.131	0.023	0.100	0.010
Lactate	0.130	0.010	0.135	0.017
Alanine	0.300	0.013	0.300	0.010
Acetate	25.565	1.503	24.677	1.422
Acetamide	0.187	0.015	0.182	0.051
Succinate	0.458	0.059	0.452	0.053
Methionine	0.038	0.008	0.034	0.008
Aspartate	0.164	0.016	0.216	0.022
Dimethylamine	0.023	0.017	0.020	0.012
Phenylalanine	0.102	0.006	0.102	0.025
Betaine	0.116	0.006	0.116	0.003
Methanol	17.886	0.899	15.282	0.605
Trehalose	0.126	0.006	0.127	0.004
Tyrosine	0.043	0.005	0.049	0.003
Formate	4.432	0.413	4.264	0.473

*: numbers indicate mean concentrations from six replicates estimated from 1D ^{1}H NMR using DSS as reference

The spectra were binned and normalized as described previously (Chapter 2, Section 2.2.27) and analysed statistically by Principal Component analysis and T-test (Figure 5.7.). The paired binned spectra that displayed statistical significance of $P \leq 0.05$ are listed in Appendix Tables A.5.1, and A.5.2. The spectra bins that displayed significant difference before and after growth in both control and sample group were disregarded. The spectra bins 1.18 and 8.71 were significantly reduced post-growth in paired comparisons where each replicate was compared before and post-incubation individually, while spectrum bin 3.35 was significantly reduced post-growth in an unpaired comparison where the before and post-incubation spectra were compared between all replicates (Figure 5.8.).

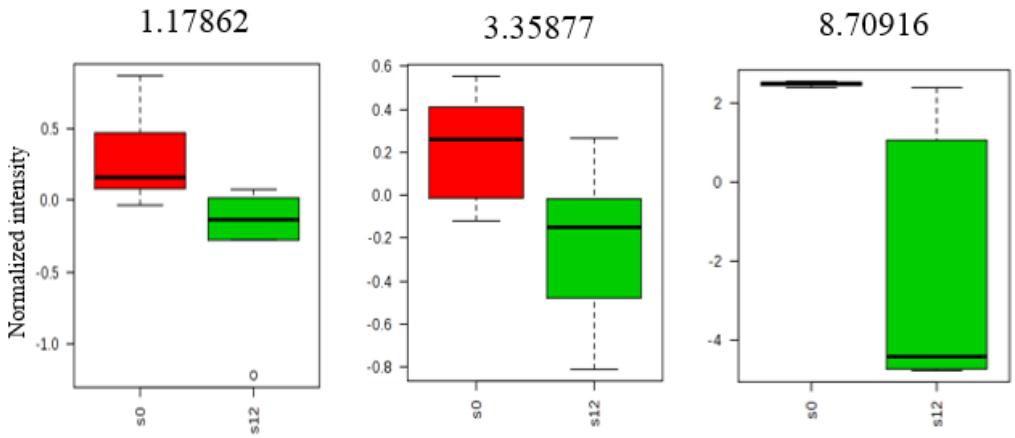


Figure 5.8. Spectra bins that were reduced significantly post-growth. s0: samples before incubation, s12: samples post-incubation. Spectra resonances were normalized by log transformation and Pareto scaling.

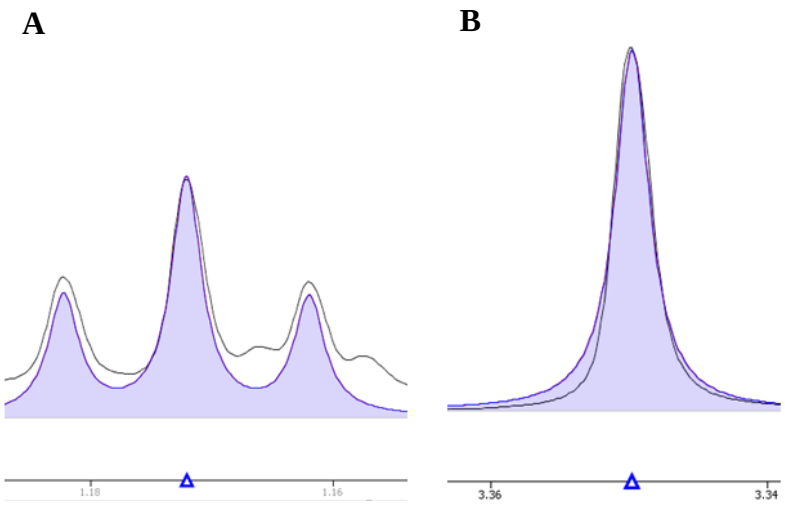


Figure 5.9. One-dimensional peak alignments. **A.** Chemical shift 1.18 aligned with standard peak of ethanol. **B.** Chemical shift 3.35 aligned with standard peak of methanol.

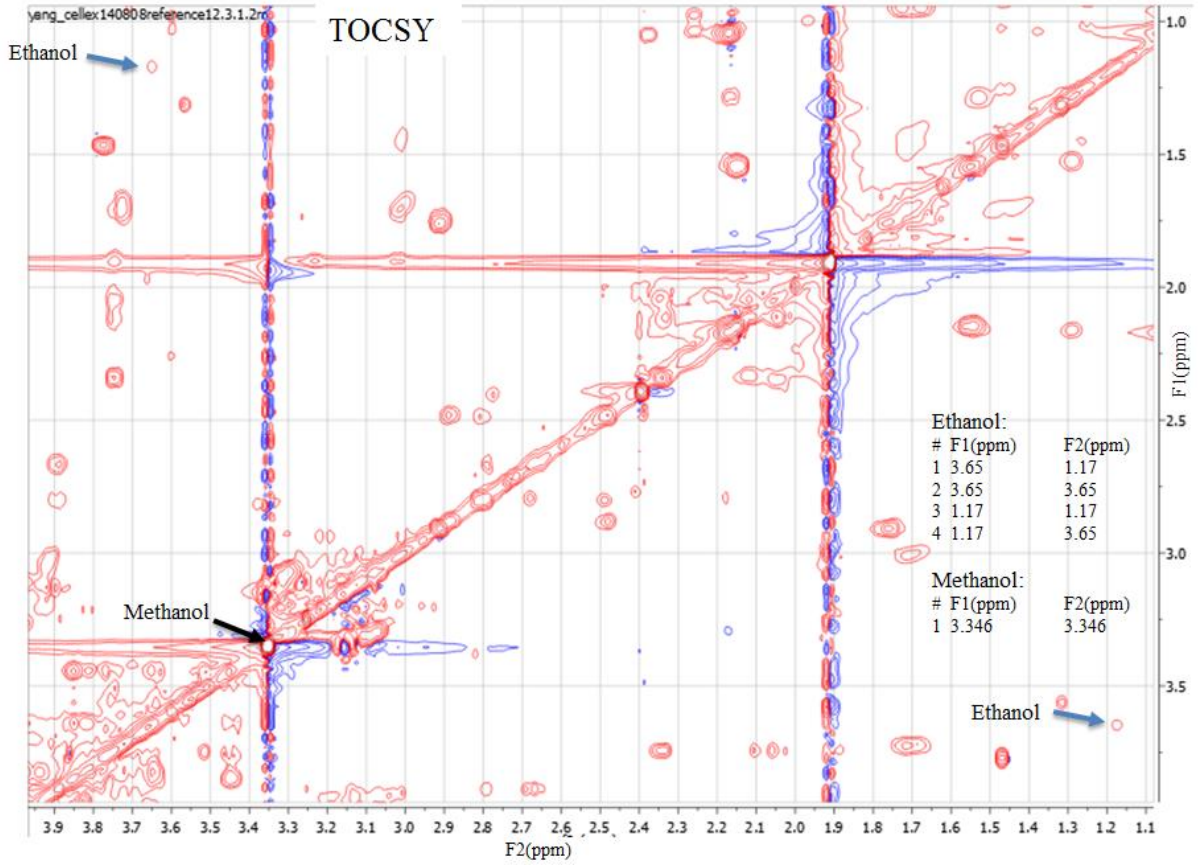


Figure 5.10. TOCSY of pooled samples. The correlation between protons (F1) and protons (F2) from samples with ethanol and methanol are displayed as positive (red) and negative (blue) peaks. The chemical shift values for two dimensional spectrum (F1 and F2) corresponding to ethanol and methanol are displayed. The sample spectrum associated with ethanol is labelled with blue arrows, the sample spectrum associated with methanol is labelled with black arrows.

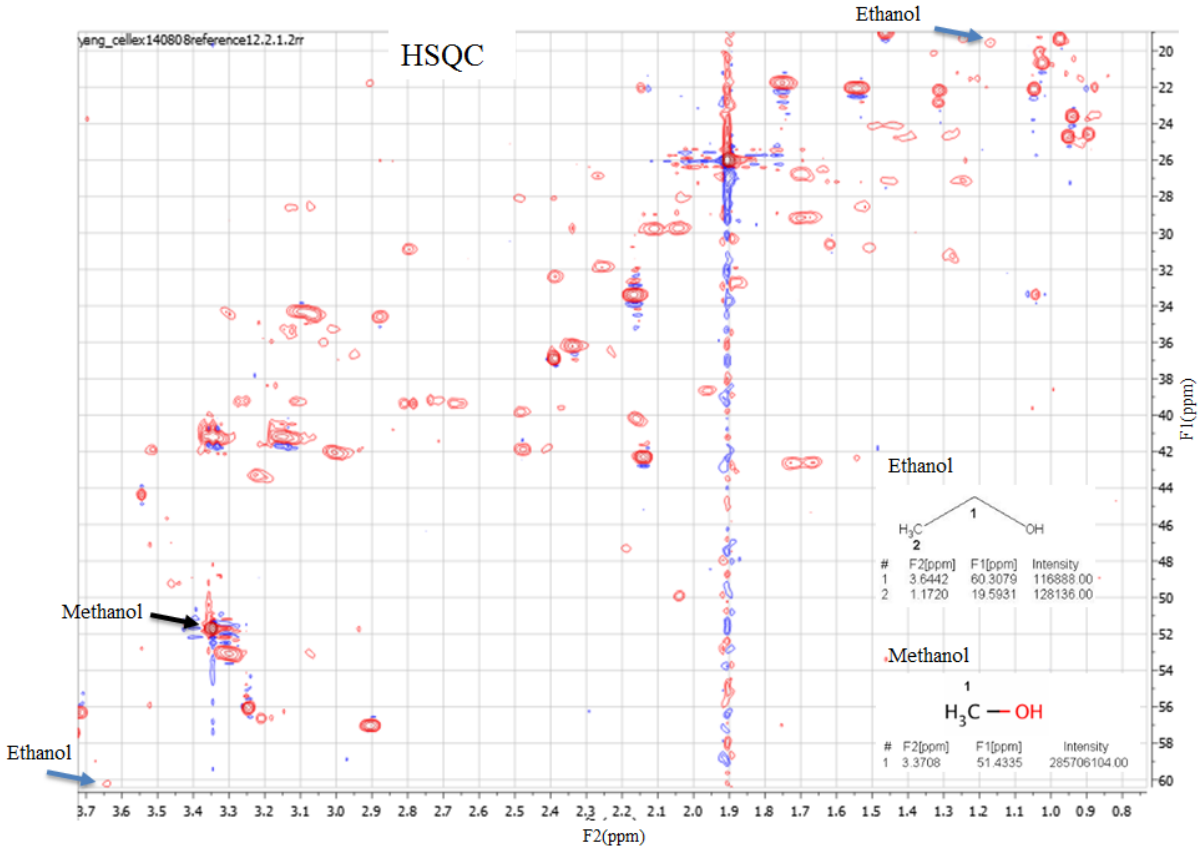


Figure 5.11. HSQC of pooled samples. The correlation between protons (F2) and carbons (F1) for ethanol and methanol are displayed as positive (red) and negative (blue) peaks. The chemical shift values for two dimensional spectrum (F1 and F2) corresponding to ethanol and methanol are displayed. The spectrum associated with ethanol is labelled with blue arrows, the spectrum associated with methanol is labelled with a black arrow.

Identification of the compounds associated with each spectrum bin was carried out by one-dimensional peak alignment, two dimensional Heteronuclear Single Quantum Coherence Spectroscopy (HSQC), and Total Correlated Spectroscopy (TOCSY) mapping with reference compounds. A compound with a chemical shift of 1.18 was identified as ethanol by one-dimensional peak alignment, and a compound with chemical shift of 3.35 was identified as methanol by one-dimensional peak alignment (Figure 5.9.). Both the ethanol and methanol were also verified in pooled sample TOCSY and HSQC (Figures 5.10., 5.11.).

The compound associated with spectrum bin 8.71 was at low concentration, and it could not be identified from one dimensional peak alignment, TOCSY or HSQC because it was too similar to background noise. Compounds that correspond to a chemical shift of 8.71±0.01 include nicotinamide, nicotinurate, pyruvic acid, isoniazid, biopterin, phenanthrene, chrysene and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. Nicotinamide, a B group vitamin, was considered the most relevant to ISO4-H5 growth. The SSPGMS preparation, which contains higher concentrations of metabolites, was analysed in a prolonged ^{1}H NMR for 72 h to amplify

the signal strength, and compared to SSPGMS spiked with nicotinamide (300 µM) (Figure 5.12.). After comparison, the compound at chemical shift 8.71 was verified as nicotinamide.

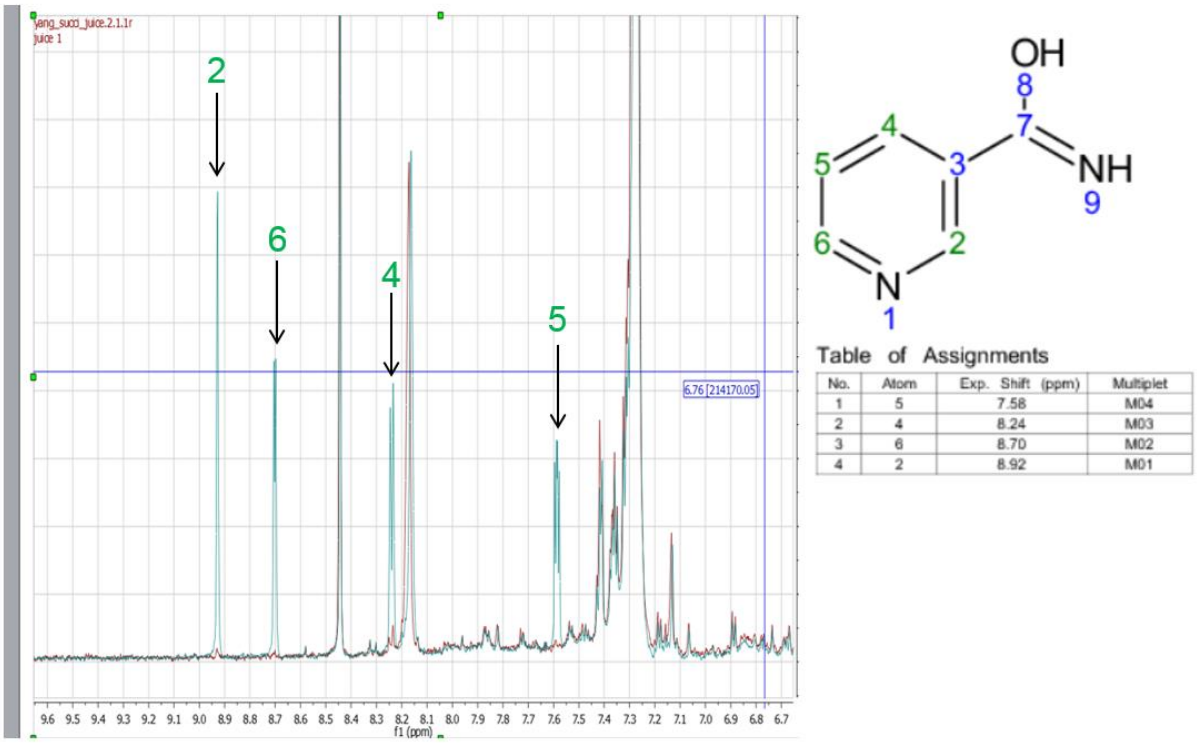


Figure 5.12. Identification of spectrum bin 8.71 as nicotinamide. The spectrum of nicotinamide-spiked SSPGMS (green) was aligned with SSPGMS alone (red). Molecular structure of nicotiniamide is shown at the right and the chemical shift of protons (F1) are listed in the table of assignments.

The nicotinamide concentration detected was close to the minimum detection limit of NMR, and it could not be quantified in the post-growth samples, therefore it is possibly limiting the growth of ISO4-H5. Nicotinic acid is already present in the Vitamin 10 mix added to the growth medium, and its final concentration in the medium is 0.1 mg/L. Additional nicotinamide and nicotinic acid were supplemented to ISO4-H5 culture at 100 fold more than the standard cultivation condition, nicotinic acid was removed from the Vitamin 10 mix in the 0 mg/L nicotinic acid treatment, together with SSPGMS supplementation, and CH_4 production was monitored.

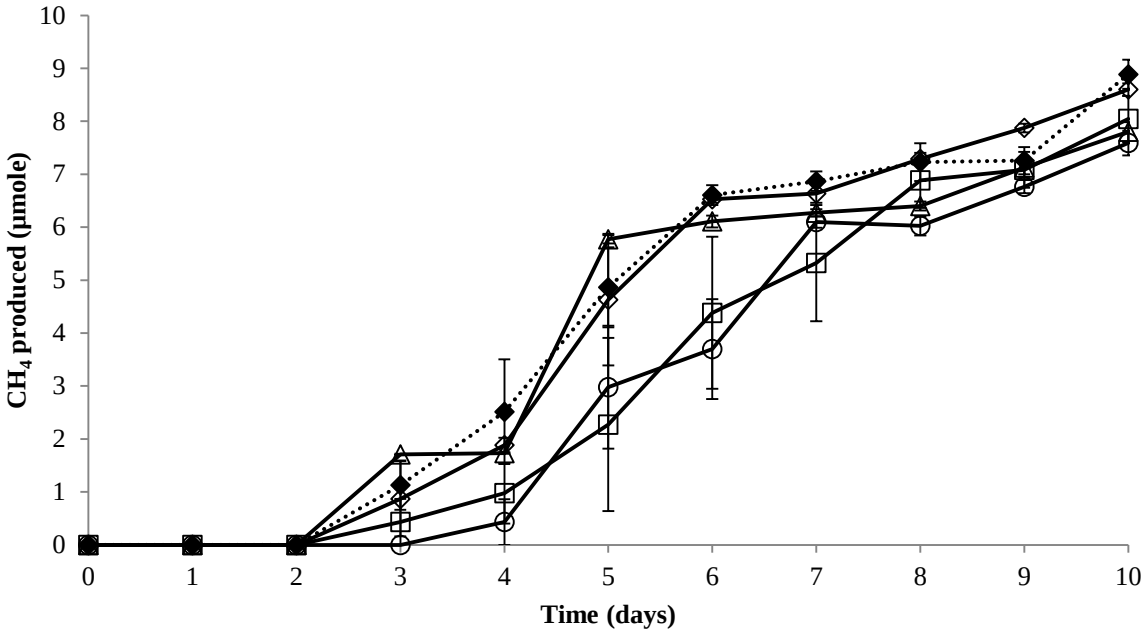


Figure 5.13. Growth of ISO4-H5 pure cultures with nicotinamide and nicotinic acid supplementation. CH_4 per mL of headspace produced was recorded. All replicates contain formate (60 mM), methanol (20 mM), acetate (20 mM), 10% (v/v) SSPGMS and CoM (10 µM). (–◊–) control with 0.1× vitamin mix, (–Δ–) nicotinamide (0.1 mg/L) with 0.1× vitamin mix, (–□–) nicotinamide (10 mg/L) with 0.1× vitamin mix, (–○–) nicotinic acid (0 mg/L) with modified 0.1× vitamin mix omitting nicotinic acid, (···♦···) nicotinic acid (10 mg/L) with modified 0.1× vitamin mix omitting nicotinic acid. Error bars represent ± SEMs.

The supplementation of nicotinamide or nicotinic acid in excess or removal of nicotinic acid resulted in CH_4 production equivalent to the control, which contains SSPGMS and vitamin mix with 0.1 mg/L nicotinic acid (Figure 5.13.).

5.2.4. Substrate utilisation

The ISO4-H5 enrichment culture was originally grown with 20 mM methanol, and the genome analysis (Chapter 3) predicted utilization of mono-, di-, tri-methylamine and methylthiol as substrates for methanogenesis. In order to assess the utilization of alternative methyl substrates, it was first necessary to establish the minimal amount of methanol required for ISO4-H5 growth, to prevent excess methanol being carried over in the inoculum during testing of alternative methyl substrates. The reduction in methanol concentration post-growth from six replicates analysed by NMR above, ranged from 4 mM to 1.8 mM, with an average reduction of 2.6 mM. Therefore the medium used for the methanol utilisation experiment was modified to contain 4 mM methanol. CH_4 production from ISO4-H5 was monitored on varying concentrations of methanol, with no added methanol as negative control and 10 mM methanol as positive control (Figure 5.14.). All other constituents of the medium were the same as

described in Chapter 2, except with the addition of 10 mM ethanol and 0.1 mg/L (final concentration) of nicotinamide.

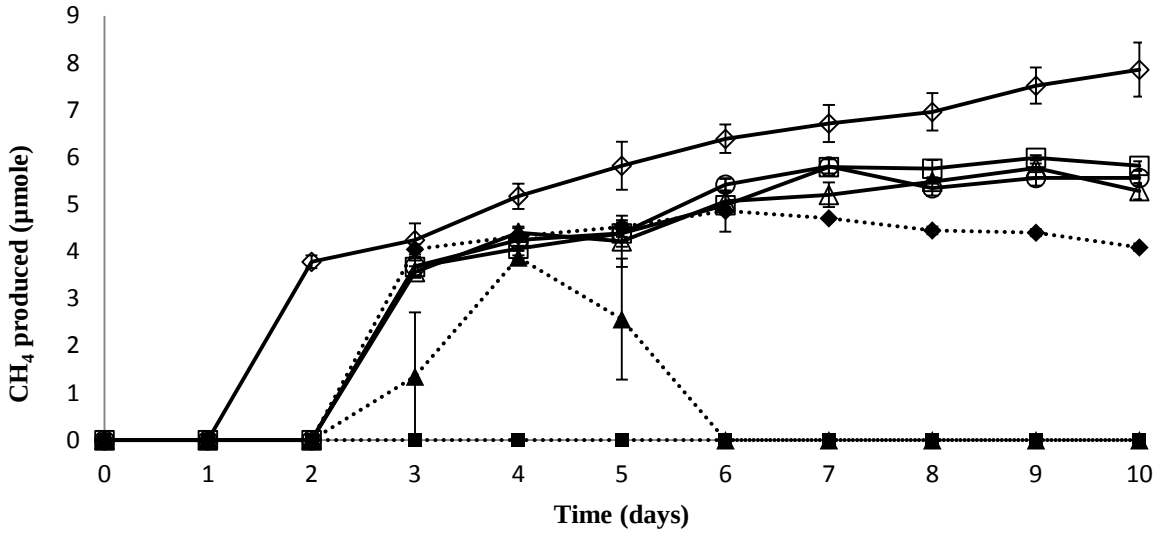


Figure 5.14. Growth of ISO4-H5 pure cultures with varying methanol concentrations. CH_4 per mL of headspace produced by ISO4-H5 over 10 days. All replicates contain formate (60 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. The starter culture was grown in medium containing 4mM methanol and 10% (v/v) SSPGMS. Culture was re-pumped with H_2/CO_2 (4:1; 200 kPa) after day five. (–◊–) methanol (10 mM) with SSPGMS, (–Δ–) methanol (10 mM) without SSPGMS, (–□–) methanol (3 mM) without SSPGMS, (–○–) methanol (2 mM) without SSPGMS, (···♦···) methanol (1 mM) without SSPGMS, (···▲···) methanol (0.5 mM) without SSPGMS, (···■···) methanol (0 mM) without SSPGMS. Error bars represent ±SEMs.

The ISO4-H5 pure culture was unable to grow without methanol (Figure 5.14.); however, it was able to grow on 0.5 mM methanol briefly before CH_4 production ceased at day five, the reduction of CH_4 on day five and six is likely due to re-pumping of the culture with H_2/CO_2 on day five. ISO4-H5 produced less CH_4 than the control when grown with 1 mM methanol, with declining CH_4 detected after day six. There was no significant difference in CH_4 production from ISO4-H5 grown with 10 mM, 3 mM or 2 mM methanol. Therefore 2 mM methanol was defined as the non-limiting concentration required for ISO4-H5 growth. The ISO4-H5 culture with SSPGMS supplement produced more CH_4 than the cultures with 10 mM methanol (Figure 5.14.) suggesting it contains growth stimulating components in addition to the methanol, ethanol and nicotinamide previously identified by NMR analysis.

Mono-, di-, trimethylamine (10 mM), and dimethylsulfide (10 mM) were tested as possible substrates for methanogenesis, and compared to growth on methanol (10 mM) in the absence of SSPGMS (Figure 5.15.). Ethanol (10 mM) and butanol (10 mM) were also tested due to ISO4-H5 encoding a homologue of a NADP-dependent alcohol dehydrogenase (AR505_0483). The CH_4 production and H_2 consumption of ISO4-H5 grown on mono-, di-,

tri-methylamine were equivalent to growth on methanol (Figure 5.15A., B.). ISO4-H5 was unable to produce CH_4 using dimethylsulfide, ethanol or butanol as substrate.

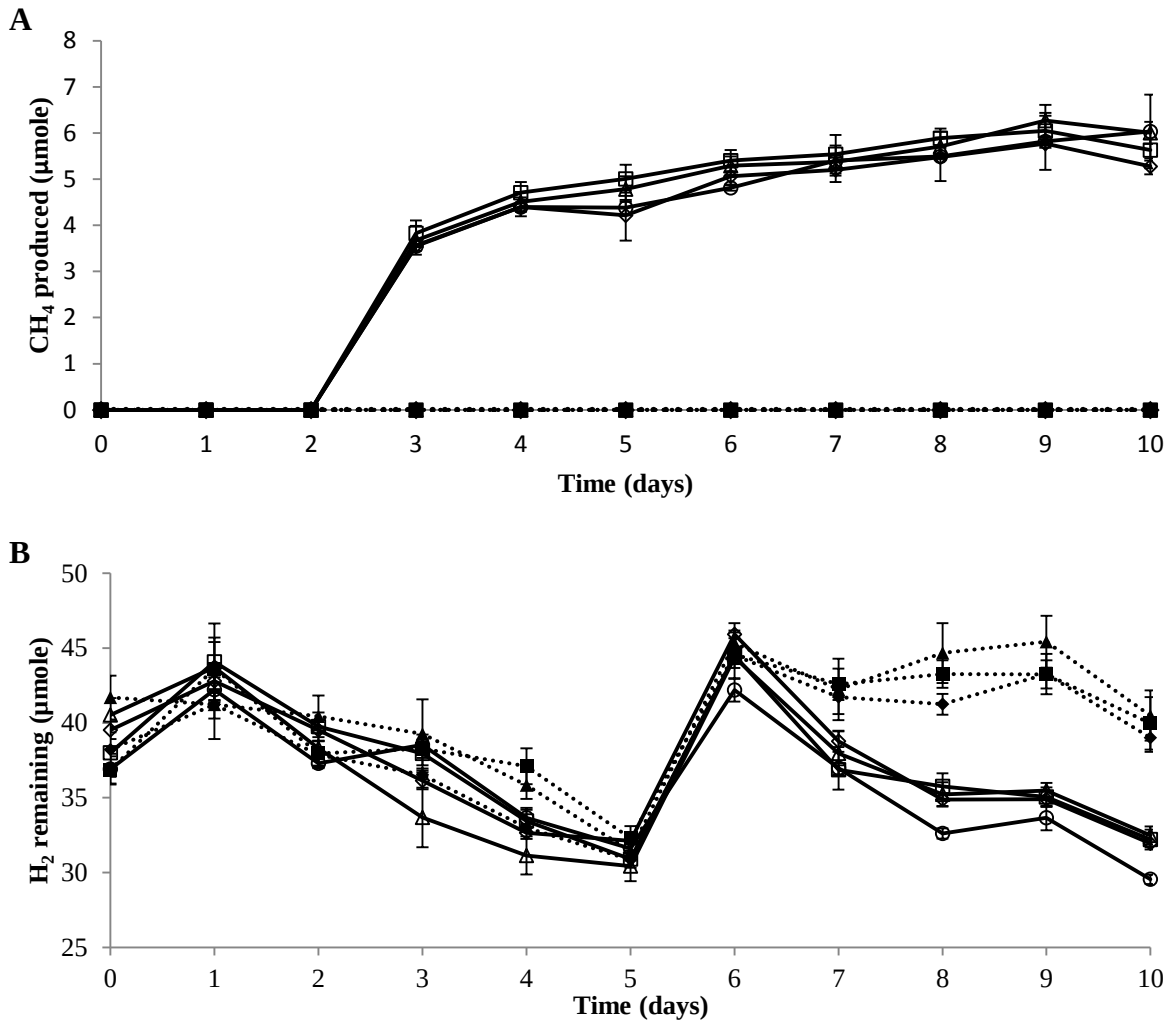


Figure 5.15. Growth of ISO4-H5 pure cultures with methyl substrates and H_2+CO_2. All replicates contain formate (60 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. (–◊–) methanol (10 mM), (–Δ–) trimethylamine (10 mM), (–□–) dimethylamine (10 mM), (–○–) methylamine (10 mM), (···♦···) dimethylsulfide (10 mM), (···▲···) ethanol (10 mM), (···■···) butanol (10 mM). **A.** CH_4 produced per mL of headspace was measured over 10 days and expressed as µmoles of CH_4 per mL of headspace. **B.** H_2 per mL of headspace used by ISO4-H5 during growth on methyl-substrates was measured and expressed as µmoles of H_2 remaining perl mL of headspace. Cultures were re-pumped after day 5 with H_2/CO_2 (4:1; 200 kPa). Error bars represent ±SEMs.

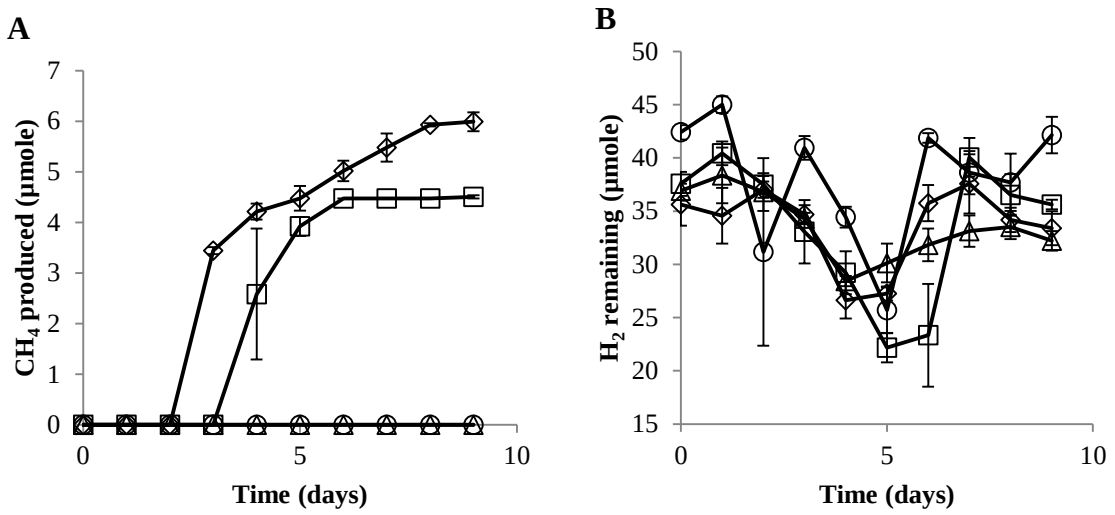


Figure 5.16. Growth of ISO4-H5 pure cultures with methyl-sources and H_2+CO_2. All replicates contain formate (10 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. (–◊–) methanol (10 mM), (–Δ–) M3SP (10 mM), (–□–) M3MSP (10 mM), (–○–) methanol (0 mM). M3SP: methylmercaptopropionate. M3MSP: methyl-3-methylthiopropionate. **A.** Amount of CH_4 produced per mL of headspace by ISO4-H5 over nine days. **B.** H_2 per mL of headspace remaining from ISO4-H5 growth on methanol. Culture was re-pumped after day five with H_2/CO_2 (4:1; 200 kPa). Error bars represents ±SEMs.

As a homologue of methylthiol: CoM methyltransferase MtsA was found in the ISO4-H5 genome, methylmercaptopropionate (M3SP) and methyl-3-methylthiopropionate (M3MSP) were also explored as possible substrates for methanogenesis. Results from these growth tests show that ISO4-H5 can use M3MSP but not M3SP. Total CH_4 production from M3MSP was less than from methanol and CH_4 formation from M3MSP was observed a day later (Figure 5.16A.). In the previous growth test it was shown that ISO4-H5 does not produce CH_4 from ethanol. As ethanol has been reported to be able to supply reducing potential to methanogenesis in *Mbb. ruminantium* $M1^T$, ethanol was tested in ISO4-H5 as a source of reducing potential for methanogenesis. Results from the growth test showed that ISO4-H5 cannot use ethanol to replace H_2 as a source of reducing potential to produce CH_4 (Figure 5.17.).

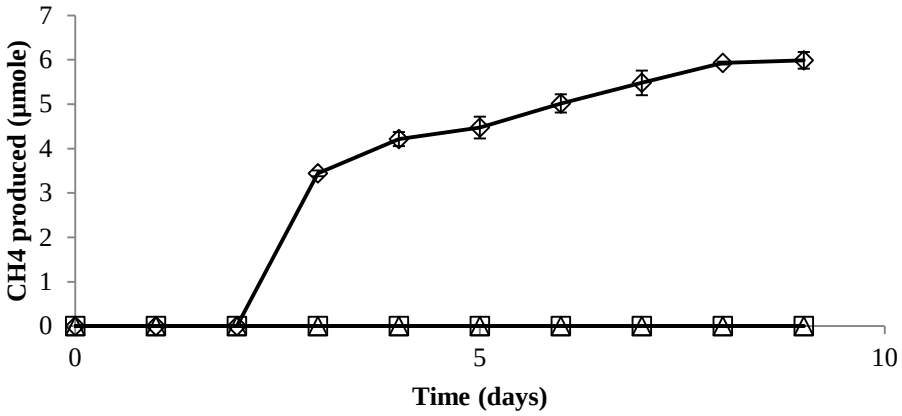


Figure 5.17. Growth of ISO4-H5 pure culture with putative reducing potential-supplying compound in absence of H_2. All replicates contain formate (10 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. (–◊–) methanol (10 mM), (–Δ–) methanol (10 mM) and ethanol (10 mM) without H_2, (–□–) methanol (10 mM) without H_2. Amount of CH_4 produced per mL of headspace by ISO4-H5 over nine days. Culture was re-pumped after day five with H_2/CO_2 (4:1; 200 kPa). Error bars represents ±SEMs.

Whether ethanol supplies reducing potential in the presence of H_2 was also explored in growth tests. The results show that the addition of 10 mM ethanol had no effect on CH_4 production or H_2 consumption (Figure 5.18A., B.). Some methanogens are capable of utilising formate by oxidising it to H_2 and CO_2 via the action of formate dehydrogenase. Although no homologue of formate dehydrogenase was identified in the ISO4-H5 genome, the effect of formate on CH_4 production was explored. The results show that ISO4-H5 did not utilise formate to produce CH_4 in presence of H_2 (Figure 5.16.), ISO4-H5 does not utilise formate in absence of H_2 to produce CH_4 (Figure 5.17.), and ISO4-H5 grows equally well in high or low formate (Figure 5.18A., B.).

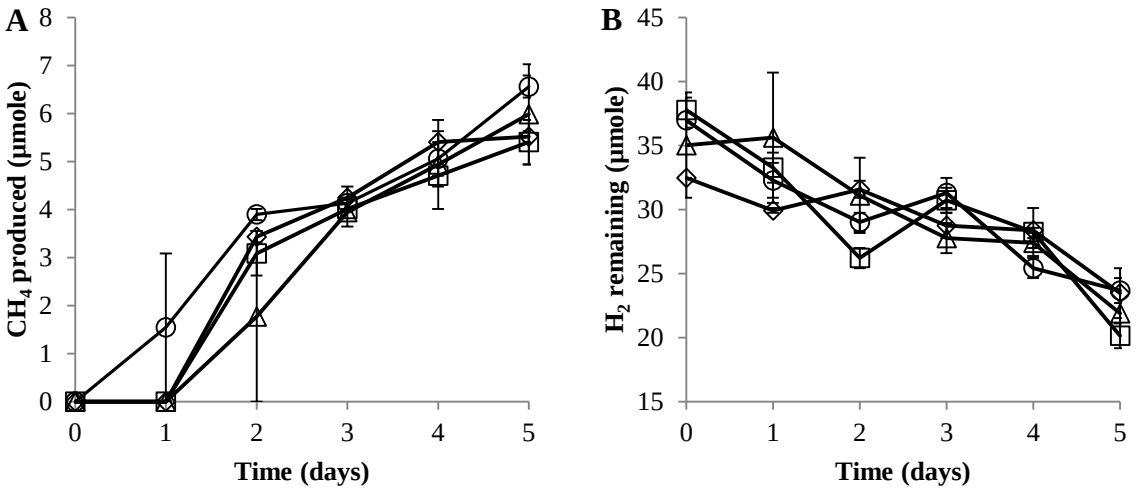


Figure 5.18. Effect of ethanol and formate addition on CH_4 formation in ISO4-H5 pure cultures. All replicates contain methanol (4 mM), acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix at H_2/CO_2 (4:1; 200 kPa) without SSPGMS. Cultures supplemented with or without additional formate and ethanol. (–◊–) ethanol (10 mM), formate (10 mM), (–Δ–) no additional ethanol, formate, (–□–) formate (10 mM), no additional ethanol, (–○–) ethanol (10 mM), no additional formate. There is a residual of 0.01 mM ethanol and 0.06 mM formate carried over from previous inoculations. **A.** Amount of CH_4 produced per mL of headspace by ISO4-H5 over five days. **B.** H_2 per mL of headspace remaining from ISO4-H5 growth on methanol over five days. Error bars represents ±SEMs.

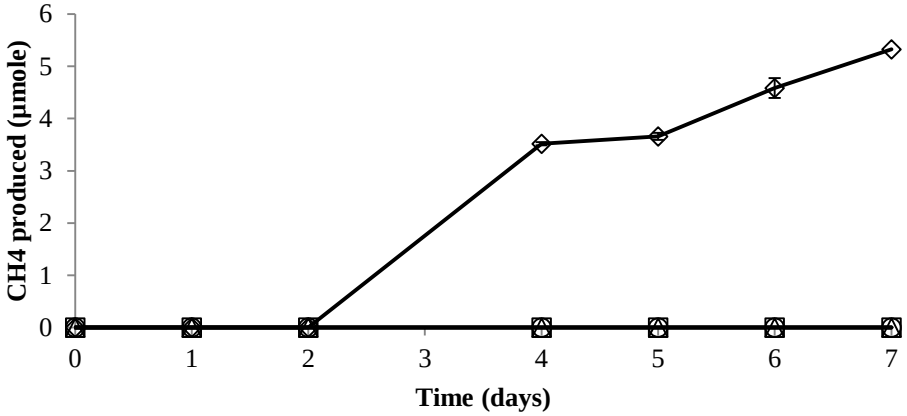


Figure 5.19. Testing growth of ISO4-H5 pure cultures on betaine or choline and H_2+CO_2 (4:1; 200 kPa). All replicates contain formate (60 mM), methanol (≤0.4 mM) carried over from previous inoculation, acetate (20 mM), CoM (10 µM) and 0.1× vitamin mix. (–◊–) Methanol (10 mM), (–Δ–) Betaine (10 mM), (–□–) Choline (10 mM), (–○–) Methanol (0 mM). Amount of CH_4 produced per mL of headspace by ISO4-H5 over seven days. Error bars represents ±SEMs.

The precursors of methylamines in the rumen are likely to be choline and/or betaine, therefore these compounds were explored as possible substrates for methanogenesis by ISO4-H5. Results from these growth tests showed that ISO4-H5 was unable to produce CH_4 using betaine or choline (Figure 5.19.).

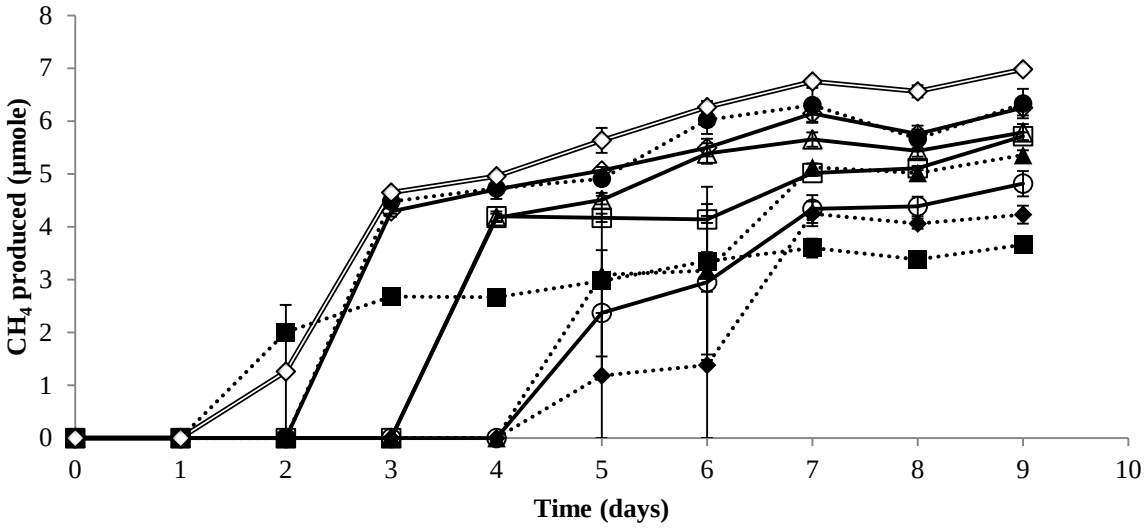


Figure 5.20. ISO4-H5 pure culture supplementation. Pure culture of ISO4-H5 supplemented with amino acids and other metabolites in triplicates. All replicates contain formate (60 mM), acetate (20 mM), methanol (20 mM), CoM (10 µM), 10% (v/v) SSPGMS and 0.1× vitamin mix. (–◊–) Control, (–Δ–) NaCl (1% w/v), (–□–) succinate (10 mM), (–○–) citrate (10 mM), (⋯♦⋯) malate (10 mM), (⋯▲⋯) fumarate (10 mM), (⋯■⋯) aspartate (10 mM), (⋯●⋯) glycine (10 mM), (=◆=) serine (10 mM). Amount of CH_4 produced per mL of headspace by ISO4-H5 over nine days. Error bars represents ±SEMs.

As the growth of ISO4-H5 *in vitro* was slow and sub-optimal, various supplements and metabolic intermediates were explored to identify growth limiting factors (Figure 5.20.). The betaine and choline were tested separately from other supplements due to requiring a valid negative control, where the carried-over methanol concentration is lower than what ISO4-H5 requires to survive, while other supplements were tested in presence of methanol as a methyl source. *Methanomassiliicoccus luminyensis* was reported as having optimal growth at higher NaCl concentrations, therefore 1% NaCl (w/v) was tested. The ISO4-H5 genome is predicted to encode an incomplete citric acid cycle, therefore succinate, citrate, malate, fumarate and aspartate (all at 10 mM) were tested as growth supplements, along with glycine and serine, as ISO4-H5 was predicted to lack a serine hydroxymethyltransferase gene required for glycine biosynthesis from serine. Results from these growth tests showed that ISO4-H5 does not produce more CH_4 with NaCl, succinate, citrate, malate, fumarate, serine or glycine supplementation (Figure 5.20.). While aspartate supplementation allowed earlier CH_4 formation in ISO4-H5, overall CH_4 production was lower than the unsupplemented control.

5.2.5. Analysis of ISO4-H5 gene expression during growth on different methanogenic substrates and under high or low H_2

Methanogen growth is dependent on the concentration of H_2 and the type of methyl substrate, therefore it was of interest to examine ISO4-H5 gene expression under conditions varying in

H_2 level and methyl substrate type. An experiment was conducted using the ISO4-H5 enrichment culture grown under high and low H_2 conditions. The high H_2 condition was provided using a headspace of H_2/CO_2 (4:1; 200 kPa), while the low H_2 environment was simulated by co-culture with the cellulose-degrading, H_2-producing bacterium, *Ruminococcus flavefaciens* FD1 grown on cellulose with a N_2 (200 kPa) atmosphere headspace (Figure 5.21C.). The experimental set up of the co-cultures is summarised in Table 5.3.

Table 5.3. Experimental set up of high and low H_2 co-cultures of ISO4-H5 enrichment grown with methanol or methylamine

Treatment	Co-culture description	Methanol	Methylamine	H_2	FD1*	Sdex*	ISO4-H5 enrichment
1	ISO4-H5 high H_2 + methanol	20 mM	-	High	-	-	+
2	ISO4-H5 + FD1 high H_2 + methanol	20 mM	-	High	+	-	+
3	ISO4-H5 + FD1 low H_2 + methanol	20 mM	-	Low	+	-	+
4	ISO4-H5 + FD1 low H_2 + methylamine	-	20 mM	Low	+	-	+
5	Sdex + FD1 high H_2	20 mM	-	High	+	+	-
6	Sdex + FD1 low H_2	20 mM	-	Low	+	+	-
7	FD1 high H_2	20 mM	-	High	+	-	-
8	FD1 low H_2	20 mM	-	Low	+	-	-
9	Sdex + FD1 low H_2 + methylamine	-	20 mM	Low	+	+	-

*FD1: *Ruminococcus flavefaciens* FD1. Sdex: *Succinivibrio dextrinosolvens* H5. Each culture was supplemented with 20 mM acetate, 1 mM CoM, 0.1× Vitamin 10 mix and 0.2% (w/v) cellulose.

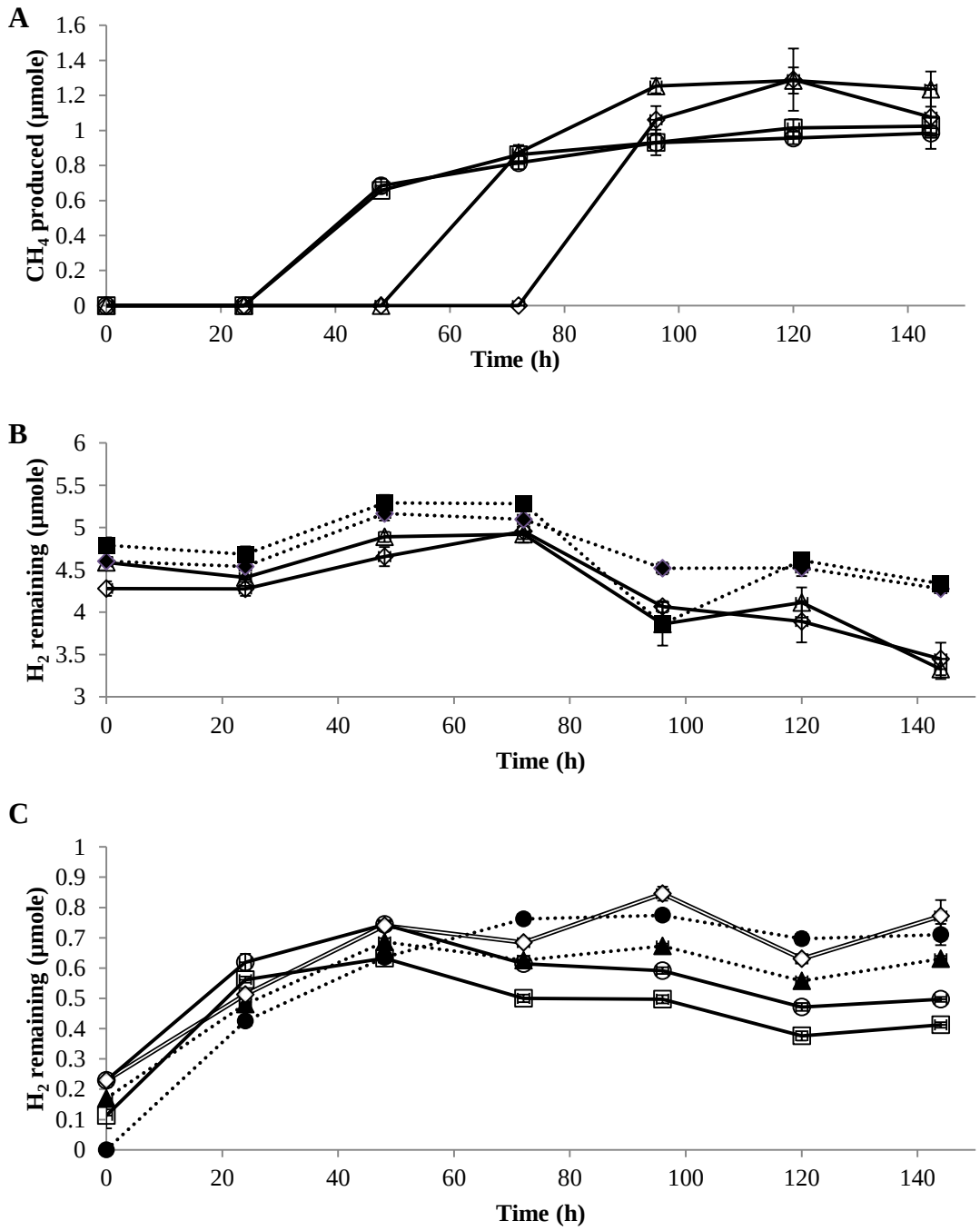


Figure 5.21. Growth of ISO4-H5 enrichment culture under high or low H_2 with either methanol or methylamine. (–◊–) ISO4-H5 high H_2 + methanol, (–Δ–) ISO4-H5 + FD1 high H_2 + methanol, (–□–) ISO4-H5 + FD1 low H_2 + methanol, (–○–) ISO4-H5 + FD1 low H_2 + methylamine, (···♦···) Sdex + FD1 high H_2, (···▲···) Sdex + FD1 low H_2, (···■···) FD1 high H_2, (···●···) FD1 low H_2, (=◊=) Sdex + FD1 low H_2 + methylamine. **A.** Amount of CH_4 produced per mL of headspace in ISO4-H5 enrichment culture under high or low H_2 conditions. **B.** Amount of H_2 per mL of headspace consumed by ISO4-H5. **C.** Amount of H_2 per mL of headspace produced by FD1 and consumed by ISO4-H5. Error bars represent ±SEMs.

CH_4 production was observed earliest when ISO4-H5 was grown on methanol under the FD1 + low H_2 condition but a higher final CH_4 concentration was observed in the stationary phase ISO4-H5 co-culture with FD1 under high H_2 conditions (Figure 5.21A.). Approximately one µmole of H_2 per mL of headspace was consumed by ISO4-H5 under high H_2 conditions (Figure

5.21B.). The H_2 produced from FD1 appears to be more than sufficient for ISO4-H5 growth, as accumulation of H_2 under low H_2 conditions was observed (Figure 5.21C.). The FD1 monoculture produced more H_2 than the co-culture of FD1 with *S. dextrinosolvens* H5 (Figure 5.21C.). ISO4-H5 grown on methanol and methylamine produced equivalent amounts of CH_4 under the low H_2 condition (Figure 5.21A.).

RNA extraction was carried out on culture samples taken at 120 h post-inoculation, as described in Chapter 2. Sufficient quantity of RNA for sequencing could only be collected from the cultures containing ISO4-H5. The quality of RNA in each replicate is displayed in Table 5.4. Three of the four replicates were transcriptome sequenced.

Table 5.4. Quality of RNA extracted from ISO4-H5 co-cultures

RNA sample	RNA integrity number
1-A (120 h)	N/A
1-C (120 h)	9.3
1-E (120 h)	9.2
1-F (120 h)	9.1
2-A (120 h)	9.3
2-C (120 h)	9.8
2-E (120 h)	8.9
2-H (120 h)	9.1
3-A (120 h)	8.9
3-B (120 h)	8.8
3-D (120 h)	8.8
3-H (120 h)	9
4-A (120 h)	9
4-B (120 h)	9
4-E (120 h)	9.2
4-F (120 h)	9.1

To calculate gene expression levels, the transcripts were trimmed and assembled to the ISO4-H5 reference genome at 98% similarity via the Rockhopper and Edge-pro pipelines (Table 5.5.). In a preliminary test of these two assemblers, the number of transcripts aligned to ISO4-H5 by Edge-pro was higher compared to Rockhopper, however, the Edge-pro alignment could not differentiate transcripts from the three co-cultured organisms, therefore Rockhopper was chosen for this task.

Table 5.5. Reads aligned to ISO4-H5, *S. dextrinosolvens* H5 and *R. flavefaciens* FD1 genomes using Rockhopper

Treatment description	Replicate	Total reads	Reads aligned to					
			ISO4-H5	% total	Sdex*	% total	FD1	% total
Trt1: ISO4-H5 high H_2 + methanol	1	9574814	7188267	75%	674275	7%		
	2	11171088	7396072	66%	1726545	15%		
	3	12762460	8004660	63%	2191541	17%		
Trt2: ISO4-H5 + FD1 high H_2 + methanol	1	8907622	5000664	56%	1266430	14%	1119464	13%
	2	8082221	5068906	63%	847775	10%	855331	11%
	3	9896974	6586875	67%	672412	7%	788202	8%
Trt3: ISO4-H5 + FD1 low H_2 + methanol	1	8704366	6066984	70%	480305	6%	376550	4%
	2	10191472	6643413	65%	613478	6%	757425	7%
	3	8012444	5698164	71%	457562	6%	448399	6%
Trt4: ISO4-H5 + FD1 low H_2 + methylamine	1	11000283	8689292	79%	643327	6%	51385	0.5%
	2	9758529	7610170	78%	646486	7%	99859	1%
	3	11799619	9246903	78%	860353	7%	62816	1%

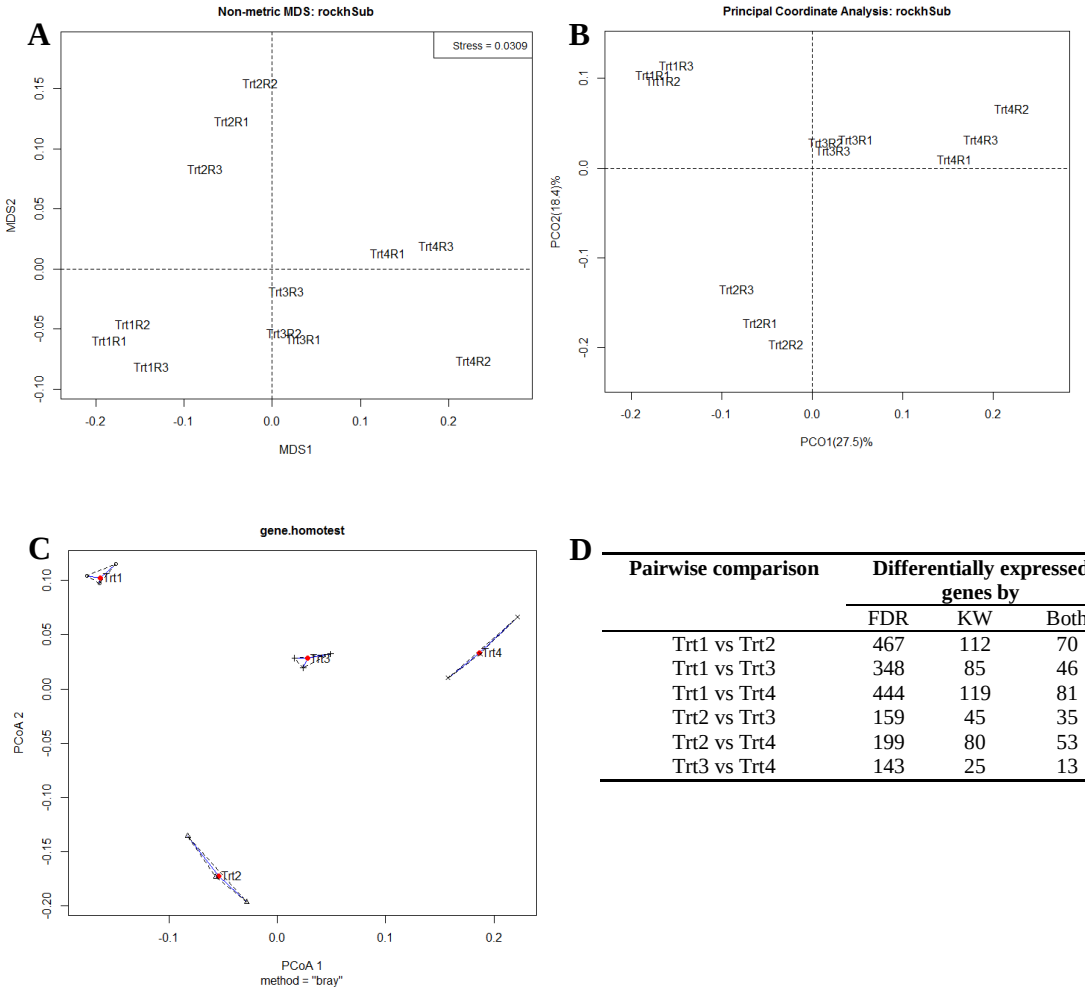
*Sdex: *S. dextrinosolvens* H5

5.2.6. Statistical analyses

The read counts of mapped transcripts were normalized and differential expression analyses were performed by Rockhopper, using a false discovery rate (FDR) cutoff of <0.05, a non-parametric Kruskal-Wallis rank sum test (KW) *P*-value cutoff of <0.05, as well as fold change >2, to define differentially expressed genes between any two conditions. Gene expression profiles of the three genomes are displayed in Tables A.5.3, A.5.4 and A.5.5.

To investigate the difference between conditions as well as replicates, multivariate analyses were performed on genes differentially expressed between two conditions (Figures 5.22., 5.23., 5.24.). The MDS plot of differentially expressed genes in ISO4-H5 showed clustering of replicates within Trt1, Trt2 and Trt3 and more dissimilarity was observed between replicates of Trt4. The *S. dextrinosolvens* H5 MDS plot showed similarity between replicates of Trt2 and Trt4, and replicate one of Trt1 appeared less similar to the two other replicates. The MDS plot of differentially expressed genes in *R. flavefaciens* FD1 displayed strong dissimilarity among replicates of Trt4.

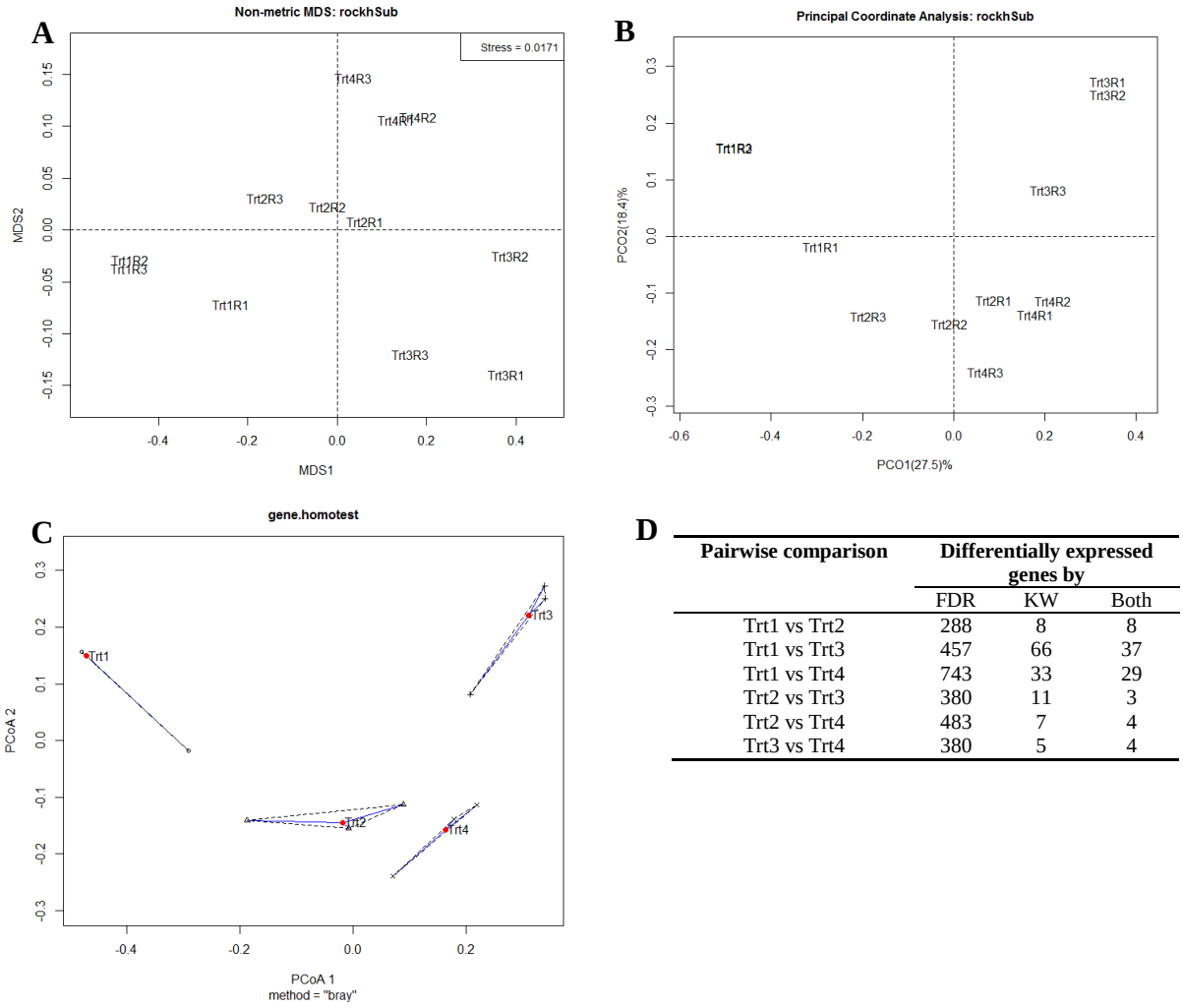
The PCoA plots of differentially expressed genes in ISO4-H5 and *R. flavefaciens* FD1 showed clustering of replicates within treatments, whereas with *S. dextrinosolvens* H5 replicates of Trt2 and Trt4 do not display obvious clustering and is similar even between treatments, while replicate 1 of Trt1 and replicate 3 of Trt3 displayed dissimilarity to other replicates within the treatment. The homogeneity plot of differentially expressed genes in ISO4-H5 and *R. flavefaciens* FD1 showed low transcriptional diversity within treatments, whereas larger dispersion within treatments was observed in *S. dextrinosolvens* H5.



D.

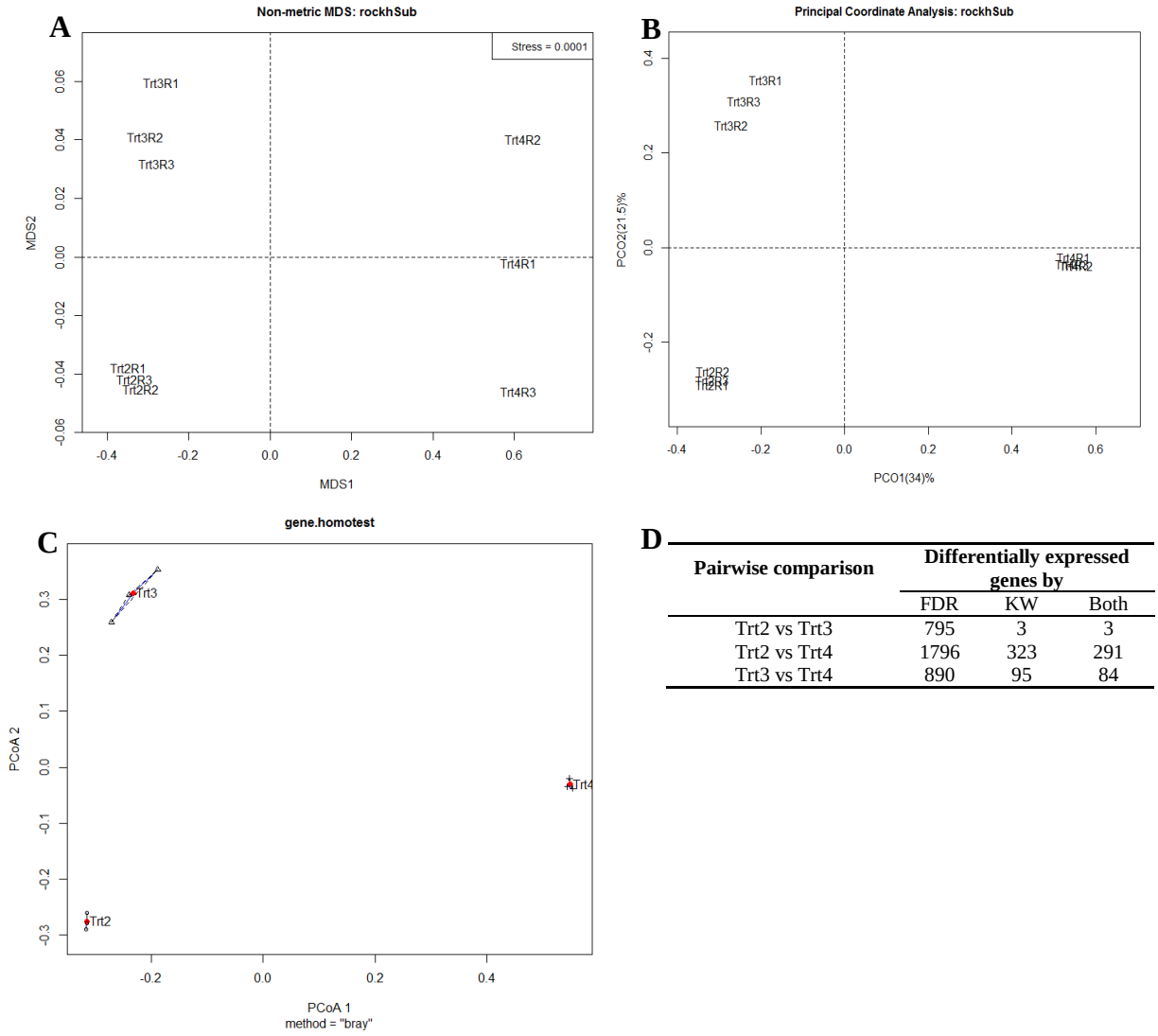
Pairwise comparison	Differentially expressed genes by		
	FDR	KW	Both
Trt1 vs Trt2	467	112	70
Trt1 vs Trt3	348	85	46
Trt1 vs Trt4	444	119	81
Trt2 vs Trt3	159	45	35
Trt2 vs Trt4	199	80	53
Trt3 vs Trt4	143	25	13

Figure 5.22. Multivariate analysis based on relative abundance of transcripts in the ISO4-H5 transcriptome via Bray Curtis distance matrix. **A.** Nonmetric multidimensional scaling plot **B.** PCoA plot. **C.** Group dispersions plot by permutation-based test of multivariate homogeneity. **D.** Table of numbers of differentially expressed genes determined by FDR and KW. Trt1-4 are as described in Table 5.3.



Pairwise comparison	Differentially expressed genes by		
	FDR	KW	Both
Trt1 vs Trt2	288	8	8
Trt1 vs Trt3	457	66	37
Trt1 vs Trt4	743	33	29
Trt2 vs Trt3	380	11	3
Trt2 vs Trt4	483	7	4
Trt3 vs Trt4	380	5	4

Figure 5.23. Multivariate analysis based on relative abundance of transcripts in the *S. dextrinosolvens* H5 transcriptome via Bray Curtis distance matrix. **A.** Nonmetric multidimensional scaling plot. **B.** PCoA plot. **C.** Group dispersions plot by permutation-based test of multivariate homogeneity. **D.** Table of differentially expressed genes. Trt1-4 are as described in Table 5.3.



D

Pairwise comparison	Differentially expressed genes by		
	FDR	KW	Both
Trt2 vs Trt3	795	3	3
Trt2 vs Trt4	1796	323	291
Trt3 vs Trt4	890	95	84

Figure 5.24. Multivariate analysis based on relative abundance of transcripts in the *R. flavefaciens* FD1 transcriptome via Bray Curtis distance matrix. **A.** Nonmetric multidimensional scaling plot. **B.** PCoA plot. **C.** Group dispersions plot by permutation-based test of multivariate homogeneity. **D.** Table of differentially expressed genes. Trt1-4 are as described in Table 5.3.

Correspondence analysis was also used to explore the association between treatments and differentially expressed genes in ISO4-H5. At a CA threshold of 0.3, only genes positively associated with Trt1 and Trt4 were observed (Figure 5.25.).

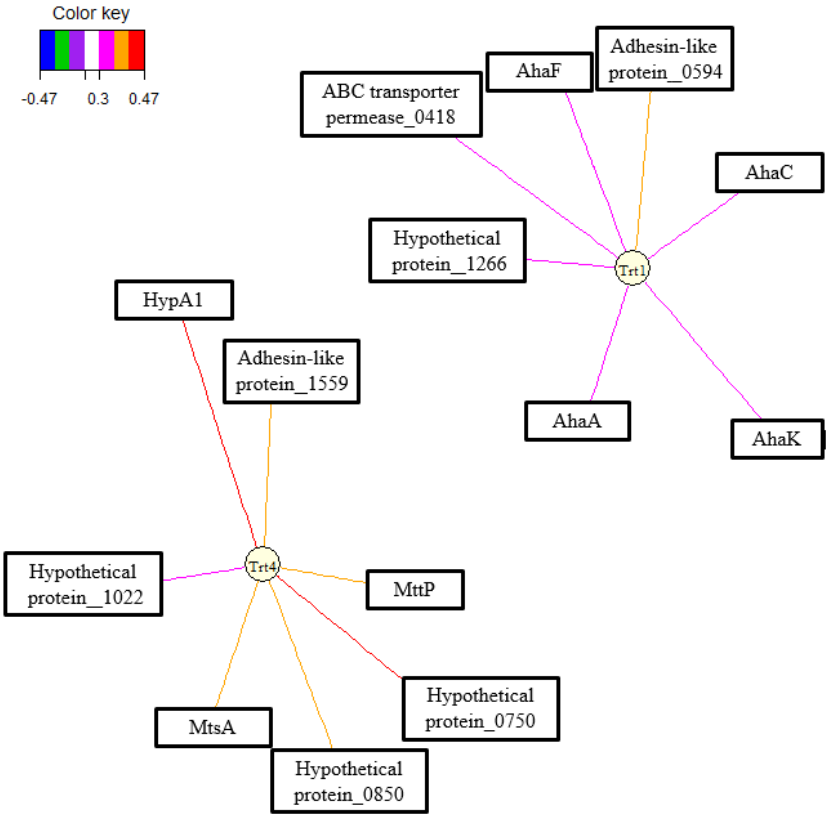


Figure 5.25. Network diagram showing the associations between genes and treatments based on Correspondence Analysis. The threshold used was 0.3. Solid lines represents positive association. Line colours signify magnitude of association as illustrated by the key at top left. A_1A_0 ATP synthase subunits (Aha). Trimethylamine permease (MttP). Methylthiol:corrinoid methyltransferase (MtsA). Hydrogenase nickel insertion protein (HypA1).

5.2.7. Differential gene expression in ISO4-H5

There were 298 differentially expressed ISO4-H5 genes that fit the aforementioned statistical cut-offs (Figure 5.22D., Table A.5.3.). In order to assess the ISO4-H5 response to H_2 level while using methanol, the 35 genes differentially expressed between Trt2 (high H_2) and Trt3 (low H_2) were investigated first (Table 5.6.).

Table 5.6. ISO4-H5 genes differentially expressed between Trt2 (high H_2) and Trt3 (low H_2)

Locus_tag	Product	Expression (RPK0.1M)			
		Trt1	Trt2	Trt3	Trt4
AR505_0038	ribosomal RNA large subunit methyltransferase J RrmJ	144	245	113	140
AR505_0126	transmembrane protein	118	172	56	92
AR505_0199	transcriptional regulator AsnC family	72	45	101	45
AR505_0225	ribosomal protein L29P Rpl29p	235	81	301	177
AR505_0226	translation initiation factor aSUI1	298	65	358	163
AR505_0234	ribosomal protein L6P Rpl6p	168	51	178	94
AR505_0273	CoB-CoM heterodisulfide reductase subunit C HdrC	93	79	681	166
AR505_0274	CoB-CoM heterodisulfide reductase subunit B HdrB1	182	110	836	279
AR505_0275	TPR repeat-containing protein	13	8	38	16
AR505_0518	hypothetical protein	37	20	45	21
AR505_0536	TPR repeat-containing protein	122	195	88	125
AR505_0675	acetylglutamate kinase ArgB	29	15	48	34
AR505_0689	transmembrane protein	150	235	104	148
AR505_0909	adhesin-like protein	128	263	114	164
AR505_1118	MTA/SAH nucleosidase MtnN	35	61	26	45
AR505_1155	quinolinate synthetase A protein NadA	113	196	98	186
AR505_1187	hypothetical protein	221	1091	159	474
AR505_1188	transmembrane protein	67	125	45	102
AR505_1216	TPR repeat-containing protein	63	98	42	51
AR505_1233	pseudouridylate synthase	48	87	38	63
AR505_1351	TPR repeat-containing protein	51	92	40	59
AR505_1386	methanogenesis marker protein 17	381	118	411	209
AR505_1387	methanogenesis marker protein 15	256	95	298	136
AR505_1388	methanogenesis marker protein 5	101	48	131	53
AR505_1395	hypothetical protein	1683	999	4926	1831
AR505_1396	methyl-CoMreductase alpha subunit MrtA	1696	953	2883	1328
AR505_1476	methyl viologen-reducing hydrogenase alpha subunit MvhA	436	320	1771	861
AR505_1477	methyl viologen-reducing hydrogenase gamma subunit MvhG	515	364	1870	1014
AR505_1478	methyl-viologen-reducing hydrogenase delta subunit	400	361	2055	1072
AR505_1479	CoB-CoM heterodisulfide reductase subunit A HdrA	381	327	1194	748
AR505_1583	tRNA nucleotidyltransferase Cca	68	138	54	108
AR505_1648	transketolase subunit B	120	65	170	87
AR505_1661	hypothetical protein	87	170	56	88
AR505_1675	radical SAM domain-containing protein	85	114	50	89
AR505_1782	DNA primase large subunit PriB	128	187	93	102

The expression of these genes and the location of their corresponding enzymes in the methylotrophic methanogenesis pathway in ISO4-H5, are summarised in Figure 5.26. There were 19 genes with higher expression under low H_2 compared to high H_2, including *mrtA* (3× up-regulated) which is central to methanogenesis, *hdrA* (3×), *hdrB* (8×), *hdrC* (9×) involved in cofactor regeneration and *mvhAGD* (5×) involved in supplying reducing potential for methanogenesis. Four additional genes associated with methanogenesis also had higher expression under low H_2: AR505_1395 (encoding a hypothetical protein immediately downstream of *mcrA*, 5×) and 3 genes encoding methanogenesis marker proteins (MMP) *mmp5*, *mmp15* and *mmp17* (3×). In addition, one set of methanol utilizing genes (*mtaB2*, *mtaC2*) and a pair of Fdxs genes *fdx1* and *fdx2* (AR505_0452, AR505_0453) have 3× higher expression under low H_2 which meet the FDR cutoff (*mtaB2*: 3.2×10^{-4}, *mtaC2*: 5.3×10^{-5}, *fdx1*: 1.4×10^{-6}, *fdx2*: 2.0×10^{-7}) but not the KW *p* value cutoff (*mtaB2*: 0.22, *mtaC2*: 0.14, *fdx1*: 0.14, *fdx2*: 0.14). Among the 16 genes that showed higher expression under high H_2,

hypothetical protein AR505_1187 had the highest fold change of 7×. Other genes differentially expressed were involved in replication and protein synthesis, a *nadA* gene involved in NAD biosynthesis from aspartate, genes presumed to be involved in regulation and yet other genes with unknown functions (Table 5.6.).

In order to assess the ISO4-H5 response to different methanogenic substrates, growth on methylamine (Trt4) was compared with methanol (Trt3). Only 13 genes were differentially expressed in this comparison, including genes encoding three Fpo-like complex proteins (*fpoN*, *fpoM* and *fpoL*). No genes directly associated with methanogenic substrate utilisation were differentially expressed under these conditions. Therefore, the genes encoding methylamine and methylthiol utilisation were investigated across all four treatments (Figure 5.26.). Higher expression of genes encoding methylamine use, such as *mtmB2* (8×), *mtmC1* (6×), dimethylamine utilising *mtbC* (3×), *mtbB* (3×), *mtbP1* (13×) and trimethylamine transporting *mttP* (42×) as well *mtsA* encoding methylthiol use, were observed in Trt4 (methylamine under low H_2), in comparison to either Trt1 or Trt2 (Figure 5.27., Table A.5.3.). Genes other than those encoding the methanogenesis pathway may also be involved in adaptation to growth on different methanogenic substrates, therefore genes with two-fold difference in expression in Trt4 compared to all three other treatments were investigated and are summarised in Table 5.7. Eight out of 11 subunits of the Fpo-like complex were found to have low expression in Trt4. However, the expression of Fpo-like complex genes is likely influenced by multiple factors, as illustrated in Figure 5.29A.

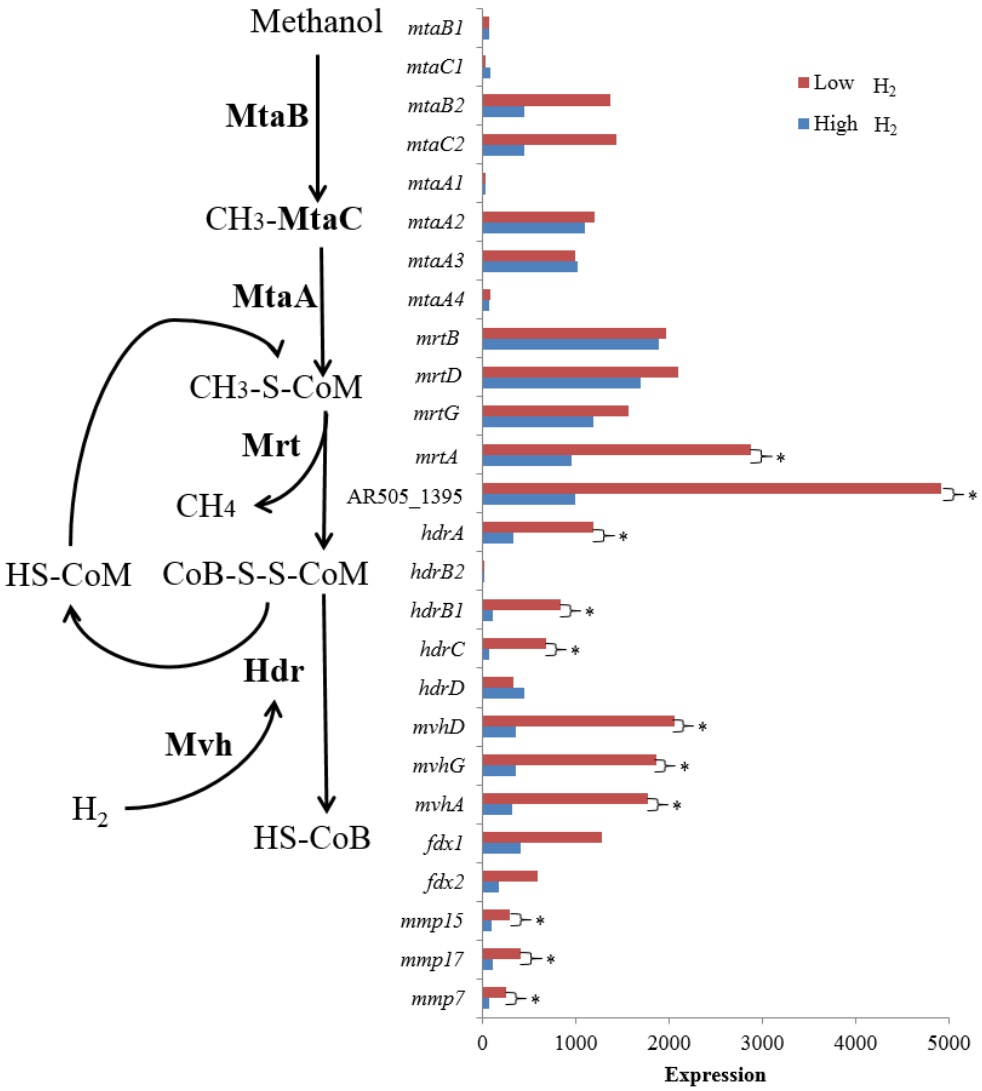


Figure 5.26. Expression of genes involved in methanol utilizing methanogenesis under high and low H_2 conditions. *: Differentially expressed genes with FDR <0.05, KW p <0.05, fold change >2 between Trt2 (high H_2) and Trt3 low H_2). Methanol:corrinoid methyltransferase (MtaB), methanol corrinoid protein (MtaC), Methyl: CoM methyltransferase (MtaA), methyl-CoM reductase complex (Mcr), hypothetical protein immediate downstream of *mcrA* gene (AR505_1395), heterodisulfide reductase (Hdr), methyl viologen hydrogenase (Mvh), ferredoxin (Fdx), methanogenesis marker protein (MMP). Gene products are displayed in bold within the pathway.

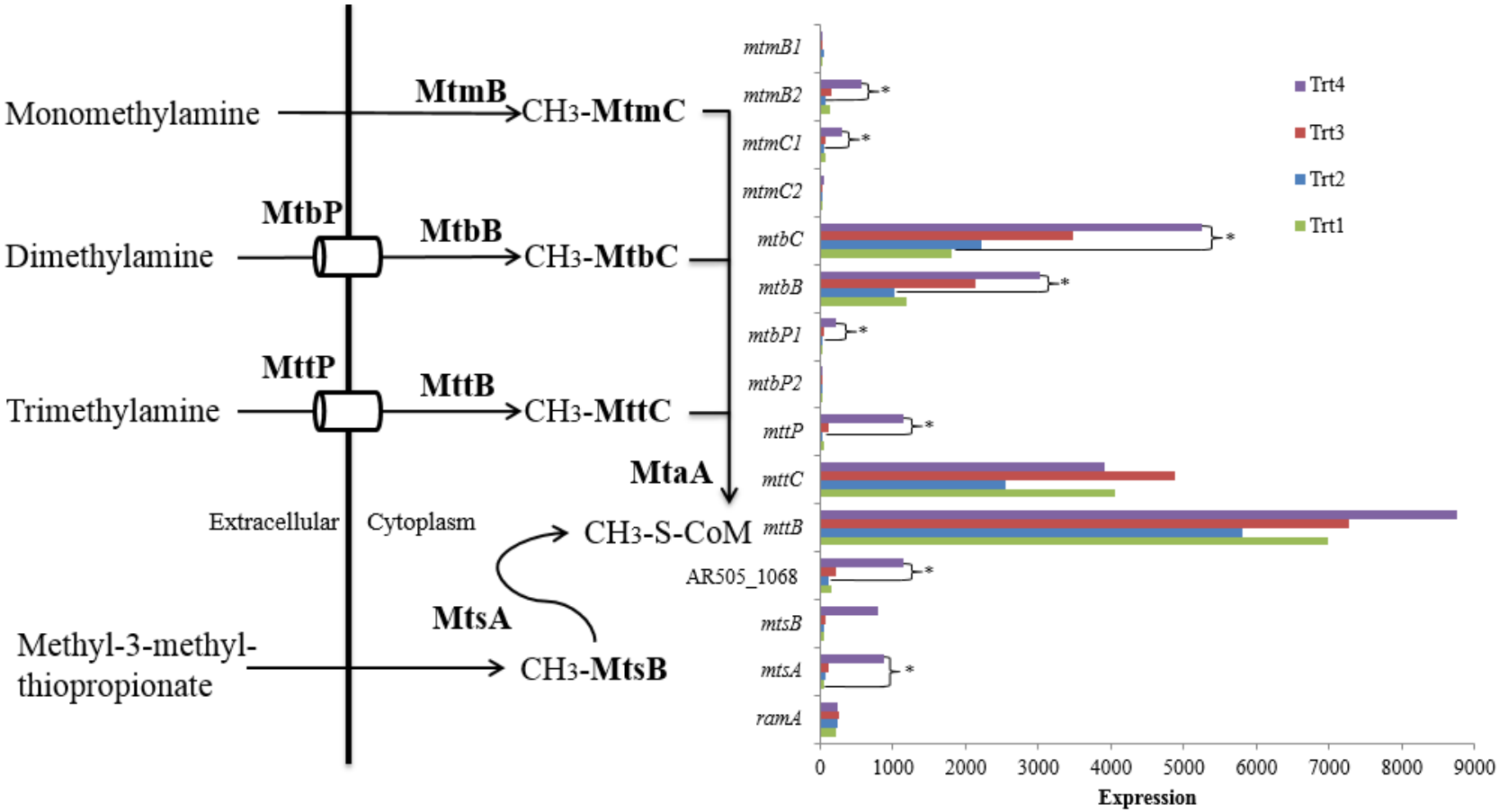


Figure 5.27. Expression of ISO4-H5 genes involved in methylamine and methylthiol use. *: Differentially expressed genes with FDR <0.05, KW p <0.05, fold change >2. Trt1: ISO4-H5 enrichment culture on methanol with high H_2. Trt2: ISO4-H5 enrichment culture +FD1 on methanol with high H_2. Trt3: ISO4-H5 enrichment culture +FD1 on methanol with low H_2. Trt4: ISO4-H5 enrichment culture +FD1 on monomethylamine with low H_2. Monomethylamine:corrinoid methyltransferase (MtmB), monomethylamine corrinoid protein (MtmC), dimethylamine:corrinoid methyltransferase (MtbB), dimethylamine corrinoid protein (MtbC), dimethylamine permease (MtbP). Trimethylamine:corrinoid methyltransferase (MttB), trimethylamine corrinoid protein (MttC), trimethylamine permease (MttP), 4Fe-4S Fdx iron-sulfur binding domain-containing protein (AR505_1068), methylthiol corrinoid protein (MtsB), bifunctional methylthiol:corrinoid methyltransferase (MtsA), methylamine methyltranferase corrinoid activation protein (RamA). Gene products are displayed in bold within the pathway. Treatment groups are defined in Table 5.3.

Table 5.7. Genes with 2-fold expression differences in Trt4 compared with other treatments

Locus_tag	Product	Expression (RPK0.1M)			
		Trt1	Trt2	Trt3	Trt4
Genes with 2-fold higher expression					
AR505_0443	universal stress protein UspA	57	62	182	475
AR505_0609	low molecular weight phosphotyrosine protein phosphatase	35	53	72	157
AR505_0749	trimethylamine permease MttP	47	27	119	1143
AR505_0750	hypothetical protein	33	27	72	955
AR505_0799	*O*-acetylserine sulfhydrylase CysM	20	15	13	43
AR505_0850	hypothetical protein	10	3	18	140
AR505_0859	transcriptional regulator TetR family	6152	7818	14364	31275
AR505_1008	SAM-dependent methyltransferases	16	42	34	89
AR505_1022	hypothetical protein	38	22	34	245
AR505_1066	bifunctional Methylthiol:corrinoid methyltransferase MtsA	52	63	105	881
AR505_1068	4Fe-4S Fdx iron-sulfur binding domain-containing protein	149	116	222	1139
AR505_1185	hydrogenase nickel insertion protein HypA1	14	15	34	1017
AR505_1242	dimethylamine permease MtbP1	29	17	58	223
AR505_1310	small multidrug resistance protein	30	11	34	163
AR505_1311	small multidrug resistance protein	29	5	30	99
AR505_1328	monomethylamine methyltransferase MtmB2	132	72	145	559
AR505_1329	methyltransferase cognate corrinoid protein MtmC1	75	49	76	291
AR505_1559	adhesin-like protein	35	28	48	418
Genes with 2-fold lower expression					
AR505_0183	hypothetical protein	43	115	63	21
AR505_0396	hypothetical protein	46	64	59	16
AR505_0466	HTH/CBS domain-containing protein	52	20	12	4
AR505_0471	phosphopyruvate hydratase Eno	48	55	28	21
AR505_1053	hypothetical protein	115	143	115	49
AR505_1219	transposase IS605 OrfB family	99	124	112	42
AR505_1622	$F_{420}H_2$ dehydrogenase subunit N FpoN	538	210	705	86
AR505_1623	$F_{420}H_2$ dehydrogenase subunit M FpoM	359	150	411	50
AR505_1624	$F_{420}H_2$ dehydrogenase subunit L FpoL	541	232	589	86
AR505_1626	hypothetical protein	453	181	439	66
AR505_1628	$F_{420}H_2$ dehydrogenase subunit I FpoI	633	214	546	82
AR505_1629	$F_{420}H_2$ dehydrogenase subunit H FpoH	376	149	289	60
AR505_1630	$F_{420}H_2$ dehydrogenase subunit D FpoD	721	317	468	134
AR505_1631	$F_{420}H_2$ dehydrogenase subunit C FpoC	765	437	519	135
AR505_1632	$F_{420}H_2$ dehydrogenase subunit B FpoB	373	230	261	69
AR505_1718	hypothetical protein	38	63	82	17
AR505_1759	ribosomal protein L7Ae Rpl7ae	466	268	221	50
AR505_1766	hypothetical protein	194	182	151	60

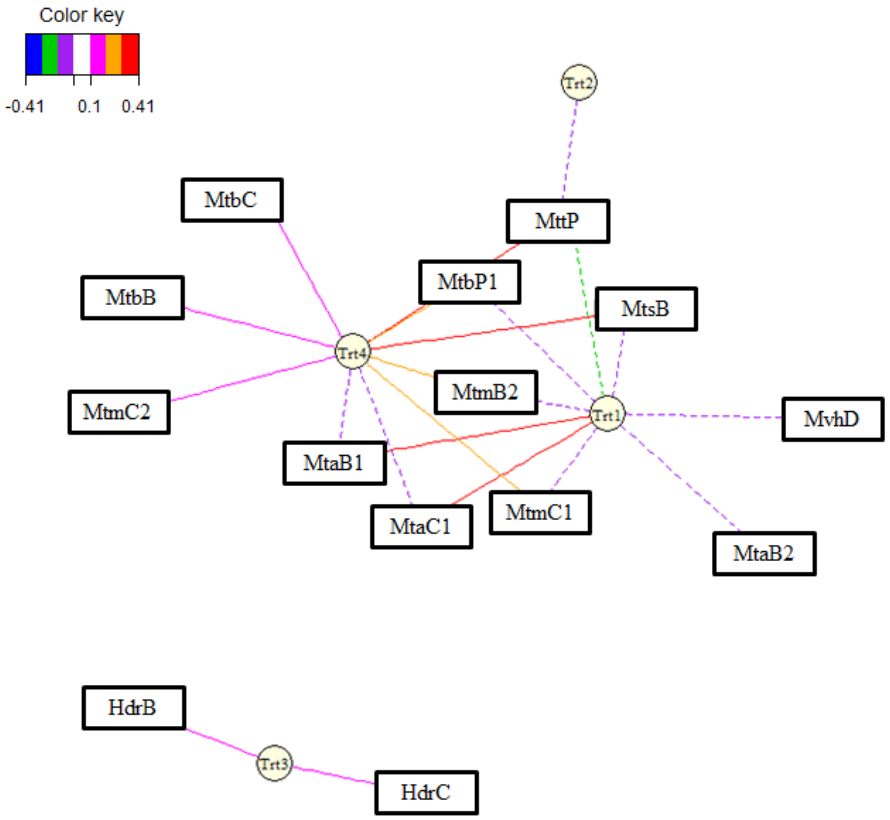


Figure 5.28. Network diagram showing associations between methanogenesis genes and treatments based on Correspondence Analysis. The threshold used was 0.1. Solid lines represents positive association, dotted lines represent negative associations. Line colours signify magnitude of association as illustrated by the key. Methanol:corrinoid methyltransferase (MtaB), methanol corrinoid protein (MtaC), monomethylamine:corrinoid methyltransferase (MtmB), monomethylamine corrinoid protein (MtmC), dimethylamine:corrinoid methyltransferase (MtbB), dimethylamine corrinoid protein (MtbC), dimethylamine permease (MtbP), trimethylamine permease (MttP), heterodisulfide reductase (Hdr). Treatment groups are defined in Table 5.3.

The association between methanogenesis genes and treatments used was examined using, Correspondence Analysis (Figure 5.28.). Trt1 and Trt4 showed the greatest number of associations with methanogenesis genes, while only one gene was associated with Trt2 and two genes were associated with Trt3.

R. flavefaciens FD1 was used to provide H_2 continuously at a low concentration, therefore the influence of FD1 on the expression of ISO4-H5 was assessed by comparing genes showing two-fold difference in expression between Trt1 and the other three treatments (Table 5.8.). Among the genes with higher expression in Trt1 were all nine subunits of the A_1A_0 ATP synthase complex (Figure 5.28B.), which were strongly up-regulated in the absence of *R. flavefaciens*.

Table 5.8. Genes with 2-fold expression differences in Trt1 compared with other treatments

Locus_tag	Product	Expression (RPK0.1M)			
		Trt1	Trt2	Trt3	Trt4
Genes with higher expression in Trt1					
AR505_0005	adhesin-like protein	462	152	153	76
AR505_0044	tRNA 2'-*O*-methylase	170	48	65	71
AR505_0093	molecular chaperone GrpE	309	134	151	100
AR505_0094	chaperone protein DnaK1	925	302	420	275
AR505_0114	methylase involved in ubiquinone/menaquinone biosynthesis	103	37	27	47
AR505_0201	Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit B GatB	72	21	34	25
AR505_0287	sodium/proline symporter PutP	141	56	39	32
AR505_0418	ABC transporter permease protein	70	10	8	11
AR505_0466	HTH/CBS domain-containing protein	52	20	12	4
AR505_0594	transmembrane protein	326	57	41	26
AR505_0601	hydroxymethylglutaryl-CoA synthase	672	197	271	188
AR505_0602	acetyl-CoA acetyltransferase	1081	337	480	277
AR505_0603	DNA-binding protein	529	141	257	91
AR505_0605	ABC transporter permease protein	142	26	38	66
AR505_0897	hypothetical protein	166	72	60	33
AR505_0954	hypothetical protein	147	30	46	45
AR505_1016	hypothetical protein	134	41	21	30
AR505_1055	site-specific recombinase	97	13	10	29
AR505_1092	CRISPR-associated protein Cas7/Cse4/CasC	67	12	21	10
AR505_1177	MMPL domain-containing protein	69	15	10	19
AR505_1222	ferrous iron transport protein B FeoB	89	43	16	32
AR505_1263	molybdenum cofactor biosynthesis protein MoaA1	189	24	64	35
AR505_1266	hypothetical protein	283	53	35	27
AR505_1297	cobalamin 5'-phosphate synthase CobS	115	45	39	26
AR505_1369	ATP-dependent DNA helicase	73	19	31	32
AR505_1758	ribosomal protein S28e Rps28e	205	92	87	49
AR505_1818	A_1A_0 ATP synthase subunit D	402	57	188	64
AR505_1819	A_1A_0 ATP synthase subunit B	353	40	131	44
AR505_1820	A_1A_0 ATP synthase subunit A	449	48	118	55
AR505_1821	A_1A_0 ATP synthase subunit F	211	29	44	33
AR505_1822	A_1A_0 ATP synthase subunit C	522	56	132	37
AR505_1823	A_1A_0 ATP synthase subunit E	395	71	150	55
AR505_1824	A_1A_0 ATP synthase subunit K	600	92	172	66
AR505_1825	A_1A_0 ATP synthase subunit I	722	189	193	107
AR505_1826	A_1A_0 ATP synthase subunit H	660	261	220	151
Genes with lower expression in Trt1					
AR505_0059	radical SAM domain protein	77	161	192	177
AR505_0377	NAD synthetase NadE	36	107	129	95
AR505_0619	transmembrane protein	11	660	349	727
AR505_0740	TATA-box-binding protein Tbp	68	258	140	207
AR505_0931	phage integrase	19	164	81	111
AR505_1008	SAM-dependent methyltransferase	16	42	34	89
AR505_1108	hypothetical protein	198	524	740	498
AR505_1157	hypothetical protein	4	48	27	23
AR505_1667	5-formaminoimidazole-4-carboxamide-1-(beta)-D- ribofuranosyl 5'-monophosphate-formate ligase PurP	116	460	365	302

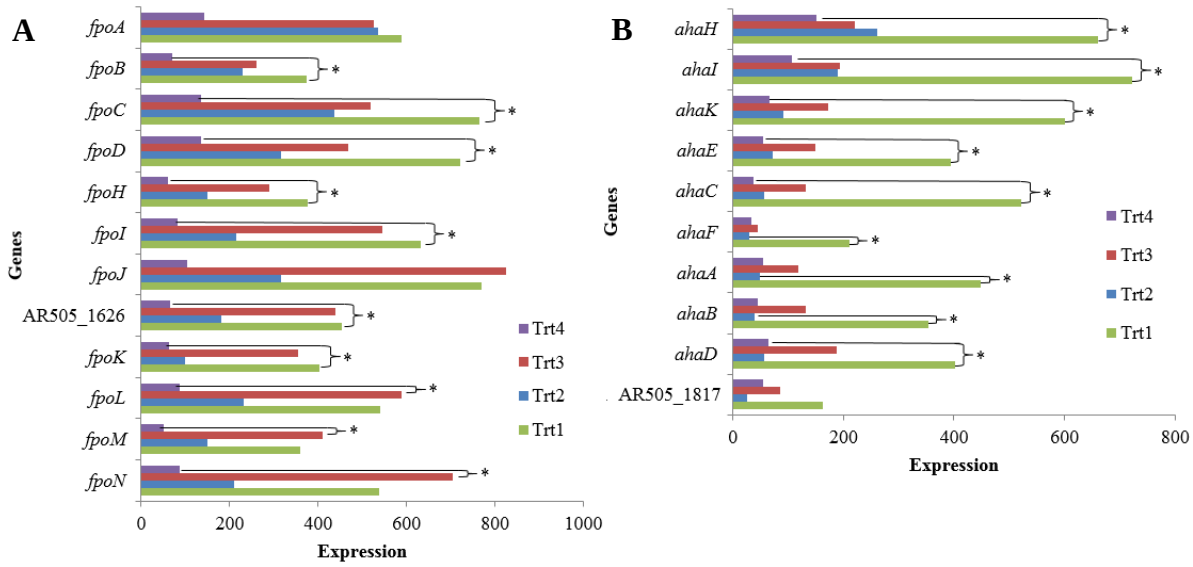


Figure 5.29. Differentially expressed ISO4-H5 genes during growth under low and high H_2 levels and different methyl sources. **A.** Expression of genes encoding the Fpo-like complex. Fpo-like complex subunits (*fpo*). hypothetical transmembrane protein (AR505_1626). **B.** Expression of genes encoding A_1A_0 ATP synthase between treatments. A_1A_0 ATP synthase subunits (*aha*). hypothetical protein immediate downstream of *ahaD* gene (AR505_1817). *: Differentially expressed genes with FDR <0.05, KW p <0.05, fold change >2. Treatment groups are defined in Table 5.3.

During the analysis of gene expression across the ISO4-H5 genome, a region from AR505_0313 to AR505_0358 (estimated region 32.3 kb) was observed displayed no gene expression at all across all the treatments. This region was identified previously in Chapter 3 as being potentially horizontally transferred, and carries several elements consistent with an integrated prophage sequence. To assess whether this region contains a prophage sequence, PCR amplification was carried out on freshly isolated ISO4-H5 gDNA with primers (H5PHX.p1ca, H5PHX.q1ca) located outside the putative prophage region facing inwards, and with five primers located within the region. A product of only ~850 bp was obtained from the inward facing PCR, and all PCR amplifications within the putative prophage region failed to give a PCR product (Figure 5.30.). This suggested that the putative prophage sequence was no longer present in the ISO4-H5 gDNA used for the PCR reactions. It is likely that this region had self-excised from the ISO4-H5 genome during the period of sub-culture between the original gDNA extraction for genome sequencing and this most recent gDNA extraction. This putative prophage region was therefore removed from the subsequent statistical assessments.

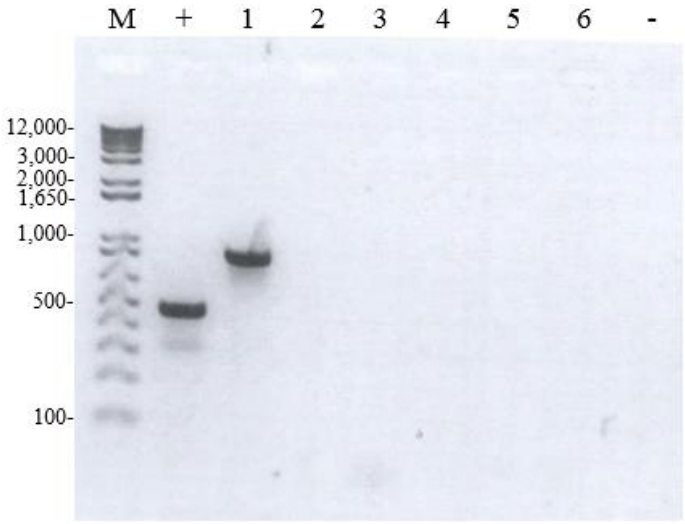


Figure 5.30. PCR amplifications flanking and within the AR505_0313 - AR505_0358 region. **M**: 1kb+ Marker, sizes are displayed in bp. **+**: Positive control 16*S* rRNA region amplified with primer pair 915af and 1386r. **1**: PCR using primer pairs H5PHX.p1ca, H5PHX.q1ca flanking the AR505_0313 – AR505_0358 region. **2**: PCR using primer pairs H5PH1.p1ca, H5PH1.q1ca within the AR505_0313 – AR505_0358 region. **3**: PCR using primer pairs H5PH2.p1ca, H5PH2.q1ca within the AR505_0313 – AR505_0358 region. **4**: PCR using primer pairs H5PH3.p1ca, H5PH3.q1ca within the AR505_0313 – AR505_0358 region. **5**: PCR using primer pairs H5PH4.p1ca, H5PH4.q1ca within the AR505_0313 – AR505_0358 region. **6**: PCR using primer pairs H5PH5.p1ca, H5PH5.q1ca within the AR505_0313 – AR505_0358 region. **-**: negative control PCR omitting genomic DNA.

5.2.8 Differential gene expressions in *S. dextrinosolvens* H5

A total of 85 *S. dextrinosolvens* H5 genes were differentially expressed between the four different treatments used in the ISO4-H5 co-culture experiments (Figure 5.22D., Table A.5.4.). of which 74 are differentially expressed between Trt1 and Trt2, Trt1 and Trt3, Trt1 and Trt4, which is contributed by presence and absence of FD1, but only 11 genes were differentially expressed between Trt2, Trt3 and Trt4. Noteworthy genes with higher expression in absence of FD1 includes succinate dehydrogenase and fumarate reductase Fe-S protein (T508DRAFT_01506), fumarate reductase (T508DRAFT_01507) and malate dehydrogenase (T508DRAFT_00914) involved in TCA cycle, pyruvate kinase (T508DRAFT_01509) and phosphoglycerate mutase (T508DRAFT_01623) involved in glycolysis, sugar phosphate permease (T508DRAFT_00592), α-1,4-glucan:α-1,4-glucan 6-glycosyltransferase (T508DRAFT_01412) and glycogen/starch/α-glucan phosphorylase (T508DRAFT_00525) involved in carbohydrate utilisation, DNA polymerase III ε subunit (T508DRAFT_00876) involved in replication, ribosomal protein S1 (T508DRAFT_01805), S9 (T508DRAFT_00953), L18 (T508DRAFT_00179), S11 (T508DRAFT_00172), S13 (T508DRAFT_00173), translation initiation factor IF-2 (T508DRAFT_01537) and translation elongation factor TU (T508DRAFT_02129, T508DRAFT_02302) involved in translation, and

conjugative transfer signal peptidase *traF* (T508DRAFT_00891) involved in conjugative transfer. Only five genes were observed with higher expression in presence of FD1, hypothetical proteins (T508DRAFT_01024, T508DRAFT_02014, T508DRAFT_02218), predicted Zn-dependent proteases (T508DRAFT_00450) and tRNA(Ile)-lysidine synthetase (T508DRAFT_00983).

5.2.9. Differential gene expression in *R. flavefaciens* FD1

There were 378 genes differentially expressed in *R. flavefaciens* FD1 (Figure 5.23D., Table A.5.5.), of which 375 were differentially expressed between Trt2 and Trt4 or between Trt3 and Trt4, but only 3 genes were differentially expressed between Trt2 and Trt3 (Table 5.9.). As FD1 is not included in Trt1, there is no Trt1 data. Large numbers of genes were identified as differentially expressed due to the generally low number of transcripts (1% or less) mapped to *R. flavefaciens* FD1 genome in Trt4 (Table 5.5.), which led to a high number of genes with no detectable expression in Trt4. A few genes in Trt4 still have comparable expression, including a glycoside hydrolase family 43 (FD1_0743), xylanase (FD1_0753), type II secretory system (FD1_0908 – FD1_0914), cytochrome b/b6 (FD1_1061), hypothetical proteins (FD1_0919, FD1_1202, FD1_1472, FD1_2347, FD1_2848, FD1_2992, FD1_3943), aspartate—ammonia ligase (FD1_1546), excisionase (FD1_1976), resolvase (FD1_1991), type 2 lantibiotic biosynthesis protein LanM (FD1_2204, 2205), L-threonine ammonia-lyase (FD1_2927), ribosome biogenesis GTP-binding protein YsxC (FD1_3379), antitoxin HicB (FD1_3635), RNA polymerase sigma factor (FD1_3938).

Table 5.9. Genes differentially expressed between Trt2 and Trt3

Locus_tag	Product	Expression (RPK0.1M)		
		Trt2	Trt3	Trt4
FD1_0077	hypothetical protein	2461	91	282
FD1_2086	histidine kinase-DNA gyrase B-and HSP90-like ATPase	8	274	81
FD1_3634	serine/threonine protein kinase	200	3846	270

In order to assess the effect of H_2 level on differential gene expression without being biased by the low transcript levels in Trt4, the analysis was carried out between Trt2 and Trt3 only, which identified 464 genes differentially expressed with FDR < 0.05, fold change > 2 (Locus_tag** in Table A.5.5.), but with a KW threshold of $p > 0.05$. Of these genes, 382 had higher expression in Trt2, and 82 genes had higher expression in Trt3 (Table A.5.5.). A large number of genes were highly expressed in Trt2, including carbohydrate utilisation (including cellulosome components), central carbon metabolism, nitrogen metabolism, electron transfer,

transporters, cell surface proteins and secretion systems, cell replication and house-keeping, cofactor biosynthesis, and amino acid biosynthesis. Most noteworthy were the highly expressed type II secretion system and type IV conjugative transfer system genes (Table A.5.6.). A smaller range of genes were up-regulated in Trt3, including a region encoding lantibiotic biosynthesis (FD1_2200 – FD1_2209) along with a number of hypothetical proteins (FD1_2191 – FD1_2197) which appear to be part of this lantibiotic biosynthesis region (Table A.5.6.). Other genes up-regulated in Trt3 encoded carbohydrate utilisation, cell surface proteins, a toxin/antitoxin system, transporters, secreted proteins, a DNA restriction/modification system and RNA polymerases.

5.2.10. Associations between the ISO4-H5 and *R. flavefaciens* FD1, *S. dextrinosolvens* H5 transcriptomes

The associations between the ISO4-H5 and *R. flavefaciens* FD1 transcriptomes were explored by Regularised Canonical Correlation analysis (rCCA), which deals with the high dimensionality and low number of replicates of transcriptomics data by adding a regularization term on the diagonal of covariance matrices, but requires the number of dimensions to be tuned.

The *R. flavefaciens* FD1 gene expression variable shared coordinates with approximately one third of the ISO4-H5 gene expression variable, as seen by the overlap of variable coordinates to the left of the correlation circle plot by the first two dimensions (Figure 5.31A.). This signifies that FD1 gene expressions have positive correlations with approximately one third of ISO4-H5 gene expressions, while negatively associated with the ISO4-H5 genes on the opposite side of the correlation circle plot. Based on the association between gene expression variables, the gene expressions in both genomes are negatively associated between Trt2 and Trt3 in dimension 2 (Figure 5.31B.). The *S. dextrinosolvens* H5 gene expression variable shared coordinates with approximately two thirds of the ISO4-H5 gene expression variable, as seen by the overlap of coordinates in the middle and to the left of the correlation circle plot (Figure 5.31C.). The overlap of gene coordinates to the left signifies a portion of genes in ISO4-H5 and *S. dextrinosolvens* H5 are positively associated, while the genes coordinates overlapped in the middle signifies these genes are neither positively or negatively associated with each other. Based on the association between gene expression variables, the gene expressions in both genomes are positively associated between Trt2 and Trt3 (Figure 5.31D.).

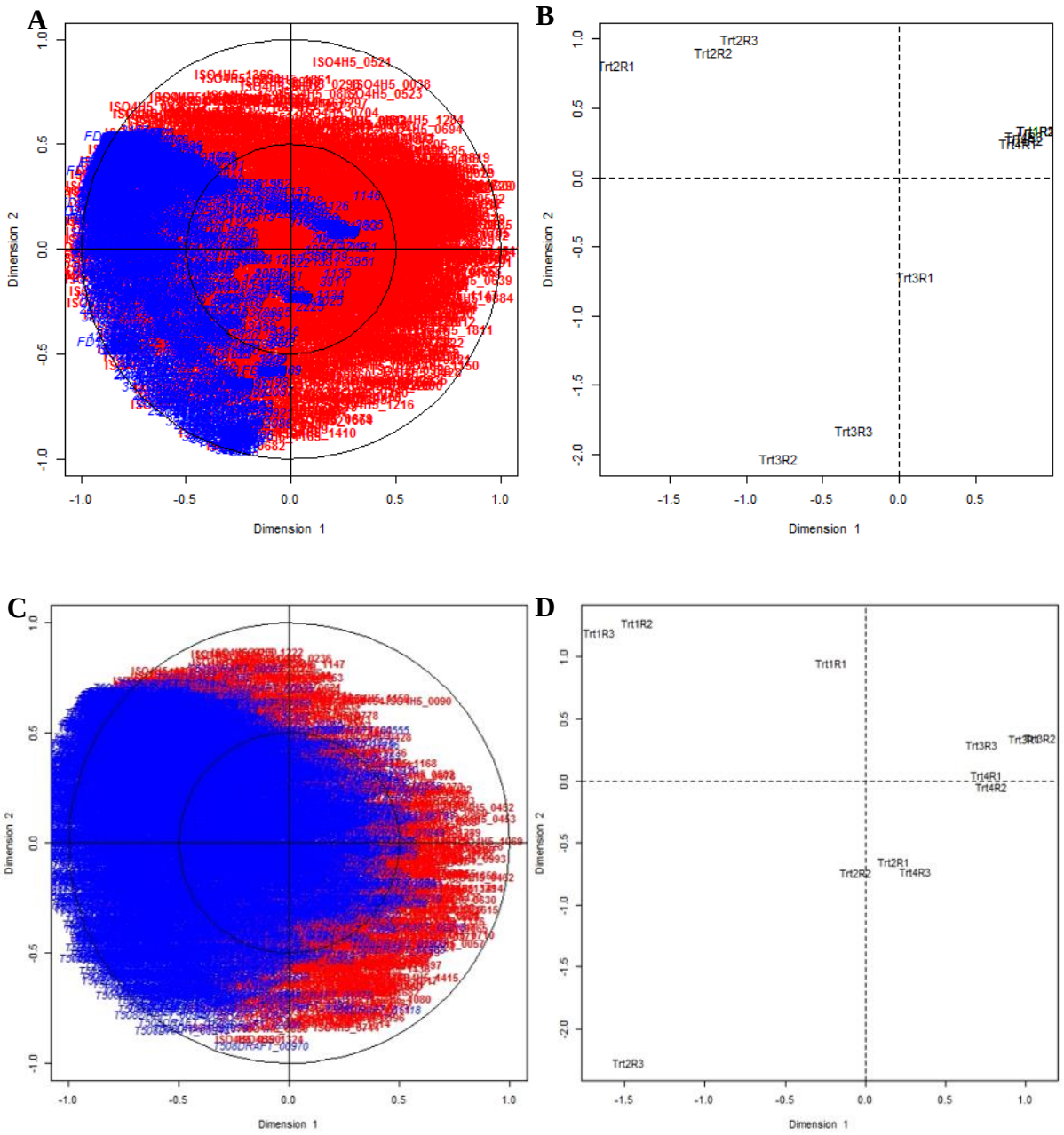
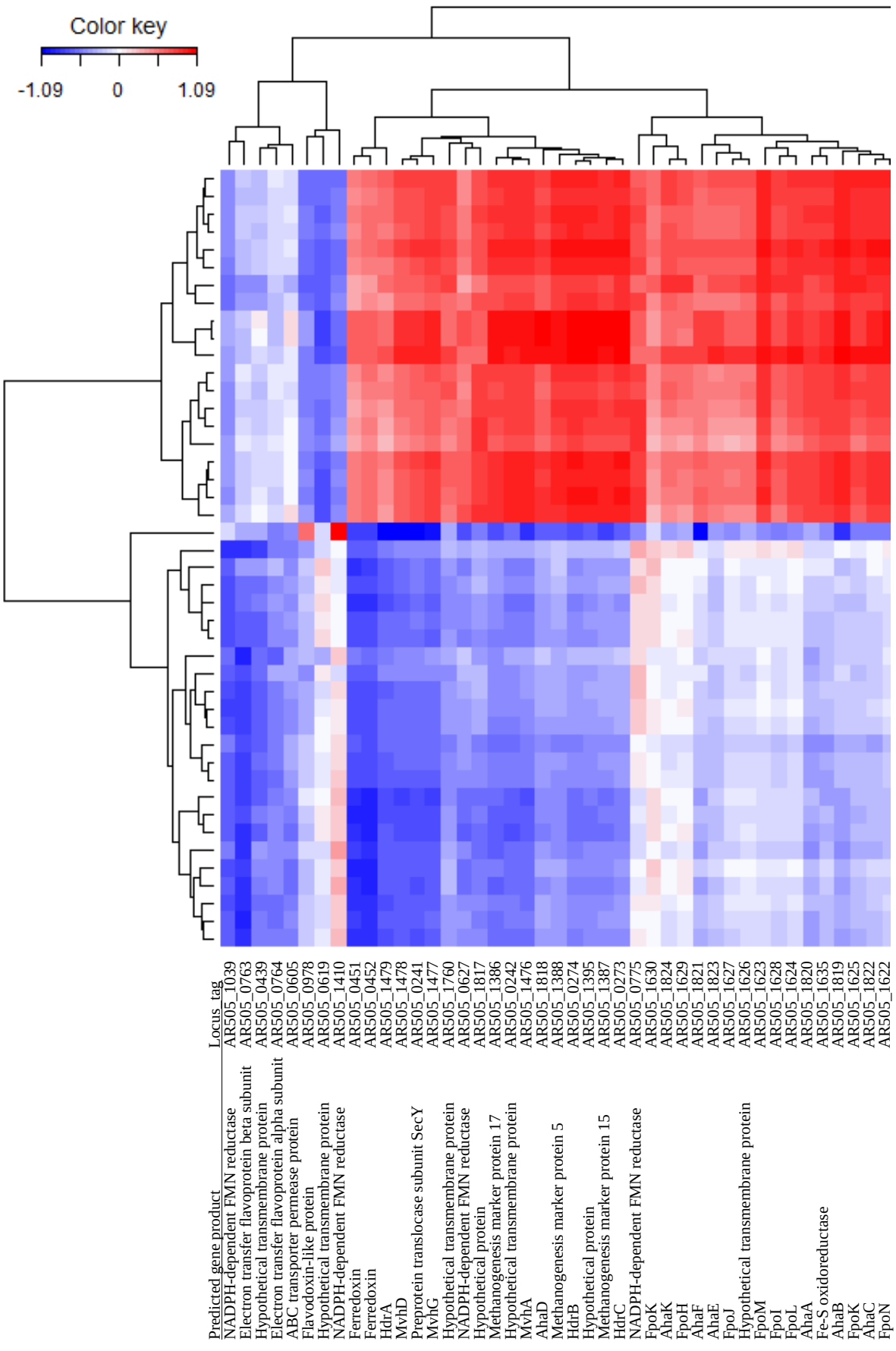
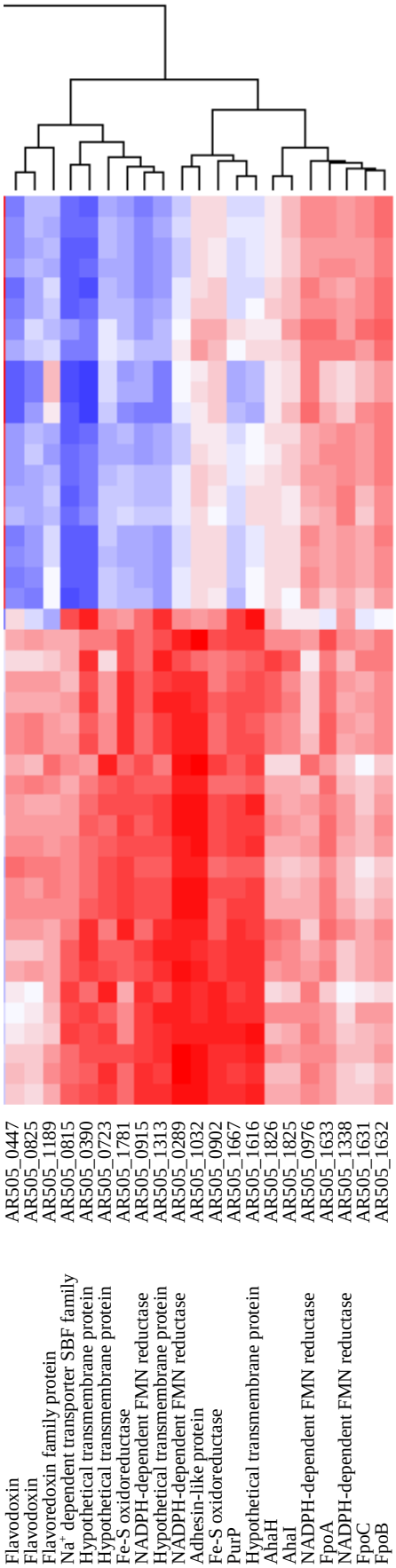


Figure 5.31. Regularised Canonical Correlation analysis of associations between ISO4-H5 and *R. flavefaciens* FD1, *S. dextrinosolvens* H5 transcriptomes. **A**. Correlation circle plot of ISO4-H5 and *R. flavefaciens* FD1 transcriptomes. Each gene vector represents variables, coordinate of vector represents correlation between variable and their associated component. ISO4-H5 genes are shown in red, *R. flavefaciens* FD1 genes are shown in blue. **B**. Correlation plot of treatments between ISO4-H5 and *R. flavefaciens* FD1. **C**. Correlation circle plot of ISO4-H5 and *S. dextrinosolvens* H5 transcriptomes. Each gene vector represents variables, coordinate of vector represents correlation between variable and their associated component. ISO4-H5 genes are shown in red, *S. dextrinosolvens* H5 genes are shown in blue. **D**. Correlation plot of treatments between ISO4-H5 and *S. dextrinosolvens* H5.


Color key
-1.09
0
1.09
Predicted gene product
Locus_tag
NADPH-dependent FMN reductase AR505_1039
Electron transfer flavoprotein beta subunit AR505_0763
Hypothetical transmembrane protein AR505_0439
Electron transfer flavoprotein alpha subunit AR505_0764
ABC transporter permease protein AR505_0605
Flavodoxin-like protein AR505_0978
Hypothetical transmembrane protein AR505_0619
NADPH-dependent FMN reductase AR505_1410
Ferredoxin AR505_0451
Ferredoxin AR505_0452
HdrA AR505_1479
MvhD AR505_1478
Preprotein translocase subunit SecY AR505_0241
MvhG AR505_1477
Hypothetical transmembrane protein AR505_1760
NADPH-dependent FMN reductase AR505_0627
Hypothetical protein AR505_1817
Methanogenesis marker protein 17 AR505_1386
Hypothetical transmembrane protein AR505_0242
MvhA AR505_1476
AhaD AR505_1818
Methanogenesis marker protein 5 AR505_1388
HdrB AR505_0274
Hypothetical protein AR505_1395
Methanogenesis marker protein 15 AR505_1387
HdrC AR505_0273
NADPH-dependent FMN reductase AR505_0775
FpoK AR505_1630
AhaK AR505_1824
FpoH AR505_1629
AhaF AR505_1821
AhaE AR505_1823
FpoJ AR505_1627
Hypothetical transmembrane protein AR505_1626
FpoM AR505_1623
FpoI AR505_1628
FpoL AR505_1624
AhaA AR505_1820
Fe-S oxidoreductase AR505_1635
AhaB AR505_1819
FpoK AR505_1625
AhaC AR505_1822
FpoN AR505_1622



Locus tag Predicted gene product

FD1_2209 lantibiotic protection ABC transporter, ATP-binding subunit
FD1_2207 ABC transporter substrate-binding protein
FD1_2208 ABC-2 type transport system permease
FD1_2203 NHLM bacteriocin system ABC transporter, peptidase
FD1_2191 Hypothetical protein
FD1_2202 Hypothetical protein
FD1_3901 Adhesin-like protein
FD1_2192 Hypothetical protein
FD1_2199 Lantibiotic lacticin
FD1_2195 Hypothetical protein
FD1_2201 Lantibiotic lacticin
FD1_2206 Hypothetical transmembrane protein
FD1_2198 Lantibiotic lacticin
FD1_2194 Hypothetical protein
FD1_2197 Hypothetical protein
FD1_2196 Hypothetical protein
FD1_2205 Type 2 lantibiotic biosynthesis protein LanM
FD1_2204 Type 2 lantibiotic biosynthesis protein LanM
FD1_2193 Hypothetical protein
FD1_2200 Lantibiotic lacticin
FD1_1061 cytochrome b/b6
FD1_0914 Hypothetical protein
FD1_3815 Dockerin
FD1_2667 Flavodoxin
FD1_0586 Flavodoxin
FD1_0909 Type IV-A pilus assembly ATPase PilB
FD1_1469 Rubrerythrin
FD1_0912 Prepilin-type cleavage/methylation protein
FD1_0908 Twitching motility protein PilT
FD1_0915 Type IV pilus assembly protein PilM
FD1_0916 Type II secretion protein F
FD1_0913 Adhesin-like protein
FD1_1719 Preprotein translocase subunit SecY
FD1_4004 Electron transport complex, RnfABCDGE type, B subunit
FD1_0911 Prepilin-type cleavage/methylation protein
FD1_3093 Sodium:proton antiporter
FD1_4109 Type IV conjugative transfer system coupling protein TraD
FD1_1689 Conjugal transfer protein TraX
FD1_4095 Conuugal transfer protein
FD1_4094 Type-IV secretion system protein TraC
FD1_3748 Sodium transporter
FD1_0910 Type II secretory pathway prepilin signal peptidase PulO
FD1_0844 Ferredoxin
FD1_4096 Conjugal transfer protein TrbL

Figure 5.32. Cluster image map of subsets of ISO4-H5 genes and *R. flavefaciens* FD1 genes involved in electron transfer, membrane potential generation and energy conservation. The image shows a graphical representation of associations between genes, and the colour in the heatmap indicates the strength or weakness of the correlation between subsets of gene expression, as illustrated by the key. The genes in both axes are ordered by hierarchical clustering, where each dendrogram illustrates the proximity between correlated variables.

The finding that the expression of the ISO4-H5 A_1A_0 ATP synthase and Fpo-like complex were negatively correlated with the presence of *R. flavefaciens* FD1 in the co-culture (Table A.5.3., Trt1 vs Trt2) was unexpected. Both the A_1A_0 ATP synthase and Fpo-like complex are transmembrane enzyme complexes, involved in electron transfer, membrane potential generation and energy conservation, therefore it was hypothesized that these functions in ISO4-H5 were influenced by the presence of *R. flavefaciens* FD1 in the co-culture. To explore this putative association, the ISO4-H5 and FD1 genes involved in electron transfer, membrane potential generation and energy conservation, that were differentially expressed between Trt2 and Trt3, were analysed using rCCA (using thresholds of $\lambda_1 = 0.01099$, $\lambda_2 = 0.11089$). *R. flavefaciens* FD1 genes also included transmembrane proteins involved in electron transfer or secretion, including Type II secretion system, Type IV conjugal transfer, lantibiotic biosynthesis and electron transfer. The gene associations gathered from rCCA are summarised in a clustered image map (Figure 5.32.). The genes encoding a putative lantibiotic biosynthetic region in FD1 displayed strong positive associations with seven subunits of the ISO4-H5 A_1A_0 ATP synthase, two Fdxs, *hdrABC, mvhAGD* genes, as well as genes of eight subunits of the Fpo-like complex. FD1 genes encoding cytochrome b/b6, Type IV conjugal transfer system and Type II secretion system displayed strong positive relationships with three hypothetical transmembrane proteins, one adhesin-like protein and two NADPH-dependent FMN reductase genes from ISO4-H5, and negative associations to the two Fdxs.

5.3. Discussion

Methanomassiliicoccales are a recently established order of methanogens, and currently includes four members identified from human and termite gut sources (Borrel *et al.* 2012; Gorlas *et al.* 2012; Borrel *et al.* 2013; Lang *et al.* 2015). Only *Methanomassiliicoccus luminyensis* B10 has been isolated as a pure culture (Dridi *et al.* 2012). In this thesis, the methanogenic archaeon ISO4-H5 enriched from the ovine rumen has been isolated as a pure culture by exploiting the metabolism of *S. dextrinosolvens* H5 that was co-enriched with ISO4-H5. The purification of ISO4-H5 allowed confirmation that it is a small coccus, without fluorescence at 420 nm. Studies on human faeces and the termite gut also identified Methanomassiliicoccales as coccus-shaped archaea, however, controversy remains in the literature regarding their autofluoresence at 420 nm. Autofluorescence has been reported in the human-sourced *Methanomassiliicoccus luminyensis* (Dridi *et al.* 2012), while an absence of autofluorescence was reported in *"Candidatus* Methanoplasma termitum" from the termite gut

(Lang *et al.* 2015). As discussed in Chapter 4, the analysis of available Methanomassiliicoccales genomes did not support their use of, or dependence on, F_{420}.

The *S. dextrinosolvens* bacterium observed in the original ISO4-H5 enrichment, was also encountered in enrichments of two other members of Methanomassiliicoccales from the ovine rumen, ISO4-G1 and ISO4-G11. All three strains of *S. dextrinosolvens* co-enriched with rumen Methanomassiliicoccales, displayed 99% 16*S* rRNA sequence identity to *Succinivibrio dextrinosolvens* strain 0554 (NR026476) (Jeyanathan 2010). These observations suggest that *S. dextrinosolvens* may play a role in the supply of metabolites that support the growth of rumen Methanomassiliicoccales. A co-culture has also been observed between an anaerobic rumen fungus and Methanomassiliicoccales (Jin *et al.* 2014) suggesting that associations between Methanomassiliicoccales and other rumen microorganisms are common.

S. dextrinosolvens was described originally as a curved rod that fermented glucose, D-xylose, L-arabinose, maltose, galactose, fructose, sucrose, mannitol, dextrin and pectin, it produced succinate, formate, acetate and lactate as principal fermentation products (Bryant and Small, 1956). The medium used for ISO4-H5 enrichment contained approximately 38 mM acetate (Saleem *et al.* 2013), and 60 mM of formate. The additional supplementation of 10 mM succinate did not improve the growth of ISO4-H5 ruling out succinate as the stimulatory factor enhancing ISO4-H5 growth (Figure 5.21.). Pectin addition to the ISO4-H5 enrichment culture stimulated CH_4 production significantly (P =5.6E-05), whereas the stimulation by glucose supplementation was less significant (P =1.4E-02) (Figure 5.1.).

This led to the conclusion that pectin-grown *S. dextrinosolvens* produced a metabolite(s) stimulatory to ISO4-H5 growth, and the stimulation of growth using SSPGMS indicated that this metabolite(s) could be supplied without the need of the *S. dextrinosolvens* organism itself. The addition of the SSPGMS combined with ampicillin enabled isolation of ISO4-H5 from the enrichment culture. The purified culture however, did not grow as well as the enrichment culture, thus the SSPGMS supplementation could not fully replace *S. dextrinosolvens* (Figure 5.4.). It is possible that certain stimulatory metabolites in the SSPGMS are unstable and degrade quickly once removed from *S. dextrinosolvens*, or possibly that physical interaction between ISO4-H5 and *S. dextrinosolvens* H5 is required to effect full stimulation of growth. However, physical contact between these two organisms has never been observed during microscopy of ISO4-H5 enrichment cultures.

The growth of *S. dextrinosolvens* H5 in various carbohydrate sources including pectin was investigated, due to the turbidity from pectin addition, the pH of culture was also monitored in addition to OD_{600}. *S. dextrinosolvens* does not grow well with pectin and rhamnose (Figure 5.5A., B.). The *S. dextrinosolvens* H5 culture can however, be continuously maintained by the pectin (data not shown). The growth of ISO4-H5 was stimulated by the supplementation of SSP/R/G/AGMS, but is most strongly stimulated by the SSP/RGMS (Figure 5.5C.), it is interesting to find ISO4-H5 was stimulated more from the growth medium of *S. dextrinosolvens* H5 that grew poorly.

In order to investigate the metabolites within SSPGMS that aided ISO4-H5 growth, NMR was carried out to identify compounds consumed during growth. Three compounds were identified as being significantly depleted after ISO4-H5 growth; methanol, ethanol and nicotinamide (Figures 5.8., 5.10., 5.11., 5.12.). The methanol was likely used as a methyl substrate for methanogenesis, and was not limiting ISO4-H5 growth as it was supplied in the medium in excess of the amount used. Ethanol is likely utilized via an alcohol dehydrogenase (AR505_0483), as described in Chapter 3, Section 3.2.4. Ethanol has been observed as a methyl source for methanogenesis in *Methanofollis ethanolicus* (Imachi *et al.* 2009), however, ISO4-H5 was unable to utilise ethanol alone as a methyl source (Figure 5.15A.). Ethanol is also known to stimulate CH_4 formation in *M. ruminantium* $M1^T$ (Leahy *et al.* 2010), presumably via a coenzyme F_{420}-dependent NADP reductase and NADP-dependent alcohol dehydrogenase (Berk and Thauer 1997). However, there is no evidence to support the utilisation of F_{420} in ISO4-H5, and no F_{420}-dependent enzymes are predicted in ISO4-H5 genome, therefore ethanol is unlikely to supply reducing potential to methanogenesis in ISO4-H5 via this mechanism. Growth tests showed that ISO4-H5 could not utilise ethanol as a source of reducing potential in the absence of H_2 (Figure 5.17.). Testing of 10 mM ethanol in the presence of H_2 did not influence the CH_4 production (Figure 5.18.), which further supports the conclusion that ISO4-H5 does not utilise ethanol to supply reducing potential to methanogenesis. The alcohol dehydrogenase found in ISO4-H5 (AR505_0483) shares homology with the multifunctional alcohol dehydrogenase from *Clostridium acetobutylicum*, which can produce butanol (Fischer *et al.* 1993). Therefore, butanol was tested as a substrate, but it did not support CH_4 formation in ISO4-H5 (Figure 5.15A.). An aldehyde dehydrogenase (AR505_1599) was also predicted in ISO4-H5 genome, and together with an alcohol dehydrogenase could potentially produce acetyl-CoA from ethanol, and utilise ethanol as a carbon source. However, as indicated above, the addition of ethanol did not stimulate the growth rate or H_2 utilisation of ISO4-H5 (Figure

5.18.). The further assessment of alternative carbon sources was impractical as the medium that supports ISO4-H5 growth contains 30% rumen fluid, which provides 38 mM acetate (Saleem *et al.* 2013). Based on the genes present in the genome, acetate is likely the main carbon source of ISO4-H5. The assessment of alternative carbon sources in ISO4-H5 depends on the development a defined medium in which acetate can be replaced.

Nicotinamide was found to be present in SSPGMS at a level barely detectable by NMR, and it is likely utilised for NAD biosynthesis. However, ISO4-H5 possesses all the genes necessary to encode the production of NAD from aspartate or nicotinic acid (Chapter 3, Table 3.27). Nicotinic acid (0.1 mg/L) was also present in the cultivation medium as part of the vitamin mix, yet ISO4-H5 consumed additional nicotinamide from the SSPGMS. The growth of ISO4-H5 was not stimulated by additional nicotinamide or nicotinic acid, and does not appear to be limited by a lack of nicotinic acid (Figure 5.13.). The use of exogenously supplied nicotinamide is likely to be preferable compared to *de novo* biosynthesis of NAD from aspartate, as it conserves energy, but it is not essential to the survival of ISO4-H5 and does not stimulate its growth.

The growth rate of ISO4-H5 culture *in vitro* is slow and it does not produce measurable turbidity at OD_{600}. This is unlikely to be the case *in vivo*, otherwise ISO4-H5 would not survive being washed out of rumen due to the dilution effects of ingested feed and water. Therefore its *in vitro* growth is likely limited by some nutrient or growth factor. The amount of methanol and ethanol used during growth was quantified, and the amounts involved are not growth rate-limiting. It appears that while methanol, ethanol and nicotinamide supplied from SSPGMS may contribute to ISO4-H5 growth, the critical metabolite(s) from *S. dextrinosolvens* that stimulates growth is likely below the detection limit of NMR (~30 µM). If that is the case, then the critical metabolite probably functions as a cofactor for ISO4-H5 rather than a substrate. Fractionation of SSPGMS by HPLC and testing the ability of individual fractions, or combination of fractions, to stimulate ISO4-H5 growth, is the logical next step in identifying the metabolite(s) involved. While the identification of the limiting metabolite would be useful, the main objective here was obtaining a pure culture of ISO4-H5 for subsequent experimentation, which was achieved.

The purification of ISO4-H5 allowed the inspection of its morphology as well as determining the substrates that support methanogenesis in this organism. The methanogen cell envelope has been found to be SDS- and protease-resistant (Kandler and Konig 1978), it consists of

pseudomurein layer covered encrusted by S-layer, a two-dimensional protein crystal (Kandler and Konig 1978; Ellen *et al.* 2010). The pseudomurein is distinct to the bacterial peptidoglycan, as it lacks muramic acid and possesses a different molar ratio of lysine:alanine:glutamate (Kandler and Konig 1978). The S-layer consisted of mostly glycoproteins containing *N*-linked glycans (Jarrell *et al.* 2010), in *Methanococcus voltae*, a 76 kDa protein in hexagonal lattice arrangement forms the S-layer (Koval and Jarrell 1987). The electron micrographies of *M. luminyensis* B10 did not show a prominent cell-wall-like structure (Dridi *et al.* 2012), furthermore, the cells were found to be SDS-sensitive (Dridi *et al.* 2012), suggesting the Methanomassiliicoccales possess a different cell envelope structure to other methanogens. ISO4-H5 appeared to have no cell wall and no observable extracellular matrix or capsule structure (Figure 5.3A.). Instead, a bi-layer membrane was observed, with a thinner membrane (approximately 5 nm) compared to that previously observed in *M. luminyensis* B10 (approximately 100 nm, (Dridi *et al.* 2012)). ISO4-H5 also appears to be smaller (300 - 600 nm diameter) compared to *M. luminyensis* B10 (approximately 1200 nm diameter, (Dridi *et al.* 2012)). The absence of a cell wall could render cells susceptible to osmotic stress, whether the size difference observed between B10 and ISO4-H5 is due to difference in osmolarity requires further investigation. The lack of observed pseudomurein and S-layer can be reflected from genomic analysis in Chapter 3. Genomic study suggested synthesis of activated mannose and cell wall associated *N*-glycans are potentially possible (Borrel *et al.* 2014). As discussed in Chapter 3 and 4, the cell envelope may contain cardiolipin and glycosylated exopolysaccharide.

The ISO4-H5 genome sequence suggested ISO4-H5 can utilize methanol, mono-, di-, and tri-methylamines as well as certain forms of methylated thiols. Mono-, di-, and tri-methylamine (Figure 5.15.), as well as methyl-3-methylthiopropionate (Figure 5.16.) were confirmed as substrates for methanogenesis in ISO4-H5, in addition to methanol. The CH_4 production in ISO4-H5 cultures grown on trimethylamine was no different to that seen from growth on dimethylamine and methylamine, confirming that the methanogenic substrate is not the limiting condition for ISO4-H5 growth *in vitro*. Interestingly, ISO4-H5 was unable to grow on dimethylsulfide or methylmercaptopropionate, both were predicted as possible substrates based on the strong homology of its *mtsA* (AR505_1066) gene to *Methanosarcina barkeri mtsA* gene (71.6% aa identity) (Tallant and Krzycki 1997). The chemical difference between methylmercaptopropionate and methyl-3-methylthiopropionate is that the methyl-3-methylthiopropionate has a methyl group attached to the sulphur atom, which might make it more available for use by ISO4-H5. Interestingly, the dimethylsulfide has two methyl groups

attached to the central sulfur atom, but the even charge distribution over the molecule may make it difficult for ISO4-H5 MtsA to attack. Betaine and choline can be broken down to trimethylamine, and methanogens have been observed to utilise betaine and choline directly (Muller *et al.* 1981; Fiebig and Gottschalk 1983; Wargo 2013; L'Haridon *et al.* 2014; Watkins *et al.* 2014). ISO4-H5 was unable to utilise choline and betaine (Figure 5.19.), which suggests ISO4-H5 is dependent on other organisms within the rumen to supply methylamines, similar to what was found previously in *Methanosarcina barkeri* (Hippe *et al.* 1979).

The co-culture experiment between the ISO4-H5 enrichment culture and *R. flavefaciens* FD1 was carried out to identify changes in gene expression when ISO4-H5 was subjected to low or high H_2 conditions, and when growing on methylamine compared to methanol. *R. flavefaciens* FD1 was able to produce H_2 in excess of that required by ISO4-H5 (Figure 5.21C.). The ISO4-H5 appears to have consumed more H_2 under the high H_2 conditions (~1.2-1.5 µmole, Figure 5.21B.) compared to the low H_2 condition. However, CH_4 production in the low H_2 treatments was not much lower than that observed under high H_2. Assuming a 1:1 stoichiometry of H_2 use to CH_4 formation as predicted for ISO4-H5 growth on methanol, the FD1-containing co-cultures under low H_2 produced ~1 µmole of H_2 , despite H_2 use appearing to be ~0.2-0.3 µmole from the headspace analysis. Microbes that ferment carbohydrate and produce H_2 may suffer feedback inhibition in high H_2 environment (De Vrije and Claassen 2003). Similarly, *R. flavefaciens* FD1 may produce different levels of H_2 in high H_2 and low H_2 environment due to feedback inhibition. The production of CH_4 by ISO4-H5 appeared to occur sooner under the low H_2 condition (Figure 5.21A.), which suggested there could be changes in gene expression that allows ISO4-H5 to scavenge H_2 earlier. The ISO4-H5 grown with the high H_2 condition produced more CH_4 at 120-h than the low H_2 condition (Figure 5.21A.), which suggests ISO4-H5 can achieve higher growth yield with a high H_2 condition. However, CH_4 production was observed earlier in both groups that grew under low H_2 condition, which suggests ISO4-H5 has higher growth rate in low H_2 condition. The presence of *R. flavefaciens* FD1 also supplies molecular hydrogen in the media in addition to the H_2 gas present; the molecular hydrogen may be utilised more readily by the ISO4-H5 than the H_2 gas and thus contributed to the higher growth rate. Within the rumen, a higher growth rate is beneficial for a micro-organism, as it allows the organism to grow more quickly and prevents it being washed out of the rumen. These observations suggest that ISO4-H5 associates with H_2 producing microorganisms to maintain a high growth rate, and therefore persist in the rumen.

Transcriptomic studies requires the accurate interpretation of enormous amount of data, and multivariate statistical analysis is helpful in prioritizing differential gene expressions, and in filtering important and genuine relationships from the experimental noise and dependencies of the biological samples. The multivariate analysis of the gene expression from the co-culture studies showed clear differences between treatments in ISO4-H5. The most notable changes in gene expression between the high H_2 Trt2 and the low H_2 Trt3 were the expression of *hdr* and *mvh* that encode enzymes involved in the supply of reducing potential to methanogenesis using H_2 (Figure 5.25.). The HdrA protein subunit harbors flavin adenine dinucleotide (FAD) necessary for reduction of CoM-S-S-CoB (Kaster *et al.* 2011), and both HdrA, HdrC also contain iron-sulfur cluster binding motifs. The HdrB subunit has 10 highly conserved cysteine motifs important in re-oxidation of CoM-S-S-CoB, while HdrD contains characteristics of both HdrB and HdrC, having the iron sulphur cluster binding motifs as well as two conserved cysteine motifs (Hedderich *et al.* 2005). The *hdrA*, *hdrB1*, *hdrC* and *mvhAGD* genes were up-regulated under low H_2 condition which suggests that ISO4-H5 up-regulates the expression of the corresponding Hdr and Mvh enzymes to scavenge additional H_2 under low H_2 conditions. The *hdrB1* and *hdrC* genes were identified as a conserved operon in all Methanomassiliicoccales genomes analysed (Table A.4.15.), while *hdrB2* is not conserved, and display a significantly lower score in BLASTp, Pfam and COG analyses. This suggests *hdrB2* may not encode a genuine HdrB, but rather an iron-sulfur motif containing protein with unknown function. Conversely, the *hdrD* gene was not differentially expressed under different H_2 condition, which suggests HdrD does not form complex with HdrABC in ISO4-H5, or its function does not require up-regulation, it has been proposed to be coupled with the Fpo-like complex and function similarly to the absent FpoF and FpoO subunit (Lang *et al.* 2015), as discussed in Chapter 4.

The genes encoding MrtA, a hypothetical protein AR505_1395 downstream of the *mrtA* gene and three methanogenesis marker proteins Mmp7, 15 and 17 were found to be differentially expressed between Trt2 and Trt3. The up-regulation of the *mrtA* gene appears to have contributed to ISO4-H5 producing CH_4 earlier under the low H_2 conditions. The Mrt complex has been structurally characterised in *Methanosarcina barkeri* as a heterohexamer of αα'ββ'γγ' (Grabarse *et al.* 2000), therefore it is unexpected that similar up-regulation was not observed from the *mrtB* and *mrtG* genes. Similarly, *mrtA* specific down-regulation has also been observed previously in cows fed with rapeseed oil-supplemented diet, while *mrtG* expression was unaffected (Poulsen *et al.* 2013). The up-regulation of AR505_1395 suggests this

hypothetical protein is inside the Mrt operon and regulated similarly to the *mrtA*, which signifies this protein could be associated with the MrtA. A recent rumen metagenome and metatranscriptome study has classified the *mcr/mrt* operons into three major groups (Shi *et al.* 2014), and Methanomassiliicoccales *mrt* operon is grouped differently to the *mcr* and *mrt* operons of *Methanobrevibacter* (Shi *et al.* 2014). In *Methanobacterium thermoautotrophicum* the Mcr is predominant under H_2 limiting conditions, and Mrt is predominant when H_2 is not growth limiting (Bonacker *et al*. 1993; Morgan *et al*. 1997). The observed up-regulation of ISO4-H5 *mrtA* gene under low H_2 condition contradicts previous findings. The classification, combined with the differential expression of *mrtA* and AR505_1395 genes, suggests the Methanomassiliicoccales Mrt complex could behave differently to the Mrt complexes previously studied, and deserve further investigations in both its structure and function.

The genes encoding methanol utilization, and the *fdx1* (AR505_0452) and *fdx2* (AR505_0453) were not differentially expressed under differing H_2 conditions according to the statistical cut off, however the patterns of gene expression displayed by *mtaC2, mtaB2* and *fdx* genes AR505_0452 and AR505_0453 appeared to be strongly influenced by H_2 level. Most gene expression is controlled by multiple factors, for example *mrtA* (Figure 5.29.), was up-regulated by low H_2 concentration but it was not up-regulated when methylamine is present. The *mtaB2* and *mtaC2* appear to be influenced by low H_2 condition, while the *mtaB1* and *mtaC1* appear to be unaffected by H_2 level, and seem to respond negatively to the presence of *R. flavefaciens* FD1. In *Methanosaeta* spp. which lack the *fpoF* gene of the Fpo-like complex, it has been proposed that Fdx can act as an electon carrier (Welte and Deppenmeier 2011). Similarly, all the Methanomassiliicoccales genomes analysed lack the *fpoF* gene (Chapter 4), and may also utilise Fdx for electron transport. The expression pattern of *fdx* genes AR505_0452 and AR505_0453 most closely match those of *mtaB2, mtaC2, mrtA* (Figure 5.33.), *hdrABC* and *mvhAGD* genes (Figure 5.26.), and this similarity of expression pattern suggests these Fdxs could be associated with Mvh, Hdr and may be directly involved in electron transfer for methanogenesis.

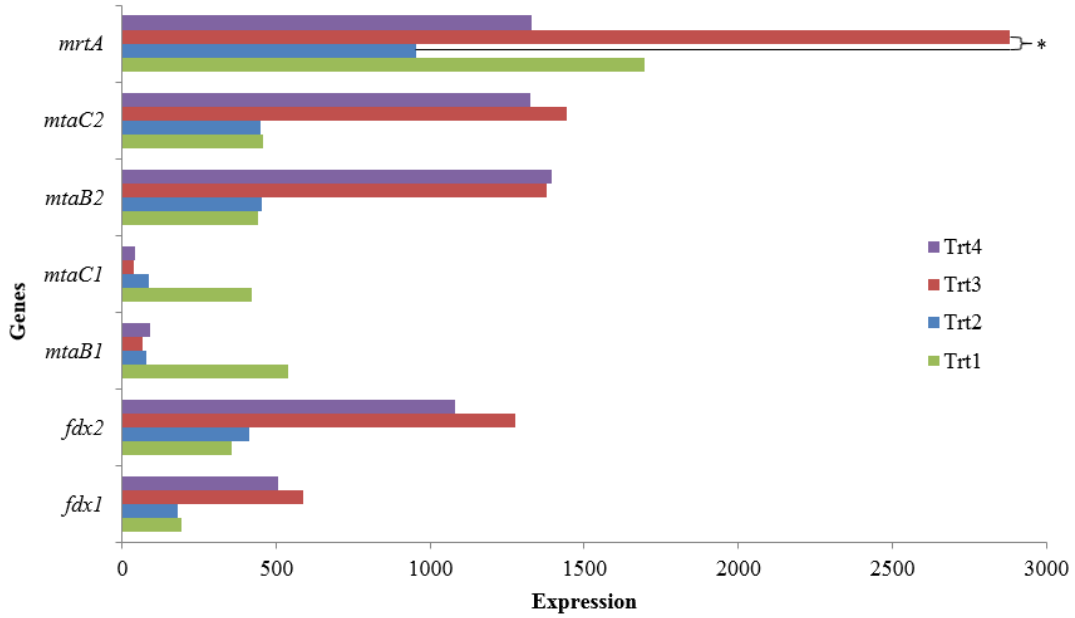


Figure 5.33. Expression of genes involved in methanol utilizing methanogenesis between high and low H_2 condition. Methanol:corrinoid methyltransferase (*mtaB*), methanol corrinoid protein (*mtaC*), methyl-CoM reductase complex (*mrt*), ferredoxin (*fdx*).

The importance of expression pattern in the co-culture study can again be seen when assessing methylamine-influenced gene expressions. Ideally, Trt3 and Trt4 should be compared as the methyl source is the only variable, however, comparison between Trt3 and Trt4 showed only 13 genes were differentially expressed, and none of the genes were directly involved in methyl substrate utilisation. When overall gene expression was analysed in a correspondence analysis, the strongest associations were linked to Trt1 and Trt4 (Figure 5.24.). This is probably due to these two conditions being the most different, in terms of substrate type and H_2 level, and when genes are regulated by multiple factors, the differences between Trt1 and Trt4 would be greater than differences between Trt3 and Trt4. The genes proposed to be involved in non-methanol methyl substrate utilisation and their gene expressions were summarised (Figure 5.26.) and the *mtmB2*, *mtmC1*, *mtbC*, *mtbB*, *mtbP1*, *mttP*, AR505_1068 and *mtsA* genes were identified as being differentially expressed, either between Trt1 and Trt4 or between Trt2 and Trt4. All of these genes show a pattern of high gene expression in Trt4, even *mtsB* gene which was not identified as differentially expressed. Although the trimethylamine utilising *mttBC* genes were not differentially expressed it appears to be highly expressed across all treatments. The up-regulation of methyl-3-methylthiopropionate utilising genes *mtsAB* is unexpected, as there is no methyl-3-methylthiopropionate supplementation in this experiment. This suggests that in the rumen the availability of methylamines for ISO4-H5 is linked to the availability of methyl-

3-methylthiopropionate, and possibly other utilisable methylated thiol compounds, and falls under the same regulation mechanism. The elevated gene expression of AR505_1068 encoding a 4Fe-4S Fdx iron-sulfur binding domain containing protein downstream of the *mtsB* gene showed it is potentially associated with MtsB; experimental validation is required to define its function.

Although Trt4 was intended to test methylamine alone as a methyl source for methanogenesis, in hindsight, approximately 2 mM of methanol would have been carried over in the inoculum in this experiment. Therefore Trt4 contained both methylamine (10 mM) and methanol (~2 mM) as methyl sources. Despite this, the differences in gene expressions between Trt3 and Trt4 were still obvious in the MDS, PCoA and group dispersion plot (Figure 5.22.) analyses, which signifies that ISO4-H5 gene expression responded to the presence of methylamine despite the presence of methanol.

In order to identify genes that are influenced by the presence of methylamine and methanol, genes with two-fold or more expression difference between Trt4 and all other treatments were selected amongst the differentially expressed genes (Table 5.6.). This identified some genes that could be highly important. This included the AR505_1559 gene encoding an adhesin-like protein, which was approximately 10 fold higher expressed in Trt4 compared to all three other treatments, and it may be involved in adherence to a methylamine/methyl-3-methylthiopropionate producing bacteria within the rumen. The hypothetical protein AR505_0750 adjacent to the *mttP* gene also showed at least 10 fold higher expression in Trt4 compared to other treatments, and may be associated with trimethylamine utilisation. Based on the high up-regulation of gene expression involved in methylamine utilisation in the presence of methanol and methylamine, methylamines could be the preferred substrate of ISO4-H5. The growth of ISO4-H5 on different methyl substrates appeared to be similar, however these growth tests were conducted under sub-optimal conditions as the growth limiting component of the medium remained unidentified. Genes encoding eight subunits of the Fpo-like complex have been identified as differentially expressed with two to seven fold lower expression in Trt4. This finding is unexpected, as theoretically the methylamines and methyl-3-methylthiopropionate shares methanogenesis genes with methanol.

Although the multivariate analysis showed clear differences between treatments in ISO4-H5, the differences were less significant in *S. dextrinosolvens* H5 (Figure 5.23.), and smaller number of genes are differentially expressed under the conditions tested (Figure 5.22D.). On

the contrary, large number of genes were found to be differentially expressed in *R. flavefaciens* FD1, although all but three are due to comparison to Trt4 (Figure 5.23D.), which is likely due to the relatively low number of transcripts aligned to *R. flavefaciens* FD1 (Table 5.5B.). Therefore the differential expressed genes identified in this manner between Trt4 and other treatments have questionable reliability.

R. flavefaciens FD1 was used to create a low H_2 environment in the co-culture with ISO4-H5. However, the *R. flavefaciens* FD1 in Trt4 has an abnormally low transcript abundance in comparison to Trt3, the low transcript abundance may not necessarily mean the lack of growth in *R. flavefaciens* FD1, as it was still able to supply sufficient quantity of H_2 to support the growth of ISO4-H5 (Figure 5.21C.). The only variable between Trt3 and Trt4 is the presence of monomethylamine. Methylamine is considered a permeable weak base (Ohkuma and Poole 1978; Ritchie and Gibson 1987), at 10 mM it inhibited *R. flavefaciens* by approximately 20% (Ricke *et al.* 1994), this may explain the abnormal transcript abundance observed in Trt4. In a rumen perspective, this suggests the presence of methylamine-consuming methylotrophic methanogens would benefit *R. flavefaciens* FD1-like organisms by uptake of methylamine. This may also explain why methylamine was predicted as the conserved substrate across all the Methanomassiliicoccales genomes analysed, as growth on methylamine would keep methylamine levels low in the rumen and encourage the growth of *R. flavefaciens*-like organisms and ensure a steady source of H_2 for growth of Methanomassiliicoccales.

The presence of *R. flavefaciens* FD1 seems to influence ISO4-H5 gene expression in an unexpected way. The most significant finding was that the ISO4-H5 A_1A_0 ATP synthase complex was highly expressed in Trt1 under high H_2 in the absence of FD1 (Table 5.7B, Figure 5.28B). The proton translocating F_1F_0 ATP synthase of *Lactobacillus acidophilus* has been observed to be up-regulated at low pH (Kullen and Klaenhammer 1999). However, no difference in pH was detected between each of the treatments, therefore the change of expression in ATP synthase could probably be associated with *R. flavefaciens* FD1. The associations between transcriptomes of ISO4-H5 and *R. flavefaciens* FD1 (Figure 5.29.) showed positive correlation between six A_1A_0 ATP synthase subunits and Fpo-like complex of ISO4-H5 with a region in *R. flavefaciens* FD1 encoding a putative lantibiotic biosynthetic island (Figure 5.32.). Lantibiotics are antimicrobial peptides with post-translational modification involving lanthionine or β-methyllanthionine (O'Sullivan *et al.* 2011). There are two types of lantibiotics, Type 1 and Type 2 (Begley *et al.* 2009). Type 1 lantibiotics are modified by LanBC, (e.g. nisin), and are involved in pore formation in cell membranes via

binding to a cell wall precursor, lipid II (Hsu *et al.* 2004). Type 2 lantibiotics are modified by LanM (Begley *et al.* 2009 7882), (e.g. lacticin 3147 which is in fact two peptides LtnA1 and LtnA2 that act synergistically), and also bind to lipid II and induce pore formation, resulting in ion leakage (Wiedemann *et al.* 2006). The lantibiotic biosynthetic island of *R. flavefaciens* FD1 genome is predicted to contain two LanM homologues FlvM1 and FlvM2, which suggests it produces Type 2 lantibiotics. Another species of the genus *Ruminococcus*, *R. gnavus* E1, could produce a Type 2A lantibiotic called ruminococcin A, which acts against bacteria phylogenetically related to *R. gnavus*, and *Clostridium* spp. (Dabard *et al.* 2001). A recent study expressed the FD1 lantibiotic operon in *E. coli* and verified that FD1 indeed produces Type 2 lantibiotics, including four structurally conserved lipid II binding peptides (Flvα.a-d) and eight structurally diverse β-peptides (Flvβ.a-g), which resembles the two-component lantibiotic system such as lacticin 3147, in particular, FlvM1 modified Flvα.a and Flvβ.g displayed synergistic antimicrobial effect against *Micrococcus luteus* DSM 1790 (Zhao and van der Donk 2016). The FD1 did not produce detectable level of lantibiotics when co-cultured with *Ruminococcus albus* 7 (Zhao and van der Donk 2016), which suggested that Flvαs and Flvβs may behave differently to ruminococcin A, it may not target phylogenetically related bacteria, or there are additional triggers of expression. This means that ISO4-H5 may be affected by the lantibiotics, which is reflected by the up-regulation of Fpo-like complex and ATP synthase, which suggests the proton gradient of ISO4-H5 was affected. This study offers no conclusive evidence on this matter, as the cell envelope structure of ISO4-H5 is unknown.

Negative correlations were also identified between the *R. flavefaciens* FD1 cytochrome b/b6, type II secretion and type IV conjugative transfer systems genes and the ISO4-H5 A_1A_0 ATP synthase subunits, Mvh, Hdr and Fdxs up-regulated under the low H_2 condition. *Geobacter sulfurreducens* has been reported as being able to transfer electrons to extracellular electron acceptors via pili (Reguera *et al.* 2005), and direct interspecies electron transfer has been observed between *Geobacter metallireducens* and two methanogens, *Methanosaeta harundinacea* (Rotaru *et al.* 2014) and *Methanosarcina barkeri* (Rotaru *et al.* 2014). Furthermore, co-cultures with these methanogens could not be established with a pilin-deficient *G. metallireducens* strain (Rotaru *et al.* 2014), which suggests the electron transfer via electrical conducting pili is an important form of syntrophy. Additionally, intercellular "wiring" between sulphate-reducing bacteria and anaerobic methanotrophic archaea was found to be essential for the anaerobic oxidation of CH_4 (Wegener *et al.* 2015). In the co-culture of *G. metallireducens* and *G. sulfurreducens*, the up-regulation of PilA, pilin domain 2 protein

and pilus-associated cytochrome OmcS were considered to be important in the direct interspecies electron transfer (Shrestha *et al.* 2013). The cytochrome b/b6 gene and genes involved in Type II secretion and Type IV conjugative transfer were up-regulated in *R. flavefaciens* FD1 during the high H_2 condition, which suggests *R. flavefaciens* FD1 may interact in a similar fashion with ISO4-H5, transferring electrons via pili and/or cytochromes. Up-regulation of these genes in FD1 may facilitate electron transport to ISO4-H5, disposing of reducing equivalents and alleviating the feedback inhibition caused by H_2 under high H_2 conditions. The ISO4-H5 A_1A_0 ATP synthase and Fpo complex genes were down-regulated in Trt2 in the presence of *R. flavefaciens* FD1, which suggests ISO4-H5 may utilise electrons transferred from *R. flavefaciens* FD1 to drive proton efflux and ATP production. The up-regulation of the Fpo complex and A_1A_0 ATP synthase in ISO4-H5 in the absence of *R. flavefaciens* FD1 may reflect its need to create its own membrane potential/proton gradient to drive ATP synthesis.

The majority of changes in the transcriptome of *S. dextrinosolvens* H5 was due to the presence of *R. flavefaciens* FD1. Most of the differentially expressed genes were up-regulated in absence of FD1, and are involved in carbohydrate utilisation, central carbon metabolism, protein biosynthesis and other house-keeping functions, which suggests that *S. dextrinosolvens* H5 may have to compete with FD1 for utilisable carbohydrate. Although the rCCA suggested there were gene associations between ISO4-H5 and *S. dextrinosolvens* H5, because all four treatments contain both ISO4-H5 and *S. dextrinosolvens* H5, it is difficult to identify the particular interaction between them. The logical next step would be to sequence the transcriptome of *S. dextrinosolvens* H5 monoculture and identify the difference in gene expression, which may give hints as to the interaction between *S. dextrinosolvens* H5 and ISO4-H5.

When comparing *R. flavefaciens* FD1 gene expression between Trt2 and Trt3, a large number of genes involved in cell replication and carbohydrate utilisation appeared to be highly expressed under the high H_2 condition (Table A.5.5.), which suggests FD1 grew better at high H_2 condition. This is supported by the generally higher number of transcripts assigned to *R. flavefaciens* FD1 in Trt2 (Table 5.5B.). However, this observation contradicts the dogma that high H_2 concentrations inhibit fibre-degrading rumen microbes, such as *R. flavefaciens* (Iannotti *et al.* 1973; Wolin 1976; Rees *et al.* 1995; Morvan *et al.* 1996). The interspecies H_2 transfer has been studied using FD1 and *Mbb. ruminantium* PS, when interspecies H_2 transfer happen between FD1 and methanogens, the molar yield of acetate in hexose fermentation

should double, and the succinate production that was used to regenerate NAD^+ in monoculture should be reduced (Wolin *et al.* 1997). In this study, the pyruvate flavodoxin oxidoreductase (FD1_0183) and acetate kinase (FD1_0693) genes involved in acetate and H_2 production were up-regulated in Trt2, as are six genes involved in glycolysis, phosphoglycerate kinase (FD1_4123), triosephosphate isomerase (FD1_0338), fructose-bisphosphate aldolase (FD1_0313), 6-phosphofructokinase (FD1_2874) and glucokinase (FD1_0252). The elevated expression of genes involved in glycolysis in Trt2 contradicts the dogma that high H_2 concentrations inhibit fermentation. The growth of ISO4-H5 in low H_2 condition is evidence for interspecies H_2 transfer, which is reflected in the expression of genes involved in acetate and H_2 production, however, the expression of these genes under high H_2 condition also suggests there could be an alternative method for FD1 to remove excess electrons, possibly by transferring the excess electrons to ISO4-H5. The succinate production pathway was used by FD1 to remove excess electrons in monoculture (Wolin *et al.* 1997), yet malate dehydrogenase (FD1_3067), fumarate hydratase (FD1_3371) and fumarate reductase (FD1_3362) involved in succinate production were up-regulated in Trt2 instead of down-regulated, which is inconsistent with the previous findings. It is possible that the electron transfer from *R. flavefaciens* FD1 to ISO4-H5 postulated above, may relieve the feedback inhibition caused by high H_2 condition, by dumping electrons from NADH and allowing NAD to be recycled. Only a few FD1 genes were up-regulated in low H_2 condition (Table 5.11B.), most notable being the genes within the putative lantibiotic biosynthesis island. Which means *R. flavefaciens* FD1 may gain a competitive advantage over other rumen bacteria due to the production of antimicrobial lantibiotics.

In conclusion, ISO4-H5 carries out H_2 dependent methylotrophic methanogenesis using methanol, mono-, di-, tri- methylamine and methyl-3-methylthiopropionate as substrate. Two sets of genes are involved in methanol utilisation, one set is up-regulated by low H_2 conditions. The methylamine permeases and the genes involved in utilising mono- and di-methylamine are up-regulated by presence of methylamine, as is the methyl-3-methylthiopropionate utilising operon *mtsAB* and AR505_1068 genes. H_2 level response has been observed in *mrtA, hdr* and *mvh* genes, two Fdxs have been identified as associated with Hdr and Mvh. ISO4-H5 also appears to consume ethanol, but it is not used as a methyl substrate or as a source of reducing potential for methanogenesis.

Chapter 6

The complete genome sequence of *Methanobrevibacter* sp. D5 and its comparison to other members of the genus *Methanobrevibacter*

6.1. Introduction

The genus *Methanobrevibacter* consists of obligate anaerobes that are able to reduce CO_2 to CH_4 using H_2 and sometimes formate as the electron donor (Miller 2001). *Methanobrevibacter* spp. cannot utilise acetate, methanol, methylamines or other organic compounds as electron donors for methanogenesis (Miller 2001). Ammonium ions are considered the major source of cell nitrogen and acetate is expected to be the major source of cell carbon (Miller 2001). Typically, additional nutrients such as casamino acid, yeast extract or B-complex vitamins are required for good growth (Miller and Lin 2002). Cells are usually short, lancet shaped cocci, short rods to long, filamentous rods that are non-sporulating and non-motile (Balch *et al.* 1979). Members of the *Methanobrevibacter* species contain pseudomurein cell walls which are composed of *N*-acetyl amino sugars, neutral sugars and peptide moieties containing L-glutamate, L-alanine, L-lysine, L-ornithine (Kandler and Konig 1978). To date, 15 *Methanobrevibacter* type strains has been identified, and their environments and nutritional requirements are summarised in Table 6.1.

Table 6.1. *Methanobrevibacter* species and their nutritional requirements

Species*	Environments	Substrates	Growth requirements	Reference
Methanobrevibacter smithii **PS^T, TS145A, TS145B, TS146A, TS146B, TS146C, TS146D, TS146E, TS147A, TS147B, TS147C, TS94B, TS94C, TS95A, TS95B, TS95C, TS95D, TS96A, TS96B, TS96C**, GMS-01, B-181, ALI, **F1**	Sewage, human colon, human faeces, rumen	CO_2, H_2, formate	Acetate, NH_4, B vitamins	Balch *et al.*,1979, Bryant *et al.*, 1971, Miller *et al.*, 1982
Methanobrevibacter ruminantium **M1^T**	Bovine rumen	CO_2, H_2, formate	Acetate, CoM, NH_4, aa, 2-methylbutyric acid, YE, TP	Smith and Hungate, 1958, Bryant *et al.*, 1971
Methanobrevibacter wolinii **SH^T**	Sheep faeces, rumen	CO_2, H_2	Acetate, YE, CoM, VFA	Miller and Lin, 2002
Methanobrevibacter thaueri CW^T	Cow faeces, rumen	CO_2, H_2	Acetate, TP, YE	Miller and Lin, 2002
Methanobrevibacter woesei GS^T	Goose faeces	CO_2, H_2, formate	Acetate, NaCl, TP, YE	Miller and Lin, 2002
Methanobrevibacter gottschalkii **HO^T,** PG	Horse faeces, pig faeces, rumen	CO_2, H_2	Acetate, NaCl, TP, YE	Miller and Lin, 2002
Methanobrevibacter oralis ZR^T, **JMR01**	Human subgingival plaque, human faeces	CO_2, H_2	Fecal extract, VFA	Ferrari *et al.*, 1994
Methanobrevibacter olleyae **KM1H5-1P^T, YLM1**	Ovine rumen	CO_2, H_2, formate	Acetate	Rea *et al.*, 2007
Methanobrevibacter millerae **ZA-10^T, SM9,** HW02, **YE315**	Bovine rumen	CO_2, H_2, formate	YE, TP	Rea *et al.*, 2007
Methanobrevibacter boviskoreani **JH1^T**, JH4, JH8, **AbM4**	Bovine rumen	CO_2, H_2, formate	YE, CoM, VFA	Lee *et al.*, 2013
Methanobrevibacter arboriphilus **DH1^T**, AZ, DC, **ANOR1,** A2, SH4, **SA**	Decaying cottonwood, paddy field soil, human faeces	CO_2, H_2, formate^v	RF, YE, TP, NaCl^v, B vitamins	Zeikus and Henning, 1975, Miller 2001
Methanobrevibacter cuticularis **RFM-1^T**	Termite gut	CO_2, H_2, formate	RF, YE, ca	Leadbetter *et al.*, 1996
Methanobrevibacter curvatus **RFM-2^T**	Termite gut	CO_2, H_2	RF, NB, YE, ca, LI, TerE	Leadbetter *et al.*, 1996
Methanobrevibacter filiformis **RFM-3^T**	Termite gut	CO_2, H_2	DTT, YE	Leadbetter *et al.*, 1998
Methanobrevibacter acididurans **ATM^T**	Anaerobic digester	CO_2, H_2	Acetate, RF	Savant *et al.*, 2002

*Type strain^T, other strains classified under genus *Methanobrevibacter* in the NCBI database. ^{v}, strain variable. coenzyme M (CoM), rumen fluid (RF), volatile fatty acid (VFA) yeast extract (YE), trypticase peptone (TP), dithiothreitol (DTT), casamino acids (ca), amino acids (aa), nutrient broth (NB), liver infusion (LI), termite extract (TerE). Strains with sequenced genomes are displayed in bold.

Members of the *Methanobrevibacter* genus identified from the rumen environment are classified phylogenetically according to 16*S* rRNA gene sequences into two major clades, the *Mbb. gottschalkii* clade and the *Mbb. ruminantium* clade. These clades account for 47% and 27% of rumen archaea, respectively, as reported in a global rumen microbial community survey based on 16*S* rRNA gene sequences (Henderson *et al.* 2015). In NZ rumen microbial populations, these two clades account for 42.4% and 32.9% of the rumen methanogen population (Seedorf *et al.* 2015). Thus far, 12 *Methanobrevibacter* species type strains and 27 other strains have been genome sequenced, the majority of these (21), reside within the *Mbb. smithii* clade (Table 6.1.). There are 18 sequenced genomes included in this study, consisting of 11 type strains, five rumen strains, and two additional non-rumen strains were included from the *Mbb. arboriphilus* and *Mbb. oralis* clades.

In order to expand our understanding of members of the rumen *Mbb. gottschalkii* clade, an isolate from this clade was obtained. A CH_4-forming enrichment culture designated H6 was obtained from the rumen of a sheep (Tag number #J472) grazing a pasture diet (G. Henderson, personal communication). From this H6 culture a member of the *Mbb. gottschalkii* clade was isolated and designated as *Methanobrevibacter* sp. D5 (Kim 2012). In this chapter the genome sequence of *Methanobrevibacter* sp. D5 was assembled, annotated, analysed and compared to other available *Methanobrevibacter* genomes.

6.2. Results

6.2.1. Cell morphology

Methanobrevibacter sp. D5 has a short rod morphology which are 0.5 to 1.2 µm in length and 0.4 to 0.7 µm in width and is often observed as chains (Figure 6.1.).

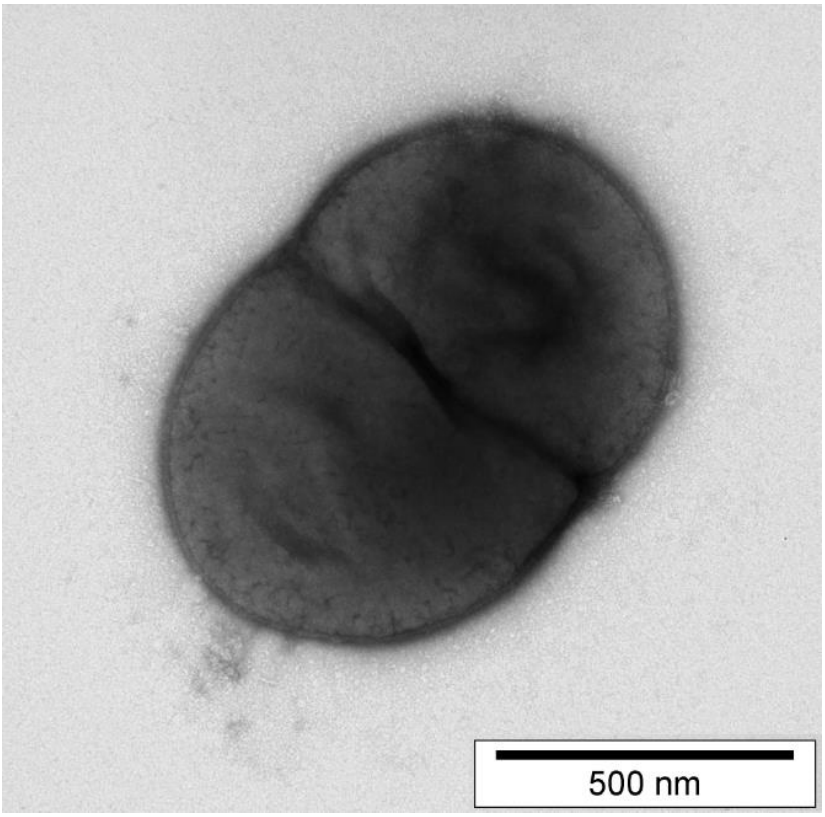


Figure 6.1. TEM of negatively stained thin sections of *Methanobrevibacter* sp. D5. The sample was prepared as previously described (Khan *et al.* 2013). Images were generated using a Philips CM10 Transmission Electron Microscope, and captured using an Olympus SIS Morada camera and SIS iTEM software (Germany).

6.2.2. Genome sequencing results and assembly

Genome assembly. Two *Methanobrevibacter* sp. D5 genome assemblies were performed during the course of this thesis; the first was performed using 454 pyrosequencing data from DNA isolated from the original H6 enrichment culture, and the second was performed using Illumina Miseq data, generated from DNA isolated from a pure culture of D5.

The sequencing of enrichment culture H6 yielded 1,407,413 reads with a total of 331,108,524 bp. The summary of the assembly statistics is displayed in Table 6.2. The assembly generated 383 scaffolds totaling 17,291,655 bp. Scaffolds belonging to the D5 genome were identified by sequence coverage, %G+C content and by the presence of ribosomal proteins. This analysis resulted in the identification of 4 scaffolds (scaffolds 1, 15, 21 and 53) containing 40 contigs

and totalling 2,756,038 bp (Figure 6.2.). In addition to the D5 genome, DNA from five putative bacterial contaminants was identified. The bacterial contaminants were identified by 16*S* rRNA gene sequence identity as *Pseudobutyrivibrio ruminis* strain 153 (99% identity), *Olsenella umbonata* strain lac31 (99% identity), *Mogibacterium pumilum* strain D2-18 (93% identity), *Lachnospiraceae* bacterium strain DJF (94% identity) and *Erysipelotrichaceae bacterium* strain NK3D112 (93% identity).

Table 6.2. Genome assembly summary of *Methanobrevibacter* sp. D5 from enrichment culture H6

Genome sequence feature	**Size or %G+C**
Enrichment culture H6	
%G+C content	40.47%
Total size	15,210,579 bp
Coverage	18.6×
Number of reads	1,407,413
Number of bases	331,108,524 bp
Number of contigs	6,150
Number of scaffolds	383
Number of contigs >500 bp	4,882
N50 contig size	50,105
Scaffold 1	
%G+C content	32.94%
Size	2,270,142 bp
Number of reads	302,839
Number of contigs	32
Scaffold 15	
%G+C content	32.15%
Size	271,401 bp
Number of reads	39,317
Number of contigs	5
Scaffold 21	
%G+C content	33.85%
Size	160,014 bp
Number of reads	24,052
Number of contigs	2
Scaffold 53	
%G+C content	36.82%
Size	54,481 bp
Number of reads	2,817
Number of contigs	1
***Methanobrevibacter* sp. D5 total from the enrichment culture**	
%G+C content	33.1%
Size	2,756,038 bp
Number of contigs	40
Coverage	31.5×

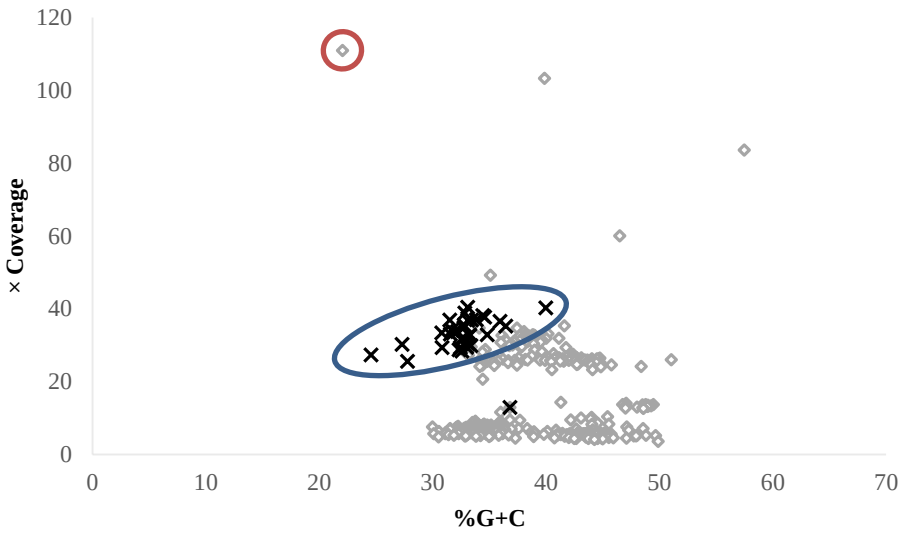


Figure 6.2. Contig coverage and %G+C of enrichment culture H6 sequencing assembly. (×) D5 contigs. (◊) Contigs of other organisms within the H6 enrichment culture. Only contigs above 3 kb are displayed. A putative prophage contig is circled in red. The cluster of D5 contigs are circled in blue.

The sequencing and assembly result of the *Methanobrevibacter* sp. D5 pure culture is summarised in Table 6.3. A hybrid assembly was produced by mapping the newly sequenced contigs onto the existing D5 assembly from the 454 assembly of the H6 enricment culture. The Miseq assembly results indicated the presence of a second genome that had a low coverage in comparison to the *Methanobrevibacter* sp. D5 contigs (Figure 6.3.). Based on the predicted ORFs and its BLASTP results, the contaminating genome likely belongs to a facultative anaerobe from the genus *Bacillus*. A D5 contig containing an rRNA operon has approximately double the coverage of other D5 contigs (Figure 6.3.).

Table 6.3. Genome assembly summary of *Methanobrevibacter* sp. D5

Features	**D5**	**Contaminating genome**
Total size	2,826,590 bp	3,741,504 bp
Coverage	567.5	8.7
Number of total reads	13,667,636	
Number of total contigs	8770	
Number of contigs >6000 bp	104	
Number of D5 contigs	33	
N50 contig size	64,460 bp	39,481 bp
G+C content	33.1	37.4

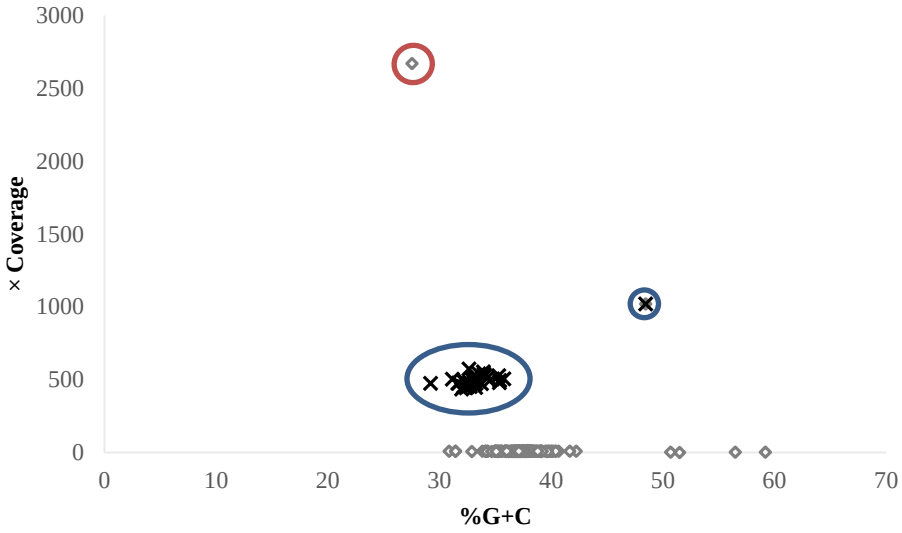


Figure 6.3. Contig coverage and %G+C of D5 pure culture sequencing assembly. (×) D5 contigs. (◊) Contigs of other organisms. Only contigs above 3 kb are displayed. A putative prophage contig is circled in red. The cluster of D5 contigs are circled in blue.

Methanobrevibacter sp. D5 gap closure. To complete the *Methanobrevibacter* sp. D5 genome as a circular chromosome, it was necessary to close the sequence gaps between contigs. The outline of the D5 genome gap closure procedure is illustrated in Figure 6.4.

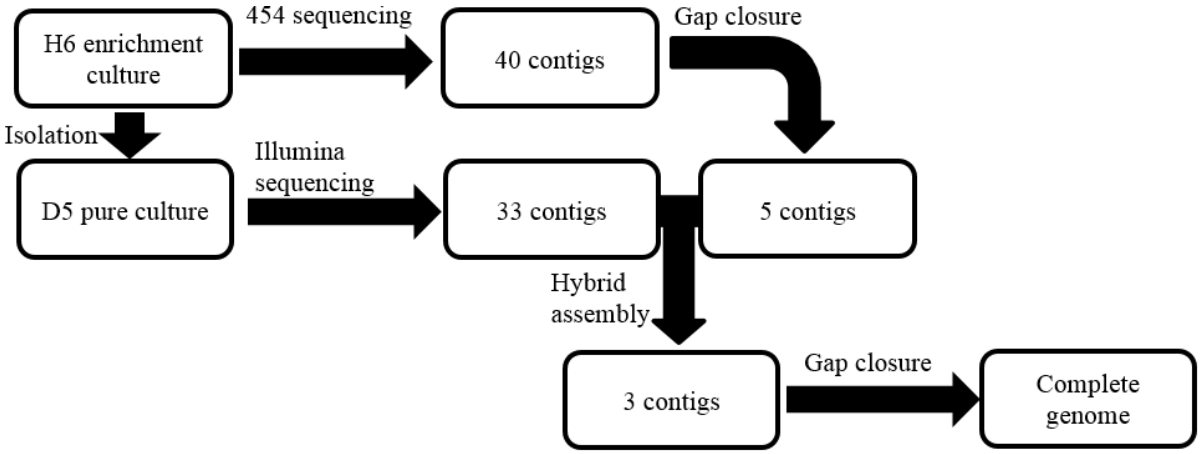


Figure 6.4. Flow chart of D5 gap closure procedure. Gap closure was performed on H6 assembly, and subsequently merged with Miseq assembly to create a hybrid assembly, further gap closure was performed on the hybrid assembly to close the genome.

The gap closure procedure is also described in Section 2.2.12. A total of 238 sequencing reactions were required to circularize the *Methanobrevibacter* sp. D5 genome as a single contig. Base conflicts in the hybrid assembly genome were validated by PCR amplification and re-sequencing.

Methanobrevibacter sp. D5 genome assembly verification. Assembly validation was confirmed by PFGE (Section 2.2.13). The restriction endonuclease *Mlu*I and *Apa*I were used (Figure 6.5.).

The *Mlu*I digestion pattern confirmed the assembly, whereas the *Apa*I digestion pattern did not. In the *Apa*I PFGE results, a 265 kb band appeared to be due to the absence of an *Apa*I restriction site at base 226,045, which caused the fused fragment of 221 kb and 44 kb. In order to assess this discrepancy, this region was validated by PCR amplification and re-sequenced and the presence of *Apa*I restriction endonuclease recognition site at base 226,045 was confirmed. Use of PFGE also revealed a 170 kb band which was observed in the digested and undigested lanes of both enzyme treatments in addition to the D5 genomic DNA.

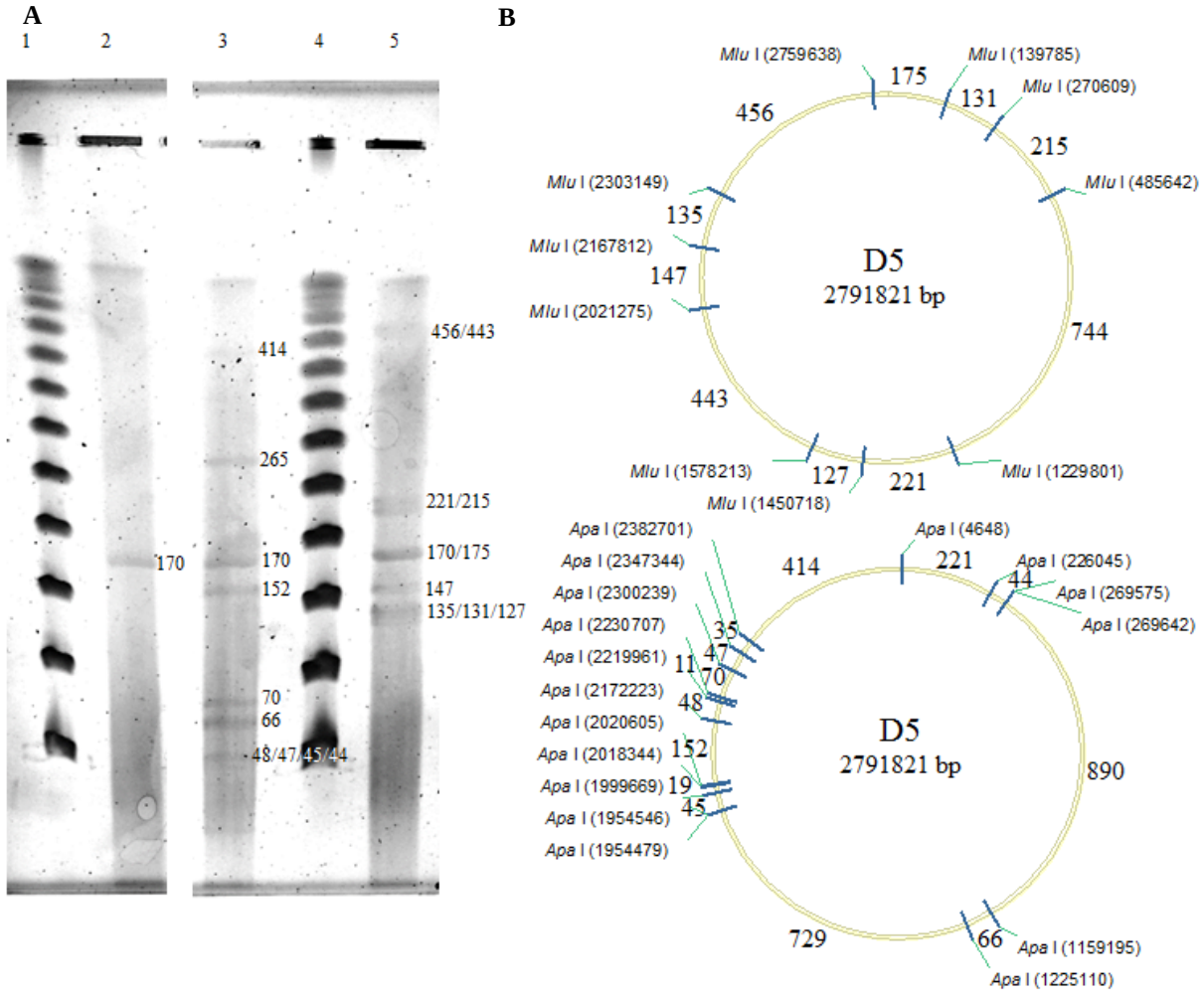


Figure 6.5. PFGE of *Methanobrevibacter* sp. D5 genomic DNA. **A.** *Methanobrevibacter* sp. D5 genomic DNA digested with restriction endonucleases. Numbers beside each band indicates size in kb. Lane 1: concatenated lambda marker (fragment sizes, from bottom in kb, are 48.5, 97, 145.5, 194, 242.5, 291, 339.5, 388 and 436.5). Lane 2: undigested genomic DNA; Lane 3: digested with *Apa*I; Lane 4: concatenated lambda marker; Lane 5: digested with *Mlu*I. **B.** In silico restriction enzyme maps of *Mlu*I and *Apa*I restriction endonuclease cleavage sites within the D5 genome. Numbers between restriction sites indicates fragment size in kb. Numbers in brackets indicates the base number of restriction sites.

6.2.3. Phylogenetic relationship of sequenced members of *Methanobrevibacter*.

The phylogenic relationships of the members of the genus *Methanobrevibacter* was inferred from their 16*S* rRNA gene sequences (Table 6.4., Figure 6.6.). A FGD tree (Figure 6.7.) and a pan-genome tree (Figure 6.8.) were created to assess the similarity between sequenced genomes based on the full complement of ORFs within each genome.

Table 6.4. Phylogenetic divergence of *Methanobrevibacter* spp. indicated by the % identity of 16*S* rRNA gene.

	ATM^T	**ANOR1**	**JH1^T**	**RFM-2^T**	**RFM-1^T**	**RFM-3^T**	**HO^T**	**SM9**	**Za-10^T**	**KM1H5-1P^T**	**JMR01**	**M1^T**	**PS^T**	**AbM4**	**D5**	**YE315**	**YLM1**	**SH^T**
ATM^T		94.9	92.7	93.9	94.1	93.7	92	92.7	93	93.8	93.5	93.9	94	92.8	92.9	92.6	93.9	92.4
ANOR1	94.9		93.2	94.8	96.4	94.7	92.6	93.6	94.2	94.4	93.7	93.1	94	92.5	93.4	93.6	94.4	92.6
JH1^T	92.7	93.2		91.6	91.7	92	91.7	92.3	92.3	93.2	92.4	93.2	93.4	**99.2**	92.6	92.6	93.4	96.3
RFM-2^T	93.9	94.8	91.6		94.9	94.1	90.8	92.1	92.4	92.9	91.7	92.9	92.3	91	91.8	91.9	92.9	91.3
RFM-1^T	94.1	96.4	91.7	94.9		94.3	91.8	93.3	93.2	93.4	93.1	93.5	93.4	91.7	92.9	93.1	93.6	91.6
RFM-3^T	93.7	94.7	92	94.1	94.3		90.9	92.2	92.2	92.5	91.9	93.1	92.3	91.9	92.3	92.4	92.7	91.6
HO^T	92	92.6	91.7	90.8	91.8	90.9		97.4	97.2	92.8	96.7	92.5	97	91.5	97.9	97.9	92.9	92.3
SM9	92.7	93.6	92.3	92.1	93.3	92.2	97.4		**99.1**	93.2	97.4	92.8	98	91.9	**98.7**	98.5	93.5	92.2
Za-10^T	93	94.2	92.3	92.4	93.2	92.2	97.2	**99.1**		93.2	97.2	93.1	97.9	92.4	**98.8**	**98.7**	93.4	92.5
KM1H5-1P^T	93.8	94.4	93.2	92.9	93.4	92.5	92.8	93.2	93.2		93.7	98.1	93.7	93	93.3	93.1	**99.9**	93.7
JMR01	93.5	93.7	92.4	91.7	93.1	91.9	96.7	97.4	97.2	93.7		92.9	98.2	92.2	97.4	96.9	93.6	92.2
M1^T	93.9	93.1	93.2	92.9	93.5	93.1	92.5	92.8	93.1	98.1	92.9		93.3	92.6	93	93.1	97.7	93.4
PS^T	94	94	93.4	92.3	93.4	92.3	97	98	97.9	93.7	98.2	93.3		93.2	98	97.4	93.7	93.5
AbM4	92.8	92.5	**99.2**	91	91.7	91.9	91.5	91.9	92.4	93	92.2	92.6	93.2		92.2	92.2	92.8	95.6
D5	92.9	93.4	92.6	91.8	92.9	92.3	97.9	**98.7**	**98.8**	93.3	97.4	93	98	92.2		**99**	93.6	92.8
YE315	92.6	93.6	92.6	91.9	93.1	92.4	97.9	98.5	**98.7**	93.1	96.9	93.1	97.4	92.2	**99**		93.5	92.5
YLM1	93.9	94.4	93.4	92.9	93.6	92.7	92.9	93.5	93.4	**99.9**	93.6	97.7	93.7	92.8	93.6	93.5		93.8
SH^T	92.4	92.6	96.3	91.3	91.6	91.6	92.3	92.2	92.5	93.7	92.2	93.4	93.5	95.6	92.8	92.5	93.8	

Identities ≥ 98.7% are displayed in bold. All positions containing gaps or missing data were eliminated. Type strains are indicated by superscript T.

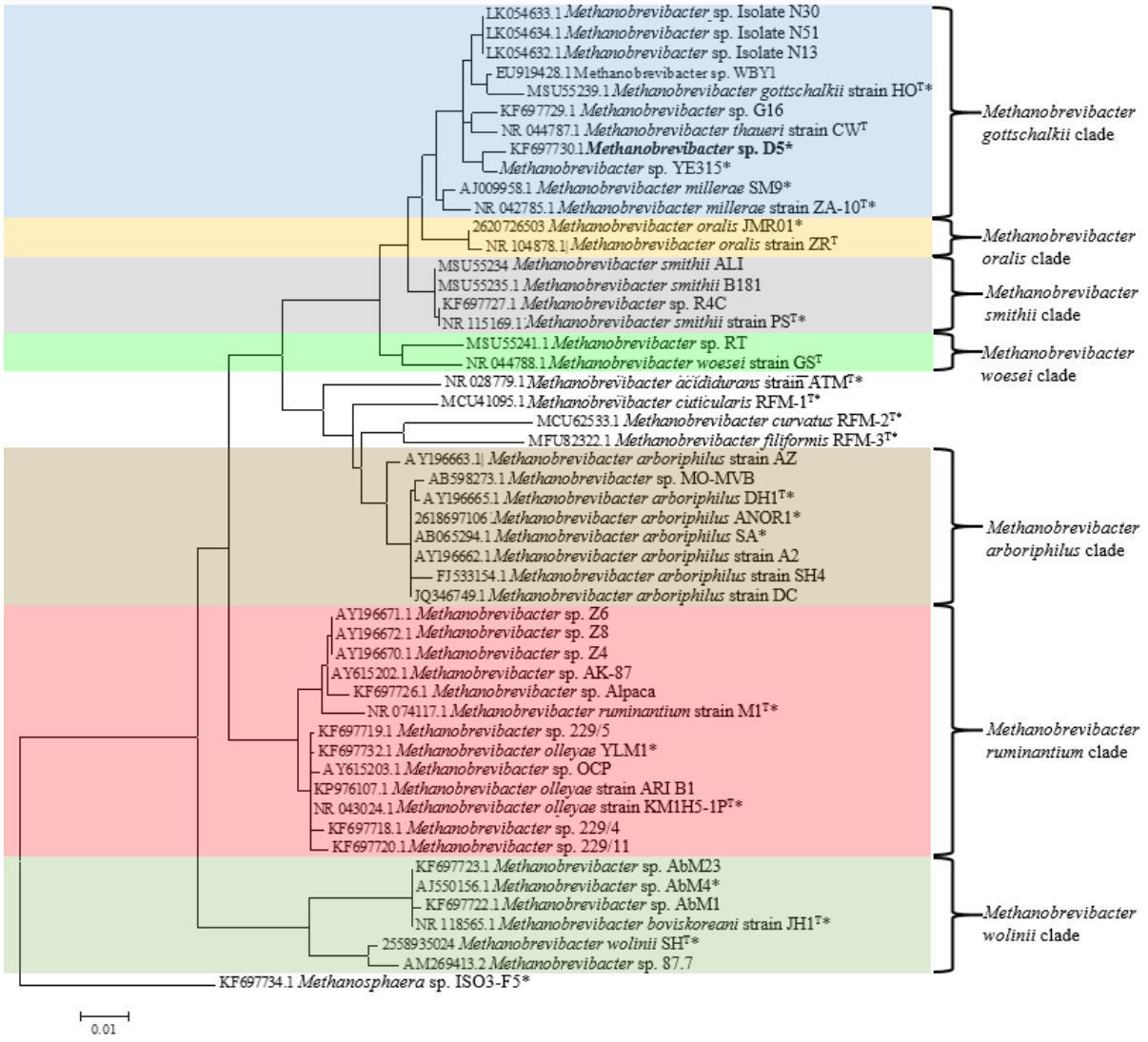


Figure 6.6. 16*S* rRNA gene-based phylogenetic tree. The phylogenetic tree showing the relationships of cultured *Methanobrevibacter* spp. in relation to *Methanobrevibacter* sp. D5 (shown in bold print). The phylogeny was inferred from 16*S* rRNA gene nucleotide sequences (1,474 bp internal region) aligned using the Maximum Likelihood method based on the Kimura 2-parameter model. Bar: 0.01 substitutions per nucleotide position. The GenBank/JGI accession numbers of sequences are displayed. Strains whose genomes have been sequenced are marked with asterisk. The initial tree for the heuristic search was obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using the MCL approach. All positions containing gaps or missing data were eliminated, giving a total of 667 positions in the final dataset. The 16*S* rRNA gene sequence from *Methanosphaera* sp. ISO3-F5 was used as an outgroup. Type strains are indicated by superscript ^{T}.

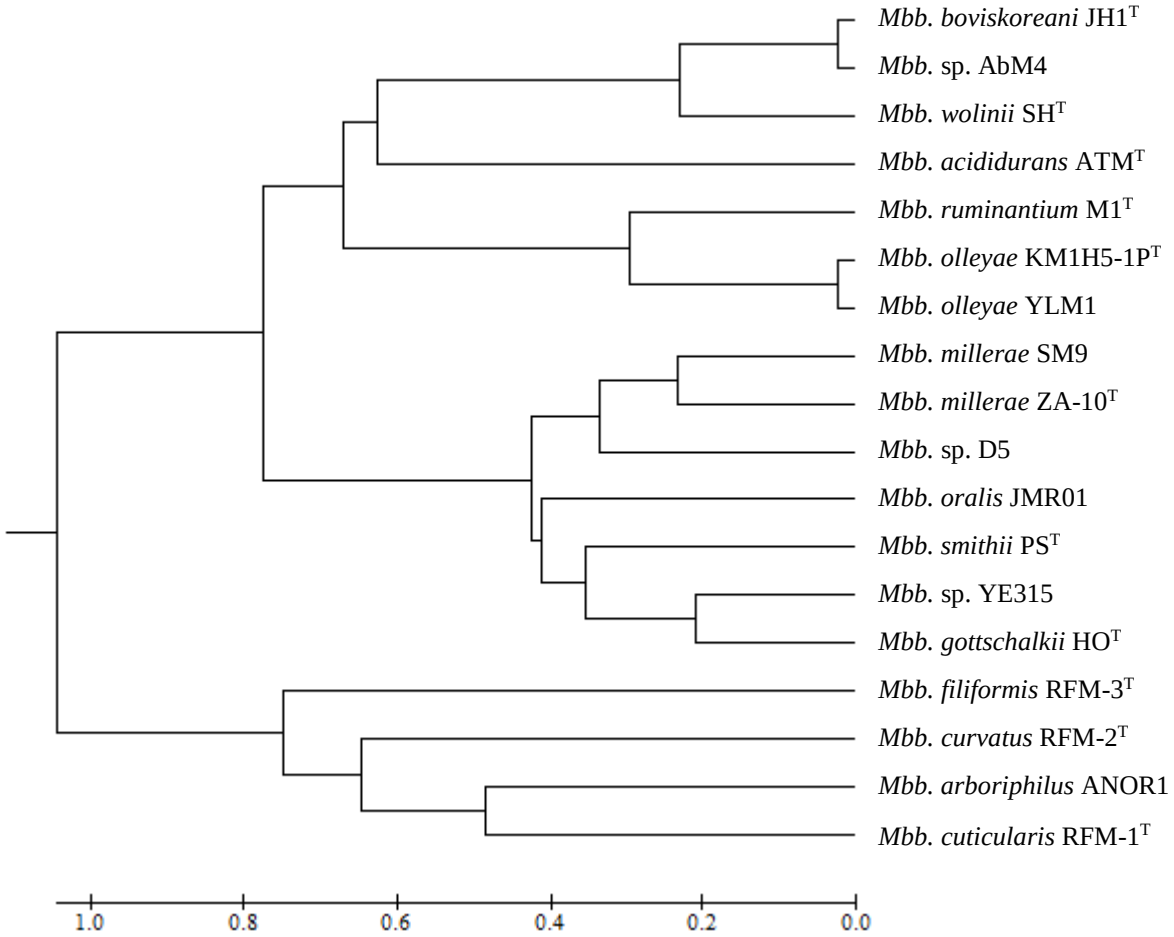


Figure 6.7. Functional Genome Distribution tree based on of the *Methanobrevibacter* spp. ORFeome. The evolutionary history was inferred using the UPGMA method (Sneath and Sokal 1962) based on the functional genome distribution (FGD) matrix (Altermann 2012). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetics tree. Type strains are indicated by superscript T.

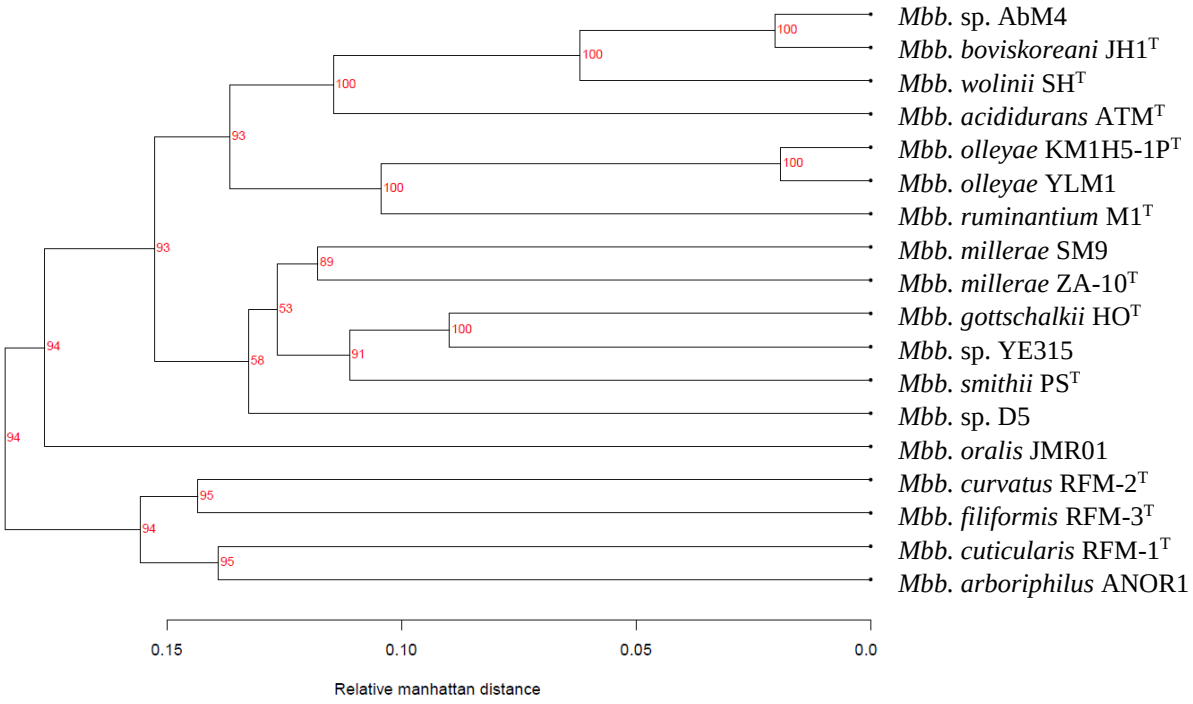


Figure 6.8. Pan-genome tree of sequenced type strains and rumen members of *Methanobrevibacter* spp. Pan-genome family tree based on the absence and presence of gene families. The relative Manhatten distance indicates the proportion of the pan-genome where genomes differ in presence/absence of gene families. Bootstrap values from 1000 resamplings are indicated as percentages at each branch point. Type strains are indicated by superscript ^{T}.

Based on this analysis, *Methanobrevibacter* sp. D5 is most closely related to *Methanobrevibacter* sp. YE315, and most distantly related to *Mbb. aboriphilus* ANOR1 (ANOR1) and the three type strains isolated from the termite hindgut. In this study, strains displaying a 16*S* rRNA gene identity of less than 98.7% are considered different species. Based on the 16*S* rRNA gene sequence identity, *Methanobrevibacter* sp. D5 can be considered as a strain of the *Mbb. millerae* species (Table 6.4.). *Methanobrevibacter* sp. AbM4 is considered the same species as JH1^{T}. YLM1 is considered the same species as *Mbb. olleyae* KM1H5-1P^{T} (KM1H5-1P^{T}) (Table 6.4.). The *Mbb. cuticularis* RFM-1^{T} (RFM-1^{T}), *Mbb. curvatus* RFM-2^{T} (RFM-2^{T}), *Mbb. filiformis* RFM-3^{T} (RFM-3^{T}) and *Mbb. acididurans* ATM^{T} (ATM^{T}) are affiliated to their own clades (Figure 6.6.). Based on the FGD and pan-genome tree, the *Mbb. smithii* PS^{T} (PS^{T}) is similar to the members of the *Mbb. gottschalkii* clade (Figures 6.7., 6.8.).

6.2.4. Genome properties of *Methanobrevibacter* sp. D5 and other sequenced *Methanobrevibacter* spp.

The general features of the circular *Methanobrevibacter* sp. D5 genome are summarised in Table 6.5 and Figure 6.9. The D5 genome is 2,791,821 bp, 33.1 %G+C and encodes 2,487 ORFs. The genome was annotated and the functional characterisation of genes are summarised in Figure 6.10. There are seven pseudogenes predicted in the genome of D5, including a pyridoxamine 5'-phosphate oxidase family protein (D5_0093), a NADH oxidase (D5_0115), an acetyltransferase GNAT family (D5_0931), an ion transport protein (D5_1201), a transporter Na^+/H^+ antiporter (D5_1397), a 5,10-methenyltetrahydromethanopterin hydrogenase Hmd (D5_1879) and a potassium channel protein (D5_2313).

Table 6.5. General features of the *Methanobrevibacter* sp. D5 genome

Feature	
Source	Ovine rumen
Project status	Complete
Genome size (bp)	2,791,821
G+C content (%)	33.1
Number of ORFs	2,487
Coding area (%)	88.8
Contigs	1
rRNAs (5s, 16s, 23s)	3, 2, 2
tRNAs (with introns)	44 (2)
Non-coding RNAs	4
IS	32
Prophage	No
CRISPR regions	1
Adhesin-like proteins	82

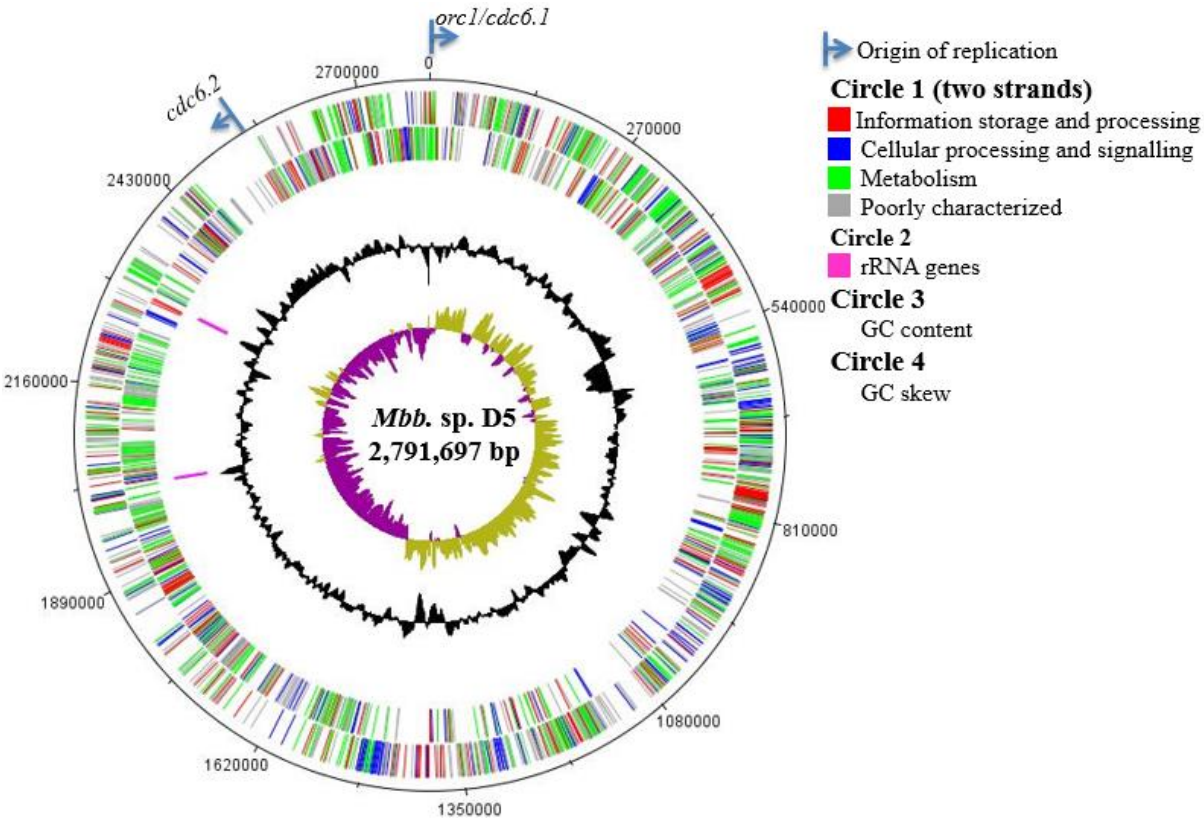


Figure 6.9. Circular representation of the *Methanobrevibacter* sp. D5 genome. Circles are referred to as 1 (outermost) to 4 (innermost). Circle 1: predicted ORFs on the + and – strands respectively. ORFs are coloured based on the COG categories. Circle 2: location of the rRNA genes. Circle 3: %G+C content. Circle 4: GC bias [(G-C)/(G+C)], khaki indicates values >1, purple <1. Origins of replication are represented as blue lines perpendicular to the outermost circle, with an arrow in direction of 5' to 3'.

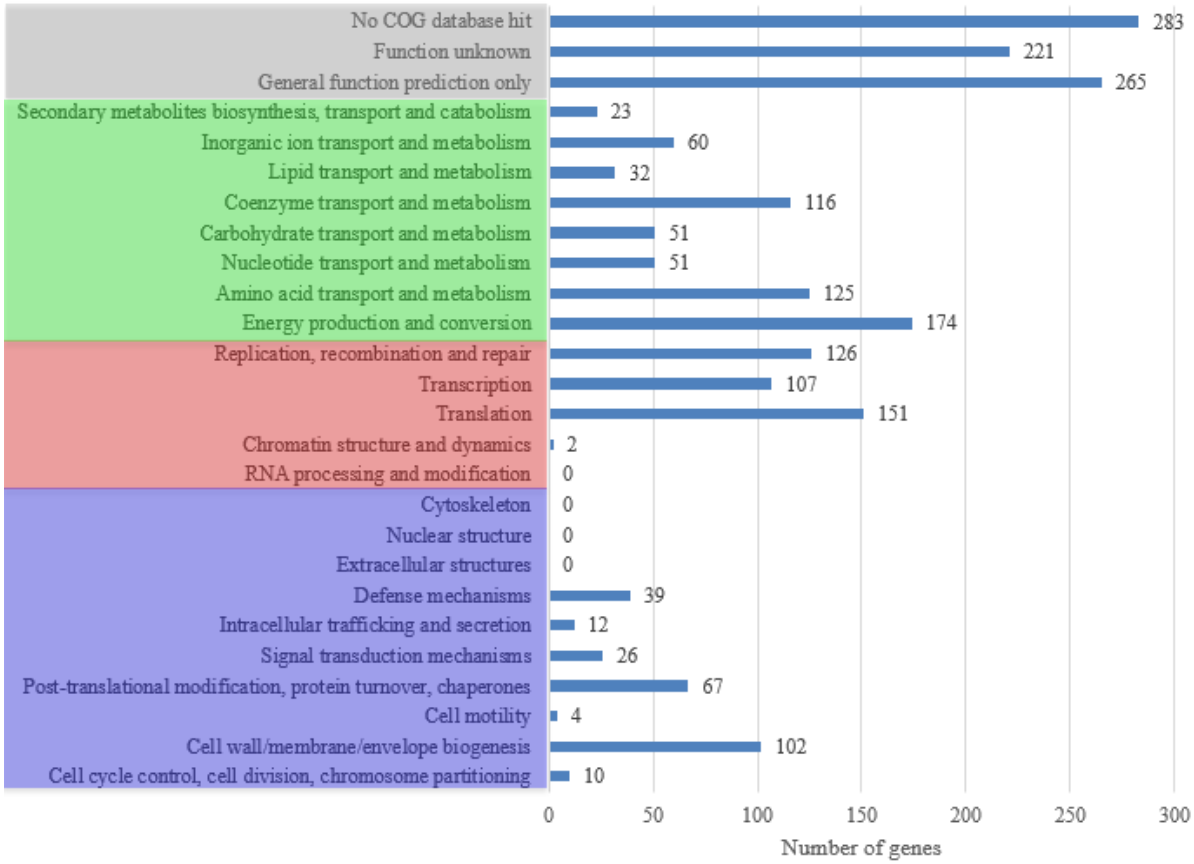


Figure 6.10. Functional classification *Methanobrevibacter* sp. D5 genes. The function classification of genes were predicted based on the COG database (Tatusov *et al.* 2001).

Genome characteristics of available Methanobrevibacter genomes. The general genome characteristics of D5 in comparison to other sequenced *Methanobrevibacter* spp. are listed in Table 6.6.

Table 6.6. General genome properties of sequenced species of the genus *Methanobrevibacter*

Species name	***Methanobrevibacter* sp. D5**	***Methanobrevibacter* sp. YE315**	***Methanobrevibacter millerae* ZA-10^T**	***Methanobrevibacter millerae* SM9**	***Methanobrevibacter gottschalkii* HO^T**	***Methanobrevibacter smithii* PS^T**
Source	Ovine rumen	Bovine rumen	Bovine rumen	Ovine rumen	Horse faeces, ovine rumen$^\alpha$	Sewage digester, human gut$^\beta$
Project status	Complete	Complete	Draft	Complete	Draft	Complete
Genome size (bp)	2,791,697	2,273,296	2,725,667*	2,543,538	1,864,477*	1,853,160
G+C content (%)	33.1	34.2	36.4	31.8	30.2	31
ORFs	2,487	1,942	2,383	2,269	1,895	1,795
Coding area (%)	88.8	83.3	89.6	89	89.3	90.8
Contigs	1	1	48	1	20	1
rRNA operons	2	2	1	2	2	2
tRNAs (with introns)	44 (2)	33 (2)	77 (3)	40 (2)	35 (2)	34 (1)
Non-coding RNA	4	4	3	8	5	2
IS	41	Nd	Nd	65	Nd	8
Prophage	No	Nd	Nd	Yes	Nd	Yes
CRISPR regions (spacers)	1 (20)	1 (6)	0*	2 (17, 20)	0*	1 (43)
Adhesin-like proteins	82	Nd	Nd	95	Nd	48
LPxTG-like motif	0	0	7	3	0	4
Sortase	1	1	1	1	1	1
NCBI accession		CP010834.1		NZ_CP011266.1		NC_009515.1
Publication	This study			Kelly *et al.*, 2016		Samuel *et al.*, 2007
Species name	***Methanobrevibacter ruminantium* M1^T**	***Methanobrevibacter olleyae* KM1H5-1P^T**	***Methanobrevibacter olleyae.* YLM1**	***Methanobrevibacter wolinii* SH^T**	***Methanobrevibacter* sp. AbM4**	***Methanobrevibacter boviskoreani* JH1^T**
Source	Bovine rumen	Ovine rumen	Ovine rumen	Sheep faeces, caprine rumen$^\gamma$	Ovine abomasum	Bovine rumen
Project status	Complete	Draft	Complete	Draft	Complete	Draft
Genome size (bp)	2,937,203	2,122,444*	2,201,192	2,041,814*	1,998,189	2,045,801
G+C content (%)	33	26.8	26.9	24.2	29	27.9
ORFs	2,115	1,813	1,839	1,700	1,671	1,756
Coding area (%)	81	76.9	75.8	76.3	75.8	78.4
Contigs	1	49	1	26	1	54
rRNA operons	2	1	2	2	3	1
tRNAs (with introns)	58 (2)	33 (2)	33 (1)	36 (2)	38 (2)	39 (2)
Non-coding RNA	3	4	5	3	4	3
IS	4	Nd	19	Nd	23	1
Prophage	Yes	Nd	Yes	Nd	No	Nd
CRISPR regions (spacers)	3 (61, 11, 38)	1* (16)	2 (211, 21)	1* (4)	1 (246)	2* (139, 9)
Adhesin-like proteins	105	Nd	53	Nd	29	Nd
LPxTG-like motif	1	1	1	2	3	3
Sortase	1	1	1	1	1	1
NCBI accession	CP001719.1		CP014265.1	JHWX00000000.1	CP004050.1	BAGX00000000.2
Publication	Leahy *et al.*, 2010		Kelly *et al.*, 2016		Leahy *et al.*, 2013	Lee *et al.*, 2013

Table 6.6. General genome properties of sequenced species of the genus *Methanobrevibacter* (continued)

Species name	***Methanobrevibacter cuticularis* RFM-1^T**	***Methanobrevibacter curvatus* RFM-2^T**	***Methanobrevibacter filiformis* RFM-3^T**	***Methanobrevibacter acididurans* ATM^T**	***Methanobrevibacter oralis* JMR01**	***Methanobrevibacter arboriphilus* ANOR1**
Source	Termite gut	Termite gut	Termite gut	Anaerobic digestor	Human faeces, human subgingival plaque$^{\Omega}$	Human faeces
Project status	Draft	Draft	Draft	Draft	Draft	Draft
Genome size (bp)	2,705,405*	2,600,395*	2,989,372*	1,655,232*	2,083,511*	2,221,072*
G+C content (%)	26.9	28.8	27.03	27.3	27.8	25.5
ORFs	2,154	2,149	2,517	1,575	2,251	1,993
Coding area (%)	71.6	76.3	76.2	85.2	82.4	74
Contigs	190	192	460	36	60	5
rRNA operons	1	1	1	1	1	2
tRNAs (with introns)	35 (2)	34 (2)	33 (2)	33 (2)	33 (2)	35 (2)
Non-coding RNA	2	2	3	3	3	3
IS	Nd	Nd	Nd	Nd	Nd	Nd
Prophage	Nd	Nd	Nd	Nd	No	No
CRISPR regions (spacers)	3 (13,6,72)	1 (50)	3* (61, 5, 3)	1* (163)	3* (33, 5, 21)	3 (2, 2, 5)*
Adhesin-like proteins	Nd	Nd	Nd	Nd	Nd	Nd
LPxTG-like motif	0	0	0	1	1	0
Sortase	1	1	1	1	1	1
NCBI accession	LWMW00000000.1	LWMV00000000.1	LWMT00000000.1		CBWS000000000.1	NZ_CBVX000000000.1
Publication					Khelaifia *et al.*, 2014	Khelaifia *et al.*, 2014

*Among currently available sequences. In addition to the environment which the sequenced organism was initially obtained, the organism has also been identified in other environments, as indicated by $^{\alpha}$. (Jeyanathan *et al.* 2011). $^{\beta}$. (Matarazzo *et al.* 2012). $^{\Omega}$. (Irbis and Ushida 2004). $^{\gamma}$. (Ferrari *et al.* 1994). $^{\#}$incomplete regions

Origin of replication. The replication origin (*orc1/cdc6*) of the D5 chromosome was identified by GC nucleotide skew [(G-C)/(G+C)] analysis (Figure 6.9.). Two *orc1/cdc6* genes were predicted in the genome (D5_0001, D5_2279). One Origin Recognition Box (ORB) motif was identified 87 bp upstream of the D5_0001 gene. Therefore, the D5_0001 gene is predicted to be the true origin of replication of D5.

In addition to D5, 12 other *Methanobrevibacter* genomes also host two copies of *cdc6* genes, whereas $M1^T$, ANOR1, AbM4, RFM-2^T and ATM^T, carry an additional truncated *cdc6* gene.

Ribosomal, non-coding and transfer RNAs. The ribosomal RNAs in completely sequenced members of the genus *Methanobrevibacter* have been found in the characteristic operon arrangement, 16*S*, 23*S* and 5*S* rRNA genes. An additional copy of the 5*S* rRNA gene is present in D5, $M1^T$, $JH1^T$, PS^T, ZA-10^T, KM1H5-$1P^T$, YE315, ANOR1, RFM-2^T, RFM-3^T and ATM^T, while two additional copies of the 5*S* rRNA gene are predicted in the genomes of *Mbb. gottschalkii* HO^T (HO^T), *Mbb. oralis* JMR01 (JMR01) and YLM1.

Four non-coding RNAS (ncRNAs) are predicted in the *Methanobrevibacter* sp. D5 genome, including an archaeal RNaseP, an archaeal signal recognition particle, a group II catalytic intron and a FMN riboswitch. The ncRNAs predicted within the other *Mbb.* genomes analysed are summarised in Table 6.7. All of the protein components of RNaseP and the signal recognition particle have been predicted in the genomes analysed.

Seventeen of the *Methanobrevibacter* genomes analysed are predicted to contain all of the tRNA genes for 20 amino acids (Table 6.8.), but the $tRNA^{Ala}$ is not detected in the *Mbb. olleyae* KM1H5-$1P^T$ draft genome. The numbers of tRNAs corresponding to each amino acid is highly conserved between the genomes analysed, with the exception of *Mbb. millerae* ZA-10^T genome and *Mbb. ruminantium* $M1^T$ genome, which contains a comparatively larger number of tRNAs. Introns have been found consistently within the tRNA genes $tRNA^{Trp}$ and $tRNA^{Met}$ across all the *Methanobrevibacter* genomes analysed, with an additional intron in the $tRNA^{Ser}$ gene of *Mbb. millerae* genome and the $tRNA^{Lys}$ gene of *Mbb. ruminantium* $M1^T$ genome.

Table 6.7. Non-coding RNAs predicted in *Methanobrevibacter* spp. genomes

Features	D5	YE315	ZA-10^T	SM9	HO^T	PS^T	JMR01	M1^T	KM1H5-1P^T	YLM1	SH^T	AbM4	JH1^T	RFM-1^T	RFM-2^T	RFM-3^T	ANOR1	ATM^T
Archaeal RNaseP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Archaeal Signal recognition particle	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Group II catalytic intron	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
FMN riboswitch (RFN element)	+	+	+	+		+	+	+	+	+	+	+	+	-	-	-	-	+
HNH endonuclease associated RNA	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
crcB RNA	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-
microRNA mir-598	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-

+ indicates the predicted presence of the ncRNA, - indicates the predicted absence of the ncRNA

Table 6.8. Predicted tRNAs in *Methanobrevibacter* spp. genomes

tRNA corresponding Amino acid	D5	YE315	ZA-10^T	SM9	HO^T	PS^T	JMR01	M1^T	KM1H5-1P^T	YLM1	SH^T	AbM4	JH1^T	RFM-1^T	RFM-2^T	RFM-3^T	ANOR1	ATM^T
Alanine	2	2	2	2	2	2	2	4	0	2	2	2	2	2	2	2	3	2
Argnine	4	4	6	4	4	4	4	5	4	4	3	4	4	5	4	4	4	3
Asparagine	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aspartate	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Cysteine	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Glutamine	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1
Glutamate	1	1	3	1	2	1	1	2	1	1	3	2	2	1	1	1	1	1
Glycine	2	2	4	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2
Histidine	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Isoleucine	2	1	2	1	1	1	1	2	1	1	1	2	2	1	1	1	1	1
Leucine	5	3	8	4	3	3	3	6	3	3	3	3	3	3	3	3	3	3
Lysine	2	1	4	2	1	1	1	4*	1	1	1	2	2	1	1	1	1	2
Methionine	3*	3*	7*	3*	3*	3*	3*	3*	3*	3*	3*	3*	3*	4**	3*	3*	3*	3*
Phenylalanine	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1
Proline	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Serine	4	3	6*	4	3	5	3	6	3	3	3	3	3	3	3	3	3	3
Threonine	2	2	6	4	2	3	2	2	2	2	2	2	2	2	2	2	2	2
Tryptophan	1*	1*	2*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*	1*
Tyrosine	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
Valine	3	2	4	2	2	2	2	4	3	3	2	2	2	2	2	2	2	2
Pseudo	2	0	11	2	1	0	0	6	2	0	2	2	2	0	0	0	0	0
Total	44	33	77	40	35	34	33	58	33	33	36	38	39	35	34	33	35	33
Intron containing tRNAs	2	2	3	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2

*intron-containing tRNAs. ** two tRNAs that contain one intron each

Codon and amino acid usage of Methanobrevibacter species. The codon usage for each genome is summarised in Table A.6.2 and visualised in Figure 6.11.

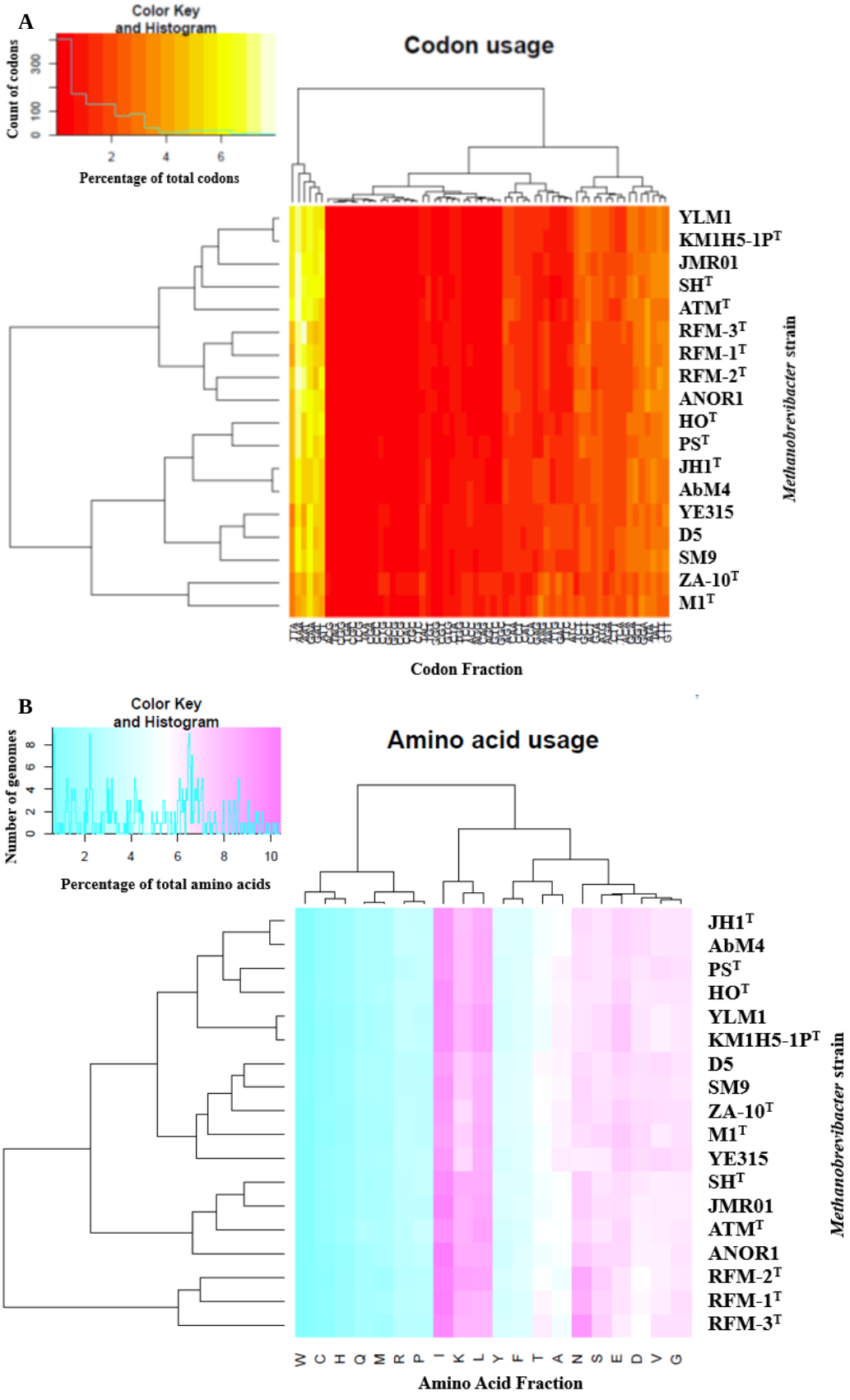


Figure 6.11. Codon and amino acid usage of the *Methanobrevibacter* genomes analysed. **A**. Codon usage heatmap of the Methanobrevibacter genomes analysed. The color key on the top left hand corner, indicates codons with the lowest percentage usage in red, and most abundant codons are displayed in pale yellow. The histogram displays the count of codons in 0.37% intervals across genomes. **B**. Amino acid usage heatmap of the Methanobrevibacter genomes analysed. The heatmap was clustered to place organisms with the shortest distance under the same branch. The color key displayed in this heatmap is illustrated on the top left hand corner, the most scarce amino acid is displayed in cyan, and most abundant amino acid is displayed in magenta. The number of genomes was tallied at intervals of 0.05% and displayed in a histogram. The figure was generated by CMG-Biotools (Vesth et al. 2013).

The *Methanobrevibacter* genomes display a high degree of homogeneity in codon and amino acid usage. The codon usage of ZA-10^T and M1^T are slightly different in comparison to other *Methanobrevibacter* analysed, as reflected in the codon usage for the amino acids leucine and isoleucine (Table A.6.2.). The RFM-1^T, RFM-2^T, RFM-3^T genomes are predicted to use a higher proportion of asparagine. Translational start codons are similar among the genomes investigated (Table A.6.2.). The KM1H5-1P^T, YLM1, SH^T, AbM4, JH1^T and ANOR1 genomes have a stronger preference for using the opal codon as a stop codon (68.2% to 74.1%) than the remainder of the genomes analysed (57.5% to 64.9%) (Table A.6.2.).

Horizontal gene transfer of Methanobrevibacter sp. *D5.* Alien Hunter (Vernikos and Parkhill, 2006) was used to identify horizontally transferred regions independent of genome annotation. There are 80 regions predicted by Alien Hunter to contain horizontally acquired DNA based on atypical sequence composition (Table 6.9.), and these regions account for 30.7% of the D5 genome. A total of 535 genes are predicted across the 80 regions. Region 30 had a moderate likelihood of being horizontally transferred.

Table 6.9. Predicted horizontal gene transfer regions of *Methanobrevibacter* sp. D5

Region	Base range	IVOM score*	%G+C	Number of genes	Average CAI*	Likelihood
1	12500..17500	24.818	35.2	1	0.779	Low
2	52500..82500	20.738	34.2	7	0.679	Low
3	150000..165000	29.713	36.0	3	0.694	Low
4	175000..180000	17.788	32.4	6	0.690	Low
5	225000..230000	17.331	36.4	6	0.802	Low
6	247500..252500	16.185	32.4	2	0.794	Low
7	282500..292500	19.14	36.3	4	0.677	Low
8	350000..360000	37.63	37.4	13	0.793	Low
9	365000..370000	22.739	34.2	4	0.809	Low
10	465000..492500	26.629	37.3	24	0.801	Low
11	495000..500000	19.895	36.3	7	0.589	Low
12	502500..512500	25.64	36.7	8	0.606	Low
13	515000..525000	27.473	37.9	4	0.639	Low
14	535000..540000	19.643	24.0	3	0.793	Low
15	547500..570000	21.184	26.2	13	0.768	Low
16	572500..590000	25.181	25.7	8	0.768	Low
17	597500..627500	29.052	37.8	28	0.622	Low
18	630000..635000	18.735	37.4	3	0.606	Low
19	655000..662500	17.717	35.8	5	0.656	Low
20	710000..715000	20.107	26.1	2	0.776	Low
21	772500..787500	20.949	33.0	26	0.817	Low
22	797500..802500	18.459	31.8	9	0.839	Low
23	815000..820000	15.563	33.7	4	0.835	Low
24	887500..892500	23.403	38.2	4	0.635	Low
25	895000..900000	24.79	38.6	4	0.655	Low

26	910000..917500	18.885	25.8	5	0.799	Low
27	935000..942500	19.447	37.5	7	0.677	Low
28	945000..950000	21.946	39.1	4	0.695	Low
29	952500..957500	17.069	35.5	3	0.637	Low
30	972500..987500	46.045	35.3	7	0.652	Moderate
31	992500..997500	18.446	34.8	2	0.675	Low
32	1000000..1007500	18.759	30.1	6	0.725	Low
33	1042500..1047500	15.802	33.6	3	0.638	Low
34	1062500..1070000	18.391	37.7	5	0.699	Low
35	1080000..1112500	21.618	34.2	3	0.762	Low
36	1115000..1135000	26.752	35.0	1	0.765	Low
37	1187500..1197500	17.395	35.0	3	0.714	Low
38	1225000..1232500	17.675	35.5	8	0.652	Low
39	1252500..1260000	26.053	34.7	2	0.750	Low
40	1275000..1280000	16.692	34.8	5	0.661	Low
41	1330000..1342500	17.087	35.6	16	0.635	Low
42	1345000..1362500	24.993	37.1	21	0.603	Low
43	1380000..1397500	18.324	27.2	12	0.754	Low
44	1402500..1410000	21.206	36.0	7	0.574	Low
45	1420000..1437500	29.047	26.8	11	0.688	Low
46	1440000..1462500	24.752	37.2	25	0.599	Low
47	1525000..1530000	16.604	36.2	3	0.668	Low
48	1580000..1592500	30.052	34.7	2	0.757	Low
49	1595000..1600000	16.993	35.3	2	0.730	Low
50	1627500..1632500	21.721	33.6	12	0.709	Low
51	1655000..1662500	18.642	33.7	3	0.748	Low
52	1665000..1675000	22.437	33.6	2	0.781	Low
53	1800000..1805000	17.586	36.3	3	0.626	Low
54	1810000..1830000	39.015	34.2	1	0.785	Low
55	1862500..1870000	18.655	31.1	4	0.809	Low
56	1932500..1942500	16.788	31.4	6	0.737	Low
57	1992500..2000000	21.047	38.2	6	0.731	Low
58	2015000..2025000	60.635	39.3	4	0.736	Low
59	2140000..2145000	24.285	36.7	4	0.797	Low
60	2195000..2200000	16.104	36.6	3	0.710	Low
61	2245000..2252500	21.284	25.74	6	0.801	Low
62	2255000..2260000	22.598	35.9	3	0.681	Low
63	2262500..2270000	18.925	36.7	11	0.717	Low
64	2295000..2307500	47.987	39.1	6	0.689	Low
65	2340000..2350000	20.18	37.6	10	0.670	Low
66	2352500..2357500	17.623	33.9	2	0.766	Low
67	2370000..2385000	45.87	39.5	5	0.774	Low
68	2417500..2427500	16.264	36.7	9	0.726	Low
69	2445000..2452500	18.62	37.1	8	0.675	Low
70	2460000..2530000	32.628	36.8	5	0.738	Low
71	2532500..2542500	29.914	37.7	1	0.708	Low
72	2550000..2555000	16.571	37.5	8	0.697	Low
73	2560000..2567500	19.277	36.4	10	0.665	Low
74	2572500..2577500	22.203	36.0	2	0.618	Low
75	2597500..2602500	15.903	36.6	9	0.662	Low
76	2627500..2632500	18.747	37.4	2	0.649	Low
77	2647500..2665000	27.298	36.2	16	0.730	Low
78	2675000..2685000	22.538	37.4	5	0.699	Low
79	2730000..2737500	19.002	37.9	3	0.698	Low
80	2742500..2750000	17.388	33.1	5	0.737	Low

*Atypical sequence composition is indicated by IVOM score, average CAI and %G+C. IVOM scores is annotation independent and it implemented variable order *k*-mers as reliable estimates of local sequence composition, as higher order motifs are more likely to capture deviation from genome background. A high regional IVOM score, abnormal % G+C and low average CAI is associated with a high likelihood of horizontal transfer.

Region 30 has a 35.3% G+C content close to the average 33.1% G+C content of the genome. This region has a high IVOM score with slightly low CAI. This region encodes six hypothetical proteins and a transposase.

Regularly Interspaced Short Palindromic Repeat (CRISPR) element. The *Methanobrevibacter* sp. D5 genome contains several CRISPR-associated genes (D5_0810, 0880, 0881, 1219 and 1929) with a CRISPR region containing 19 repeats (bases 1,381,409 to 1,382,733). The direct

repeat is 29 bp in length with the consensus sequence GTTTAAAATAGACTTAATAGTATGGAAAT. The presence of CRISPR-associated genes *cas*1, 2, 4, 6 and TIGR02710 suggested the D5 CRISPR/Cas system belongs to type U. A total of 20 predicted spacer sequences with species specific protospacer adjacent motif (PAM) were used to identify the potential targets of CRISPR RNAs (crRNAs). The predicted crRNA targets are displayed in Table 6.10. Spacers with no database matches (spacers 1, 6, 7, 8, 9, 15, 17, 18, 19) were omitted.

Table 6.10. CRISPR spacer homology of *Methanobrevibacter* sp. D5

Spacer Number	Spacer sequence match*	Accession number	Score
2, 20	*Providencia rettgeri* strain 09ACRGNY2001 plasmid pPrY2001	NC_022589	22
3	Cyanophage P-RSM1	HQ634175	18
4	*Escherichia* phage JMPW2	KU194205	18
5	*Bacillus cereus* strain S2-8 plasmid pBFR_2	NZ_CP009606	18
10	Periwinkle leaf yellowing phytoplasma plasmid p09PLY-1	NC_019245	23
11	*Listeria* phage LP-048	KJ094033	20
12	*Zymomonas mobilis* subsp. Mobilis str. CP4 = NRRL B-14023 plasmid pZM1402301	NZ_CP003716	16
13	*Candidatus* pantoea carbekii strain US plasmid pBMSBPS3	NZ_CP010910	18
14	*Synechococcus* phage S-SSM5	GU071097	22
16	*Clostridium perfringens* strain JP838 plasmid pJFP838A	NZ_CP013615	17

* Spacer sequence identified by BLAST based homology screening (Biswas *et al*. 2013). Score was calculated by matches (+1) and mismatches (-1) across the whole length of the spacer without gaps.

The CRISPR elements were predicted in 16 of the genomes analysed, containing from 1 to 3 CRISPR elements, with 4 to 246 spacer sequences (Table 6.6.). No CRISPR elements have been predicted in the genomes of ZA-10^T and HO^T. The CRISPR associated genes have been predicted and are summarised in Table A.6.3.

Insertion sequence (IS) elements of Methanobrevibacter sp. D5. IS elements were annotated as described in Section 2.2.22. A total of 32 IS elements were identified in the genome of *Methanobrevibacter* sp. D5, and these elements account for 1.8% of the coding DNA sequence. Two types of IS are found; transposons and phage integrases. The IS elements found in *Methanobrevibacter* sp. D5 belong to families IS200/IS605, IS1380, IS1182, IS5, IS630 and ISNCY, with the IS200/IS605 family predicted with the highest copy number (Table 6.11.).

Table 6.11. IS elements of *Methanobrevibacter* sp. D5

Locus tag	Family	Closest aligned source
D5_0467	IS1182	*Bacillus cereus*
D5_0936	IS1182	*Uncultured archaeon*
D5_1082	IS1182	*Methanosarcina acetivorans*
D5_1480	IS1182	*Microscilla sp.*
D5_0440	IS1380	*Rhodococcus opacus*
D5_0439	IS200/IS605	*Caldicellulosiruptor saccharolyticus*
D5_0441	IS200/IS605	*Methanosarcina acetivorans*
D5_0442	IS200/IS605	*Methanosarcina mazei*
D5_0506	IS200/IS605	*Caldicellulosiruptor saccharolyticus*
D5_0829	IS200/IS605	*Enterococcus faecium*
D5_0830	IS200/IS605	*Enterococcus faecium*
D5_1102	IS200/IS605	*Halobacillus halophilus*
D5_1177	IS200/IS605	*Halobacillus halophilus*
D5_1330	IS200/IS605	*Halobacillus halophilus*
D5_1331	IS200/IS605	*Enterococcus faecium*
D5_1622	IS200/IS605	*Halobacillus halophilus*
D5_1994	IS200/IS605	*Enterococcus faecium*
D5_2272	IS200/IS605	*Methanosarcina mazei*
D5_0892	IS5	*Methanobrevibacter ruminantium*
D5_2002	IS5	*Methanohalophilus mahii*
D5_2474	IS5	*Methanohalobium evestigatum*
D5_0019	IS630	*Archaeoglobus fulgidus*
D5_1402	IS630	*Archaeoglobus fulgidus*
D5_1403	IS630	*Uncultured archaeon*
D5_1766	IS630	*Archaeoglobus fulgidus*
D5_0495	ISNCY	*Methanobrevibacter smithii*
D5_0709	ISNCY	*Methanobrevibacter smithii*
D5_0710	ISNCY	*Methanobrevibacter smithii*
D5_0711	ISNCY	*Methanobrevibacter smithii*
D5_1223	ISNCY	*Methanobrevibacter smithii*
D5_1505	ISNCY	*Methanobrevibacter smithii*
D5_2341	ISNCY	*Methanobrevibacter smithii*

Secretome. The *Methanobrevibacter* sp. D5 genome encodes 173 ORFs predicted to contain a signal peptide (Figure 6.12A.). Among the ORFeome proteins that could be exported, most ORF functions have been classified under the poorly characterized COG category (Figure 6.12B.). Some archaea utilise ESCRT-III (eukaryotic endosomal complex required for transport) protein for vesicle secretion, but no homologues of ESCRT-III subunits involved in vesicle secretion were detected in the *Methanobrevibacter* sp. D5 genome.

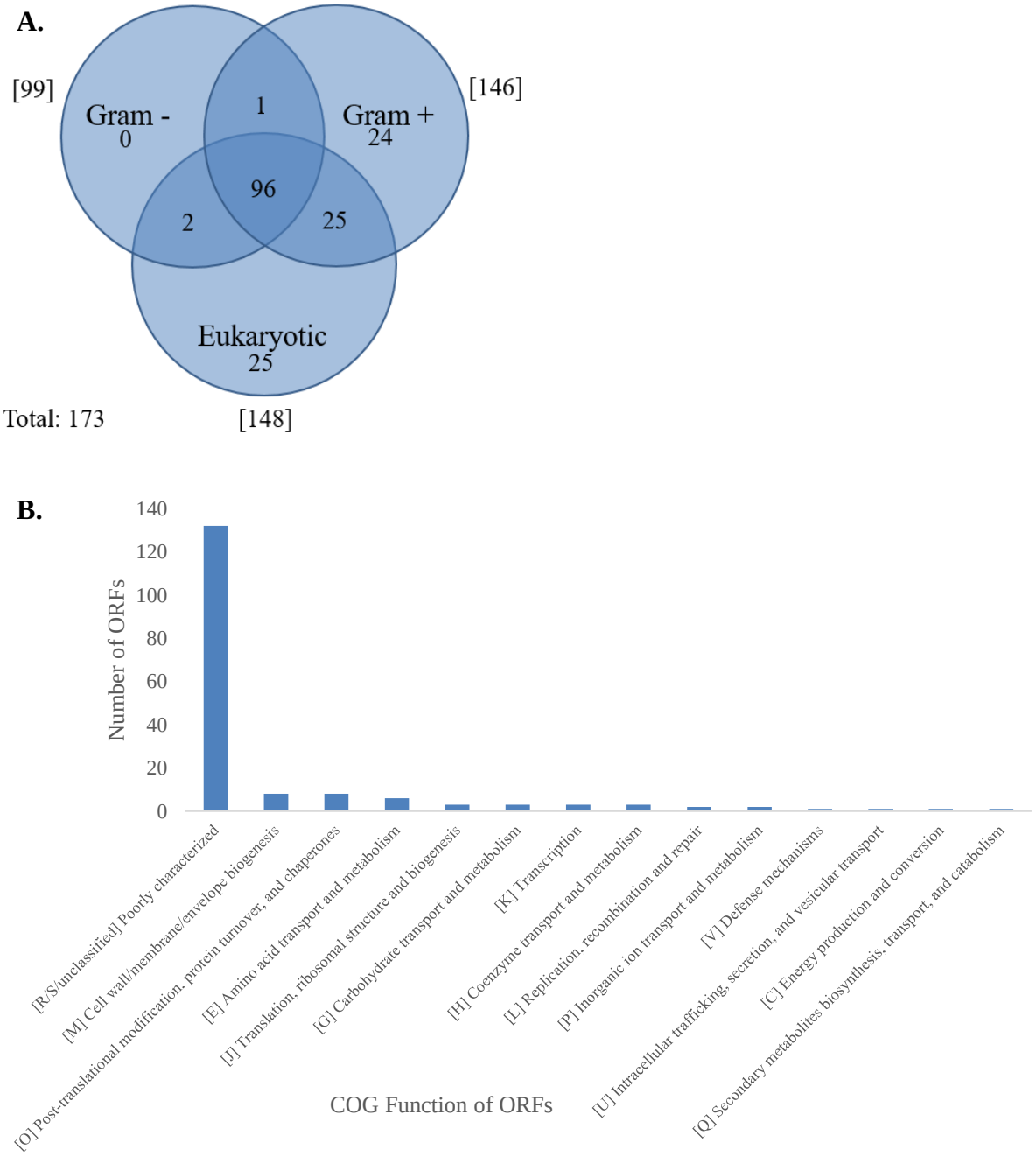


Figure 6.12. Distribution of *Methanobrevibacter* sp. D5 signal peptides. **A**. Number of signal peptide ORFs classified according to SignalP training model. **B**. COG classification of D5 ORFs predicted with signal peptide.

The predicted secretome (surface associated and/or released from the cell) proteins of the 18 *Methanobrevibacter* genomes are summarised in Table 6.12 and Table A.6.4. The genomes analysed are predicted to devote between 3.5% and 7% of their ORFs to the export of extracellular and surface-associated proteins. A large proportion of the secretomes examined are predicted to have ORFs with motifs or domains for attachment to the cell surface. Amongst the genomes analysed, between 64 to147 genes per genome are predicted to be membrane anchored by one to two transmembrane helices or by an N-terminal lipobox. Around 3 to 27 genes per genome are predicted to be either unattached (i.e. secreted) or associated with the cell wall using other (unknown) mechanisms. The extracellular proteins were analysed for the presence of repeat domains and periplasmic domains (the periplasmic domain belongs to the substrate binding subunit of ABC transporters) and the results are listed in Table A.6.4.

Table 6.12. Predicted secretome in *Methanobrevibacter* species

Secretome	**Integral membrane protein**	**Lipobox**	**LPxTG**	**Two TMH Membrane anchor**	**C-terminal membrane anchor**	**N-terminal membrane anchor**	**Secreted**	**Total (% of total ORFs)**
D5	13	9	0	20	2	103	26	173 (7)
YE315	11	10	0	19	1	65	15	121 (6.2)
ZA-10^T	11	18	7	7	3	82	27	155 (6.5)
SM9	9	16	3	16	5	67	19	135 (5.9)
HO^T	9	6	0	11	1	47	9	83 (4.4)
PS^T	3	5	4	11	5	41	8	77 (4.2)
JMR01^T	7	4	1	7	0	68	13	100 (4.4)
M1^T	9	8	1	10	1	90	14	133 (6.3)
KM1H5-1P^T	7	3	1	11	3	47	10	82 (4.5)
YLM1	9	4	1	15	4	46	11	90 (4.9)
SH^T	6	4	2	18	1	26	3	60 (3.5)
AbM4	7	3	3	12	3	28	4	60 (3.6)
JH1^T	8	3	3	11	2	33	5	65 (3.7)
RFM-1^T	10	12	0	21	4	106	23	176 (8.2)
RFM-2^T	5	13	0	12	2	80	12	124 (5.8)
RFM-3^T	5	11	0	14	1	73	22	126 (5.0)
ATM^T	7	9	0	3	1	38	7	65 (4.1)
ANOR1	6	1	0	13	1	43	12	76 (3.8)

Prediction E value cutoff 1 x 10^{-05}.

Adhesin-like proteins of Methanobrevibacter sp. D5. There are 81 genes predicted to be adhesin-like proteins in the D5 genome and these are displayed in Table 6.13, showing their predicted subcellular location, molecular weight, transmembrane helices, signal peptide and protein domains based on Pfams. Four adhesin-like proteins are predicted to contain a transglutaminase domain (D5_0392, 0657, 1669, 1926), and one adhesin-like protein is predicted to contain a collagen-binding surface protein (D5_1479).

Synteny. The synteny between completed genomes was analysed using D5 as a reference (Figure 6.13.). D5 shares a high degree of synteny with YE315 and SM9, the syntenic breakpoint in SM9 included a region downstream of an adhesin-like protein (D5_0186) and a region upstream of a hypothetical protein (D5_2337), encompassing approximately 2.3 Mb (Figure 6.13E.). The syntenic breakpoint in YE315 included an acety-CoA acetyltransferase gene (D5_0873) and a region downstream of a NRPS (D5_1744), and encompasses approximately 1 Mb (Figure 6.13F.), whereas a weaker X-shaped alignment was observed between D5 and AbM4 with multiple syntenic breakpoints (Figure 6.13A.).

Table 6.13. Adhesin-like protein prediction of *Methanobrevibacter* sp. D5

Locus_tag	Size (kDa*)	TMH#	SignalP	SPAAN P_{ad}-value	Predicted gene product	LOCTree3	PSORTb	SubLoc	Domain
D5_0014	**210.13**	Yes	Yes	0.92	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0015	**242.85**	Yes	Yes	0.88	Adhesin-like protein	Cytoplasm	Cellwall	Extracellular	Bacterial Ig-like, domain 1
D5_0016	**206.08**	Yes	No	0.90	Adhesin-like protein	Secreted	Extracellular	Periplasmic	Pectin lyase fold
D5_0036	**353.28**	Yes	Yes	0.92	Adhesin-like protein	Secreted	Cellwall	Periplasmic	Domain of unknown function DUF11
D5_0038	**466.91**	Yes	Yes	0.98	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_0051	22.80	Yes	Yes	0.94	Adhesin-like protein	Secreted	Extracellular	Extracellular	Bacterial Ig-like, domain 1
D5_0111	131.09	Yes	Yes	0.84	Adhesin-like protein	Cytoplasm	Cellwall	Extracellular	Pectin lyase fold
D5_0112	**555.26**	Yes	Yes	0.93	Adhesin-like protein	Secreted	Cellwall	Cytoplasmic	Pectin lyase fold
D5_0127	106.98	Yes	Yes	0.78	Adhesin-like protein	Secreted	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_0173	16.91	Yes	Yes	0.89	Adhesin-like protein	Secreted	Extracellular	Extracellular	
D5_0188	155.41	Yes	Yes	0.85	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_0189	136.01	No	No	0.77	Adhesin-like protein	Secreted	Unknown	Extracellular	Carbohydrate-binding-like fold
D5_0218	188.19	Yes	Yes	0.86	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0219	78.79	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Pectin lyase fold
D5_0252	20.19	Yes	Yes	0.83	Adhesin-like protein	Plasma membrane	Unknown	Extracellular	
D5_0262	81.64	Yes	Yes	0.72	Adhesin-like protein	Plasma membrane	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_0392	22.21	No	No	0.80	Adhesin-like protein	Secreted	Extracellular	Extracellular	Bacterial Ig-like, domain 1
D5_0393	28.28	No	No	0.84	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Bacterial Ig-like, domain 1
D5_0394	30.02	No	No	0.80	Adhesin-like protein	Secreted	Extracellular	Extracellular	Transglutaminase-like
D5_0452	183.19	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	CAP domain
D5_0453	93.69	Yes	Yes	0.70	Adhesin-like protein	Secreted	Extracellular	Extracellular	Bacterial Ig-like, domain 1
D5_0458	69.68	Yes	Yes	0.70	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_0459	106.59	Yes	Yes	0.79	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_0478	19.38	Yes	No	0.78	Adhesin-like protein	Secreted	Unknown	Cytoplasm	
D5_0640	33.91	Yes	Yes	0.79	Adhesin-like protein	Secreted	Cellwall	Extracellular	Bacterial Ig-like, domain 1
D5_0660	94.35	No	No	0.87	Adhesin-like protein	Secreted	Unknown	Extracellular	Transglutaminase-like, Bacterial Ig-like, domain 1
D5_0715	84.41	No	Yes	0.77	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0821	58.59	No	No	0.86	Adhesin-like protein	Secreted	Extracellular	Extracellular	Pectin lyase fold
D5_0825	52.82	Yes	Yes	0.90	Adhesin-like protein	Secreted	Extracellular	Extracellular	P22 tailspike C-terminal domain-like
D5_0867	67.09	Yes	Yes	0.84	Adhesin-like protein	Secreted	Extracellular	Extracellular	Carboxypeptidase, regulatory domain
D5_0965	26.38	Yes	Yes	0.71	Adhesin-like protein	Secreted	Extracellular	Extracellular	
D5_0970	**387.18**	Yes	Yes	0.90	Adhesin-like protein	Secreted	Cellwall	Extracellular	Quinonprotein alcohol dehydrogenase-like superfamily
D5_0971	**674.93**	Yes	Yes	0.94	Adhesin-like protein	Plasma membrane	Cellwall	Extracellular	Pectin lyase fold
D5_0973	**257.92**	Yes	Yes	0.83	Adhesin-like protein	Plasma membrane	Cellwall	Extracellular	Peptidase C1A
D5_0974	**569.24**	Yes	Yes	0.97	Adhesin-like protein	Cytoplasm	Cellwall	Extracellular	Pectin lyase fold
D5_0976	188.57	Yes	Yes	0.86	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_0978	90.69	Yes	Yes	0.81	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_0979	84.77	Yes	Yes	0.72	Adhesin-like protein	Secreted	Unknown	Extracellular	
D5_1071	38.67	Yes	Yes	0.71	Adhesin-like protein	Secreted	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_1092	**251.97**	No	Yes	0.96	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_1152	93.71	Yes	Yes	0.70	Adhesin-like protein	Secreted	Unknown	Extracellular	Domain of unknown function DUF3344
D5_1322	42.03	Yes	Yes	0.72	Adhesin-like protein	Secreted	Unknown	Extracellular	

D5_1393	**214.15**	No	Yes	0.83	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1400	19.34	Yes	Yes	0.93	Adhesin-like protein	Secreted	Extracellular	Extracellular	
D5_1406	**438.14**	No	Yes	0.88	Adhesin-like protein	Secreted	Cellwall	Extracellular	Filamin repeat
D5_1411	154.47	Yes	Yes	0.84	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1412	94.65	No	Yes	0.84	Adhesin-like protein	Secreted	Cellwall	Extracellular	Bacterial Ig-like, domain 1
D5_1413	168.85	Yes	Yes	0.87	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1423	126.25	Yes	Yes	0.84	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1424	186.87	Yes	Yes	0.81	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1473	**239.48**	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	Collagen-binding surface protein
D5_1476	**440.17**	Yes	Yes	0.96	Adhesin-like protein	Secreted	Cellwall	Extracellular	Bacterial Ig-like, group 2
D5_1494	44.9	No	No	0.71	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1610	30.23	Yes	Yes	0.84	Adhesin-like protein	Secreted	Extracellular	Periplasmic	Domain of unknown function DUF1002
D5_1618	134.96	Yes	Yes	0.86	Adhesin-like protein	Secreted	Unknown	Extracellular	Peptidase C1A
D5_1619	**687.06**	Yes	Yes	0.97	Adhesin-like protein	too long	Cellwall	Extracellular	Domain of unknown function DUF11
D5_1662	142.47	Yes	Yes	0.91	Adhesin-like protein	Secreted	Unknown	Extracellular	Transglutaminase-like, Bacterial Ig-like, domain 1
D5_1689	22.57	Yes	No	0.78	Adhesin-like protein	Secreted	Extracellular	Periplasmic	Zinc-ribbon domain
D5_1715	62.73	Yes	Yes	0.79	Adhesin-like protein	Cytoplasm	Unknown	Cytoplasmic	Pectin lyase fold
D5_1716	105.69	Yes	Yes	0.86	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1746	13.68	Yes	No	0.81	Adhesin-like protein	Secreted	Extracellular	Extracellular	Immunoglobulin-like fold
D5_1758	72.15	Yes	No	0.71	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_1764	67.74	Yes	Yes	0.76	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Peptidase C39
D5_1896	37.36	No	Yes	0.70	Adhesin-like protein	Secreted	Cellwall	Extracellular	
D5_1912	**206.63**	Yes	Yes	0.76	Adhesin-like protein	Secreted	Unknown	Periplasmic	Domain of unknown function DUF11
D5_1921	104.46	Yes	Yes	0.87	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Transglutaminase-like, Bacterial Ig-like, domain 1
D5_2120	14.17	Yes	Yes	0.86	Adhesin-like protein	Secreted	Unknown	Extracellular	
D5_2160	38.47	No	No	0.70	Adhesin-like protein	Secreted	Unknown	Extracellular	Bacterial Ig-like, domain 1
D5_2173	194.14	No	Yes	0.78	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2175	**201.09**	Yes	Yes	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2176	**231.5**	Yes	Yes	0.75	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2261	**803.2**	Yes	Yes	0.99	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2264	**1022.02**	Yes	Yes	0.99	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2265	**514.69**	Yes	Yes	0.99	Adhesin-like protein	Secreted	Cytoplasmic	Extracellular	Pectin lyase fold
D5_2266	**384.67**	Yes	Yes	0.95	Adhesin-like protein	Secreted	Unknown	Extracellular	Pectin lyase fold
D5_2335	46.85	Yes	Yes	0.76	Adhesin-like protein	Secreted	Cytoplasmic	Periplasmic	Bacterial Ig-like, domain 1
D5_2337	66.75	Yes	Yes	0.82	Adhesin-like protein	Secreted	Extracelular	Extracellular	Glycoside hydrolase superfamily
D5_2354	44.76	No	No	0.72	Adhesin-like protein	Secreted	Extracellular	Extracellular	Peptidase C1A
D5_2356	11.62	No	No	0.77	Adhesin-like protein	Secreted	Cellwall	Extracellular	Pectin lyase fold
D5_2357	18.51	No	No	0.90	Adhesin-like protein	Cytoplasm	Cytoplasmic	Extracellular	Pectin lyase fold
D5_2407	131.47	No	No	0.70	Adhesin-like protein	Cytoplasm	Extracellular	Extracellular	Peptidase C1A

[*] Predicted molecular weight in kDa, proteins above 200 kDa are displayed in bold. [#]. TMH: Transmembrane helices.

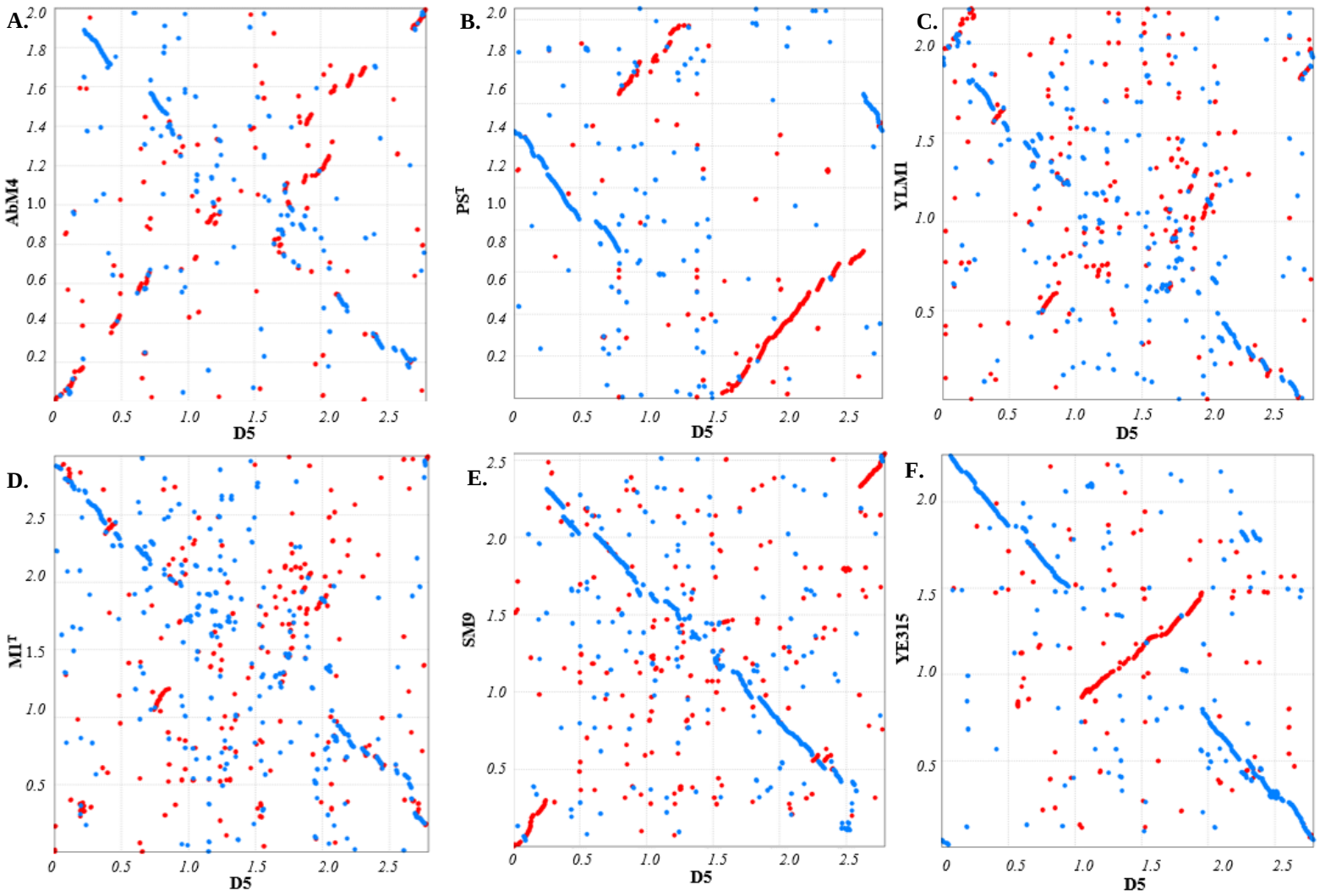


Figure 6.13. Gene synteny plots for completed genomes of *Methanobrevibacter* spp. PROmer alignments of the D5 genome against genomes from *Methanobrevibacter* spp. are shown. The alignments were plotted using MUMmer (Delcher *et al.* 2003) with forward matches shown in red and reverse matches in blue. The units displayed on both axes are in Mb.

6.2.5. Comparative analysis of gene families in *Methanobrevibacter* spp.

The term core-genome was used to describe the genes or gene families present in all *Methanobrevibacter* genomes analysed while the pan-genome was used to define the full complement of genes or gene families present in all sequenced *Methanobrevibacter* genomes. The core genome consists of 764 gene families and the pan-genome consists of 9,382 gene families (Figure 6.14.).

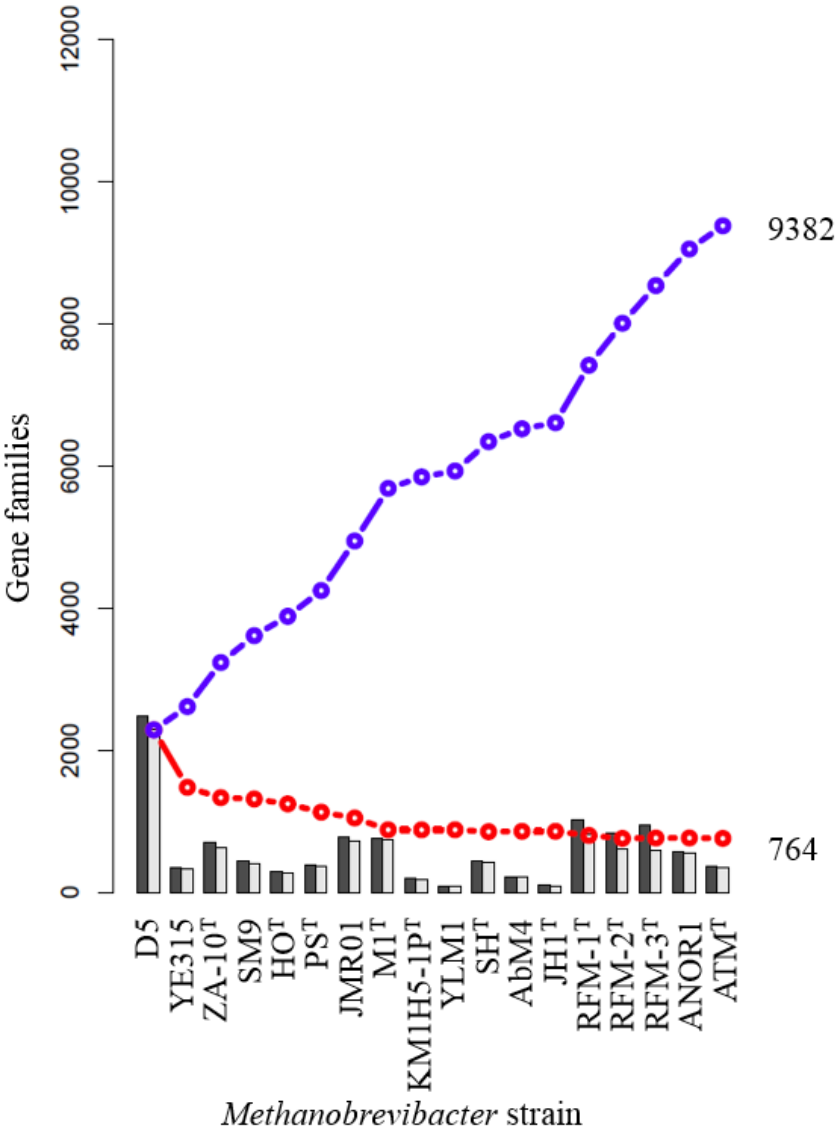


Figure 6.14. Pan-genome and core-genome plot of *Methanobrevibacter* genomes. The blue line indicates the cumulative pan-genome, and the red line indicates the core-genome. The black bars indicate the number of new genes with the addition of each genome across the x-axis; the light grey bars indicates new gene families.

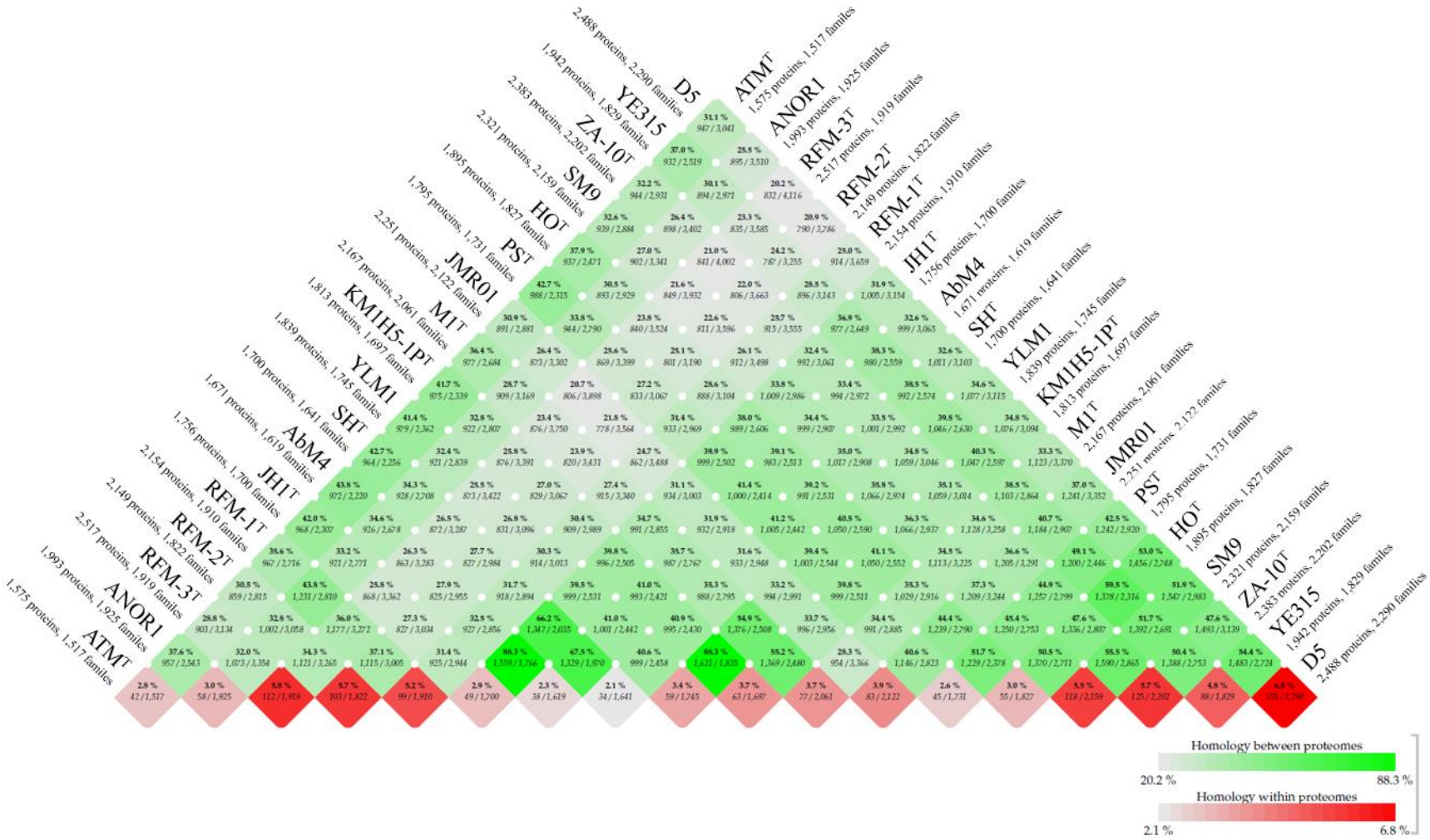
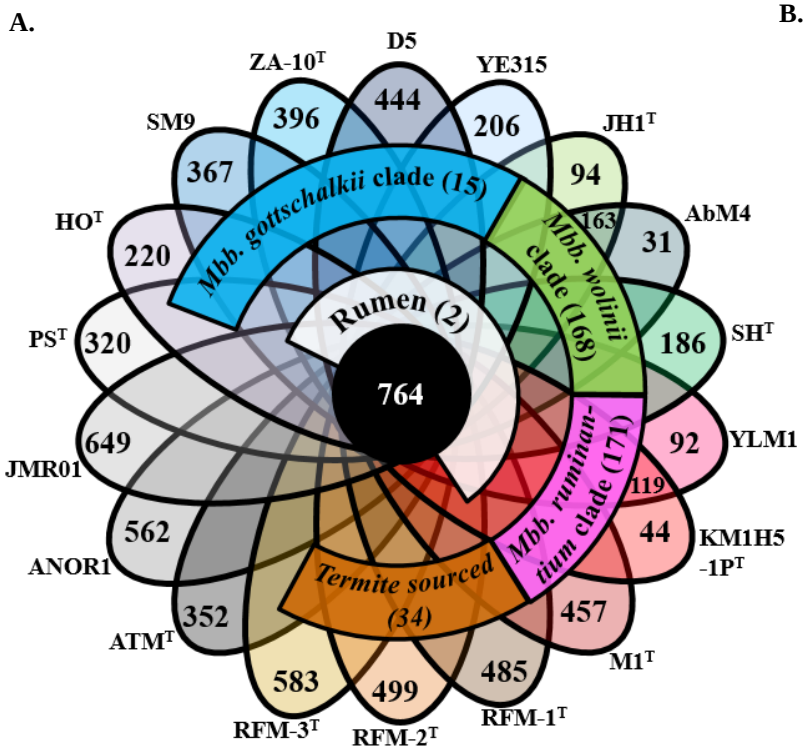


Figure 6.15. BLASTP matrix illustrating the number of conserved protein families between *Methanobrevibacter* genomes. Conservation is defined as 50% coverage and 50% identity. The color intensities are based on the relative percentage of conserved gene families; green depicts conserved protein families between genomes, red depicts protein homology within a genome.

The pattern of gene family conservation as illustrated in Figure 6.15 shows SH^T, AbM4 and $JH1^T$ are similar to each other, while $KM1H5\text{-}1P^T$ and YLM1 are more similar to each other. The analysis shows that D5 has the highest percentage of paralogs. The highest predicted gene family conservation occurs between the $JH1^T$/AbM4 pair and between the YLM1/$KM1H5\text{-}1P^T$ pair at 88.3% (Figure 6.15.).

A.



B.

Genome	Gene families
D5	2,290
YE315	1,829
SM9	2,159
$ZA\text{-}10^T$	2,202
HO^T	1,827
SH^T	1,641
AbM4	1,619
$JH1^T$	1,700
$M1^T$	2,061
YLM1	1,745
$KM1H5\text{-}1P^T$	1,697
JMR01	2,122
PS^T	1,731
ANOR1	1,925
$RFM\text{-}1^T$	1,910
$RFM\text{-}2^T$	1,822
$RFM\text{-}3^T$	1,919
ATM^T	1,517

Figure 6.16. Conserved and novel gene families among the 18 *Methanobrevibacter* genomes. **A**. Venn diagram indicating the number of conserved gene families between genomes based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria. The conserved gene families within each clade and conserved gene families between all genomes of rumen origin are presented in brackets. **B**. Table listing the total number of gene families in each of the *Methanobrevibacter* genomes.

The number of gene families in the core-genome corresponds with the phylogenetic relationship between the organisms, as does the number of gene families specific to each genome. The number of gene families specific to JMR01 is high, with 649 (30.6%) of its total gene families being specific to its genome (Figure 6.16A.).

Analysis of the Methanobrevibacter core genome. The number of gene families in the core genome is summarised in Figure 6.17 and Table 6.14. A full list of core genome gene families is presented in the Appendix (Table A.6.5.). The core genome lacks the tryptophan and proline biosynthesis operons. The JMR01 genome is predicted to lack tryptophan biosynthesis genes, and all *Methanobrevibacter* genomes are predicted to lack proline biosynthesis genes.

Table 6.14. Conserved and novel genes of *Methanobrevibacter* genomes by functional categories

COG	Code	Gene families						
		D5 Total	Core	Rumen	*Mbb. gottschalkii* clade	*Mbb. ruminantium* clade*	*Mbb. wolinii* clade*	D5 novel
Translation	J	149	118	0	0	1	0	1
RNA processing and modification	A	0	0	0	0	0	0	0
Transcription	K	106	32	0	0	3	3	5
Replication, recombination and repair	L	124	37	0	0	5	4	14
Chromatin structure and dynamics	B	2	1	0	0	0	0	0
Cell cycle control, mitosis and meiosis	D	10	4	0	0	0	1	4
Nuclear structure	Y	0	0	0	0	0	0	0
Defense mechanisms	V	39	2	0	0	1	3	9
Signal transduction mechanisms	T	26	2	0	0	0	3	4
Cell wall/membrane biogenesis	M	106	24	1	1	3	4	26
Cell motility	N	4	0	0	0	0	0	0
Cytoskeleton	Z	0	0	0	0	0	0	0
Extracellular structures	W	0	0	0	0	0	0	0
Intracellular trafficking and secretion	U	11	7	0	0	0	0	0
Post-translational modification, protein turnover, chaperones	O	66	23	0	0	4	3	12
Energy production and conversion	C	173	85	1	1	4	4	9
Carbohydrate transport and metabolism	G	51	27	0	1	0	1	2
Amino acid transport and metabolism	E	126	64	0	0	4	9	7
Nucleotide transport and metabolism	F	51	35	0	0	0	2	0
Coenzyme transport and metabolism	H	116	56	0	0	17	3	2
Lipid transport and metabolism	I	32	10	0	1	1	1	5
Inorganic ion transport and metabolism	P	59	21	0	0	3	7	2
Secondary metabolites biosynthesis, transport and catabolism	Q	23	2	0	0	0	0	2
General function prediction only	R	262	90	0	1	16	29	28
Function unknown	S	223	87	0	2	11	16	23
Unclassified COG	Unclassified COG	283	21	0	4	32	33	74
Not in COGs	-	445	16	0	4	66	42	215

*The gene families in the *Mbb. ruminantium* clade used $M1^T$ as reference for COG identification, and the gene families in the *Mbb. wolinii* clade used AbM4 as reference.

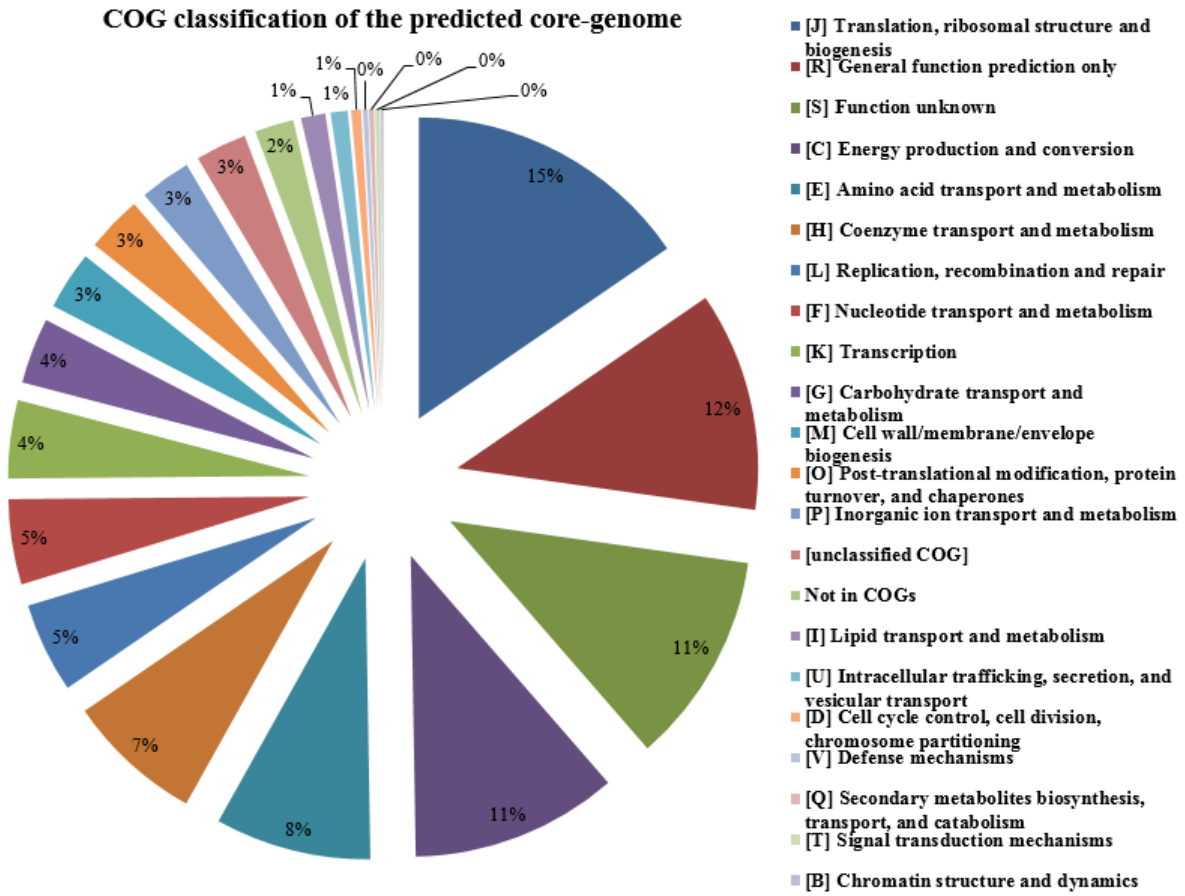


Figure 6.17. Chart displaying the COG classification of the predicted core-genome for members of the genus *Methanobrevibacter*. COG categories that do not contain a gene family in the core-genome are not displayed.

Adaptation to the rumen environment. Two gene families are uniquely conserved between all 11 *Methanobrevibacter* spp. sourced from the rumen (Table 6.15.).

Table 6.15. *Methanobrevibacter* gene families uniquely conserved in the rumen by functional category

Locus_tag	Predicted gene product	COG category
D5_0505	Nitroreductase family protein	[C]
D5_1307	Glycosyl transferase GT2 family	[M]

*only one gene from D5 for each gene family is represented

Uniquely conserved gene families in the genomes of Mbb. gottschalkii clade. Amongst the 15 gene families conserved within the *Mbb. gottschalkii* clade, 10 are classified as poorly characterised or not in COG (Table 6.16.).

Table 6.16. Gene families uniquely conserved in the *Mbb. gottschalkii* clade by functional category

Locus_tag	Predicted gene product	COG category
D5_0388	hypothetical protein	[unclassified]
D5_0628	conserved hypothetical transmembrane protein	[unclassified]
D5_0738	formate dehydrogenase alpha subunit, flpA	[C]
D5_0824	NADPH-dependent FMN reductase	[R]
D5_0940	hypothetical transmembrane protein	[S]
D5_1099	ATP-grasp domain containing protein	[I]
D5_1317	hypothetical transmembrane protein	[unclassified]
D5_1344	conserved hypothetical transmembrane protein	[unclassified]
D5_1640	hypothetical protein	[M]
D5_1755	hypothetical protein	Not in COGs
D5_1759	MFS transporter	[G]
D5_1853	hypothetical protein	Not in COGs
D5_1895	hypothetical protein	Not in COGs
D5_1990	hypothetical protein	[S]
D5_2064	hypothetical protein	Not in COGs

*only one gene from D5 for each gene family is represented

Uniquely conserved gene families in the genomes of the Mbb. ruminantium clade. Amongst the 171 gene families conserved within the *Mbb. ruminantium* clade, 125 are either classified as poorly characterised or not in COG, and 17 are classified under coenzyme transport and metabolism (Tables 6.14., A.6.6.), including genes encoding a 6-carboxyhexanoate—CoA ligase BioW (MRU_RS10255, 2595159567, YLM1_1498), 8-amino-7-oxonoanoate synthase BioF (MRU_RS10260, 2595159566, YLM1_1499) and adenosylmethionine--8-amino-7-oxononanoate aminotransferase BioA (MRU_RS10475, 2595159351, YLM1_1816), which are involved in biotin biosynthesis using pimelate as a precursor.

Uniquely conserved gene families in the genomes of the Mbb. wolinii clade. Amongst the 168 gene families conserved within the *Mbb. wolinii* clade, 120 are either classified as poorly characterised or not in COG, and nine gene families are classified under amino acid transport and metabolism (Tables 6.14., A.6.7.), including five genes involved in tryptophan biosynthesis *trpACDEF*. Despite 17 *Methanobrevibacter* genomes possessing a complete tryptophan biosynthesis operon, the *trpACDEF* genes in members of *Mbb. wolinii* clade are different to the *trpACDEF* genes in 14 other *Methanobrevibacter* genomes analysed. Using *trpC* and *trpD* genes as examples, *trpC* gene has an average of 41.4% aa identity between *Mbb. wolinii* clade and other *Methanobrevibacter* spp., *trpD* gene has an average of 39.1% aa identity between *Mbb. wolinii* clade and other *Methanobrevibacter* spp., whereas the *trpC* genes and *trpD* genes are similar when compared within *Mbb. wolinii* clade (79.7%, 84% aa respectively) and compared within members outside of the *Mbb. wolinii* clade (67%, 71% aa respectively). In members of *Mbb. wolinii* clade, *trpCD* genes shares a high homology to members of the genera *Methanosphaera* and *Methanobacterium*, while *trpCD* genes in other *Methanobrevibacter* genomes share a high homology to members of the phylum *Firmicutes* and the genus *Clostridium*. Four gene families are classified under cell

wall/membrane/envelope biogenesis, including a gene encoding aspartate racemase (2540853571, 2558933547, 2553938151) involved in conversion of L-aspartate to D-aspartate, this gene not predicted outside of *Mbb. wolinii* clade.

Novel gene families in the Methanobrevibacter sp. D5 genome. The analysis of gene family conservation identified 444 unique gene families in *Methanobrevibacter* sp. D5 (Table A.6.8.), of which, 342 are classified as poorly characterised, 26 genes are classified as cell wall/membrane/envelope biogenesis genes, including six glycosyl transferase genes (D5_0223, D5_0462, D5_0463, D5_0484, D5_0488, D5_1305) that may be involved in exopolysaccharide formation and four adhesin-like protein encoding genes (D5_0459, D5_0965, D5_1473, D5_2175). In addition, 46 other adhesin-like protein encoding genes were also identified as unique. Two genes predicted to be an endonuclease (D5_2164) and a methylase (D5_2165) of a type III restriction system are considered novel under the criteria used in this study. A non-ribosomal surfactin synthetase (D5_0482) was also identified as unique.

6.2.6. Metabolic pathway construction of *Methanobrevibacter* sp. D5

D5 genes with a predicted metabolic function are listed in Table A.6.1. A total of 1,232 genes, representing 49.5% of the genome, have been assigned to a metabolic pathway.

Energy metabolism

Methanogenesis. Based on the genome sequence, methanogenesis is the sole energy generating mechanism of *Methanobrevibacter* sp. D5. The genes predicted to be involved in methanogenesis are summarized in Figure 6.18 and Table A.6.1. The D5 genome encodes genes required for energy generation via hydrogenotrophic methanogenesis (Table A.6.1.), including the formate transporter FdhC (D5_2393), formate dehydrogenase FdhAB (D5_2391, D5_2392), FlpAB (D5_0738, D5_0739) and its accessory protein FdhD (D5_02394) and FlpD (D5_0772), suggesting D5 can utilise CO_2, H_2 and formate as substrates (Figure 6.18.).

The H_2 dependent fixation of CO_2 is carried out by a tungsten containing formylmethanofuran dehydrogenase Fwd, encoded by *fwd*HFGDBAC (D5_2382 – D5_2388) and *fwd*E (D5_1080, D5_1751) genes. Fwd binds CO_2 to a MF and reduces it to formyl-MF using the reducing potential supplied from the membrane bound energy conserving [NiFe]-hydrogenase EhaA (D5_1770 – D5_1787) with H_2 via Fdx (Figure 6.18.).

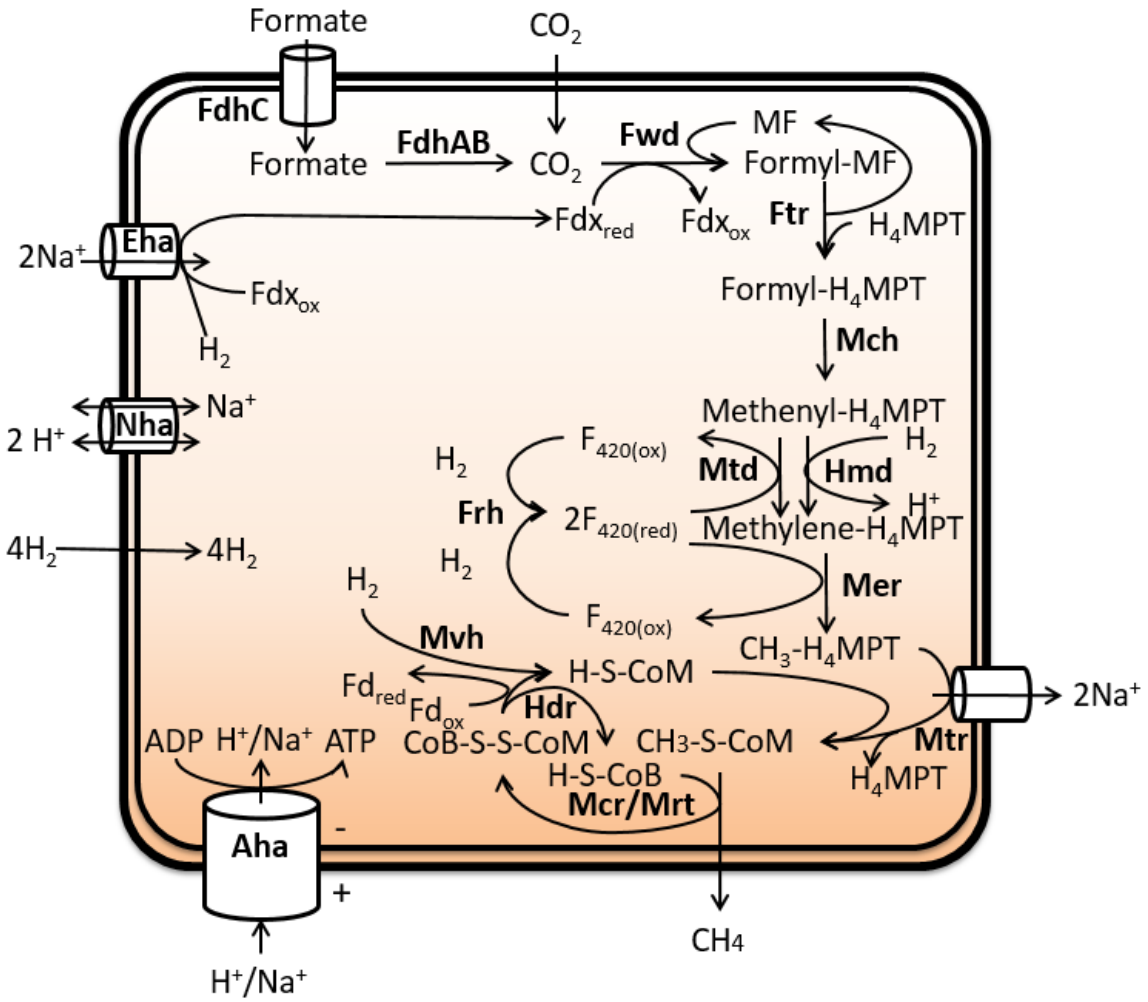


Figure 6.18. Hydrogenotrophic methanogenesis pathway of *Methanobrevibacter* sp. D5. Enzyme names are displayed in bold. The enzymes involved are: formate transporter (FdhC), formate dehydrogenase (FdhAB), formylmethanofuran dehydrogenase (Fwd), formylmethanofuran- H_4MPT formyl transferase (Ftr). N^5, N^{10}-methenyl-H_4MPT cyclohydrolase (Mch), F_{420}-dependent methylene-H_4MPT dehydrogenase (Mtd), F_{420} reducing hydrogenase (Frh), methylene-H_4MPT reductase (Mer), H_4MPT S-methyltransferase (Mtr), H_2-dependent methylene-H_4MPT dehydrogenase (Hmd), Methyl-CoM reductase (Mcr/Mrt), heterodisulfide reductase (Hdr), methyl-viologen hydrogenase (Mvh). The cofactors involved are: methanofuran (MF), ferredoxin (Fdx), tetrahydromethanopterin (H_4MPT), coenzyme M (CoM), coenzyme B (CoB), and cofactor F_{420} (F_{420}). Components involved in membrane potential generation and energy generation are energy conserving hydrogenase (Eha), Na^+/H^+ antiporter (Nha), H_4MPT *S*-methyltransferase (Mtr) and A_1A_0 ATP synthase (Aha).

The formyl group is then transferred to the H_4MPT by formylmethanofuran-H_4MPT formyl transferase Ftr (D5_0210, D5_1769), producing formyl-H_4MPT. The formyl-H_4MPT is reversibly catalysed to N^5, N^{10}-methenyl-H_4MPT by N^5, N^{10}-methenyl-H_4MPT cyclohydrolase Mch (D5_1626). Methenyl-H_4MPT is reduced to Methylene-H_4MPT by either F_{420}-dependent methylene-H_4MPT dehydrogenase Mtd (D5_0079, D5_0741) or H_2-dependent methylene-H_4MPT dehydrogenase Hmd (D5_2132). Methylene-H_4MPT is subsequently reduced to methyl-H_4MPT by Methylene-H_4MPT reductase Mer (D5_2110), this reaction is reversible

and F_{420} dependent. The F_{420} hydrogenase Frh encoded by *frh*ADGB (D5_0174 – D5_0177) is responsible for reducing F_{420} using H_2, supplying reducing potential for methanogenesis. The methyl group is transferred from methyl-H_4MPT to CoM via H_4MPT S-methyltransferase complex Mtr (D5_0289 – D5_0296), producing methyl-CoM. Mtr is coupled to the generation of a Na^+ gradient across the cytoplasmic membrane, and is one of the enzyme complexes responsible for the generation of a membrane potential. An additional copy of *mtrA* (D5_0510) and *mtrH* (D5_0350) genes were identified. The reduction of methyl-CoM to CH_4 is carried out by the methyl-CoM reductase enzyme complex; two isoenzymes are present in the D5 genome, a methy-CoM reductase I Mcr (D5_0284 – D5_0288) and a methyl-CoM reductase II Mrt (D5_0742 – D5_0745).

Methanogenesis cofactors. In the hydrogenotrophic methanogenesis pathway, three cofactors act as the C_1 moiety is transferred sequentially from CO_2 to MF, H_4MPT and CoM, and reduced to CH_4. The cofactors involved in the last step of methanogenesis are the same as described in ISO4-H5, where the F_{430} acts as the nickel-porphinoid group central to Mcr/Mrt catalytic activity, and CoB is used by the Mcr and Mrt complex for methyl-CoM reduction to produce CH_4 (Figure 6.18.). This process utilises the reducing potential carried by HS-CoB and generates CoM-S-S-CoB (heterodisulfide). Both cofactors need to be recycled into an active state, and this is carried out by a heterodisulfide reductase (Hdr). D5 encodes two sets of Hdr genes *hdr*ABC (D5_0649 - 0650, D5_1502 – 1504, D5_2467) with an additional copy of *hdr*B (D5_1005) gene. D5 encodes the methyl viologen hydrogenase Mvh required for coupling H_2 to provide reducing potential for the reducing CoM-S-S-CoB and Fdx via electron bifurcation MvhDGAB (D5_0302 – 0305) (Figure 6.18.). A second cofactor involved in the coupling of H_2 to provide reducing potential is F_{420}, Mtd and Mer depend on $F_{420}H_2$ as the source of reductant.

Generation of membrane potential from methanogenesis and potential sources of reducing potential driving methanogenesis. ATP synthase in *Methanobrevibacter* spp. can be driven by either Na^+ or H^+ gradient. Methanogenesis is coupled to the generation of a Na^+ gradient, the conversion of methyl-H_4MPT to methyl-CoM by Mtr (D5_0289 – 0296) is coupled to the translocation of Na^+ across the cell membrane, thereby generating a Na^+ gradient. The first step of hydrogenotrophic methanogenesis is carried out by a Fwd, which utilises the reducing potential supplied from reduced Fdx (Fdx_{red}), the reduction of Fdx is catalysed by membrane associated energy conserving hydrogenase A complex (EhaA) (D5_1770 – D5_1787) driven by a Na^+ motive force in the hydrogenotrophic methanogens. A second energy conserving

hydrogenase complex (EhaB) (D5_0225 – 0241) was identified in the D5 genome, which is associated with regeneration of electron carriers involved in autotrophic CO_2 fixation and acetate assimilation. Na^+/H^+ antiporters (Nha) (D5_0516, D5_1713) are also involved in ion gradient formation, they are responsible for coupling and cycling of Na^+, H^+ and pH homeostasis. Nha could compete with the ATP synthase for Na^+. Frh and $F_{420}H_2$ supplies reducing potential to methanogenesis independently of membrane potential. In addition, a gene encoding NADP-dependent F_{420} reductase, NpdG (D5_1461), has been predicted in the D5 genome, as well as two alcohol dehydrogenase genes (D5_1302, D5_1857) that may work in conjunction with NpdG to supply reducing potential.

Methanogenesis marker proteins. Homologues of methanogenesis marker proteins 1 through 17 are present within the D5 and all of the *Methanobrevibacter* spp. genomes analysed.

Central carbon metabolism

D5 is predicted to encode the genes required for a partial TCA cycle, gluconeogenesis and the pentose phosphate pathway.

One-carbon metabolism. D5 has a complement of genes allowing it to utilise acetate, bicarbonate, formate and CO_2 as a potential carbon source (Figure 6.9.). D5 encodes two acetyl-CoA synthetase, Acs, a SSS family solute/Na^+ symporter homologous to an acetate permease, ActP, as well as a holoenzyme of pyruvate synthase, PorABCD (Table A.6.1.) that allows the utilisation of acetate and the subsequent production of acetyl-CoA and pyruvate. Por utilises a Fdx_{red} as the source of its reducing potential.

Formate is involved in hydrogenotrophic methanogenesis as well as central carbon metabolism. D5 encodes a pyruvate formate-lyase, Pfl, and the associated Pfl-activating protein, allowing the conversion of formate to pyruvate. In addition, D5 is predicted to possess a set of bicarbonate ABC transporter genes, *btcABC*, allowing bicarbonate to be utilised as a carbon source by Por, or interconverted to CO_2 by carbonic anhydrase, Cab.

TCA cycle. The D5 genome is predicted to have an incomplete reductive TCA cycle from oxaloacetate to 2-oxoglutarate (Figure 6.19., Table A.6.1.). D5 has the 2-oxoglutarate Fdx oxidoreductase *kor*DABC genes required to convert succinyl-CoA to 2-oxoglutarate, this reaction utilises Fdx_{red} as the source of its reducing potential. The Fdxs associated with Kor and Por are reduced by the energy conserving hydrogenase B (EchB).

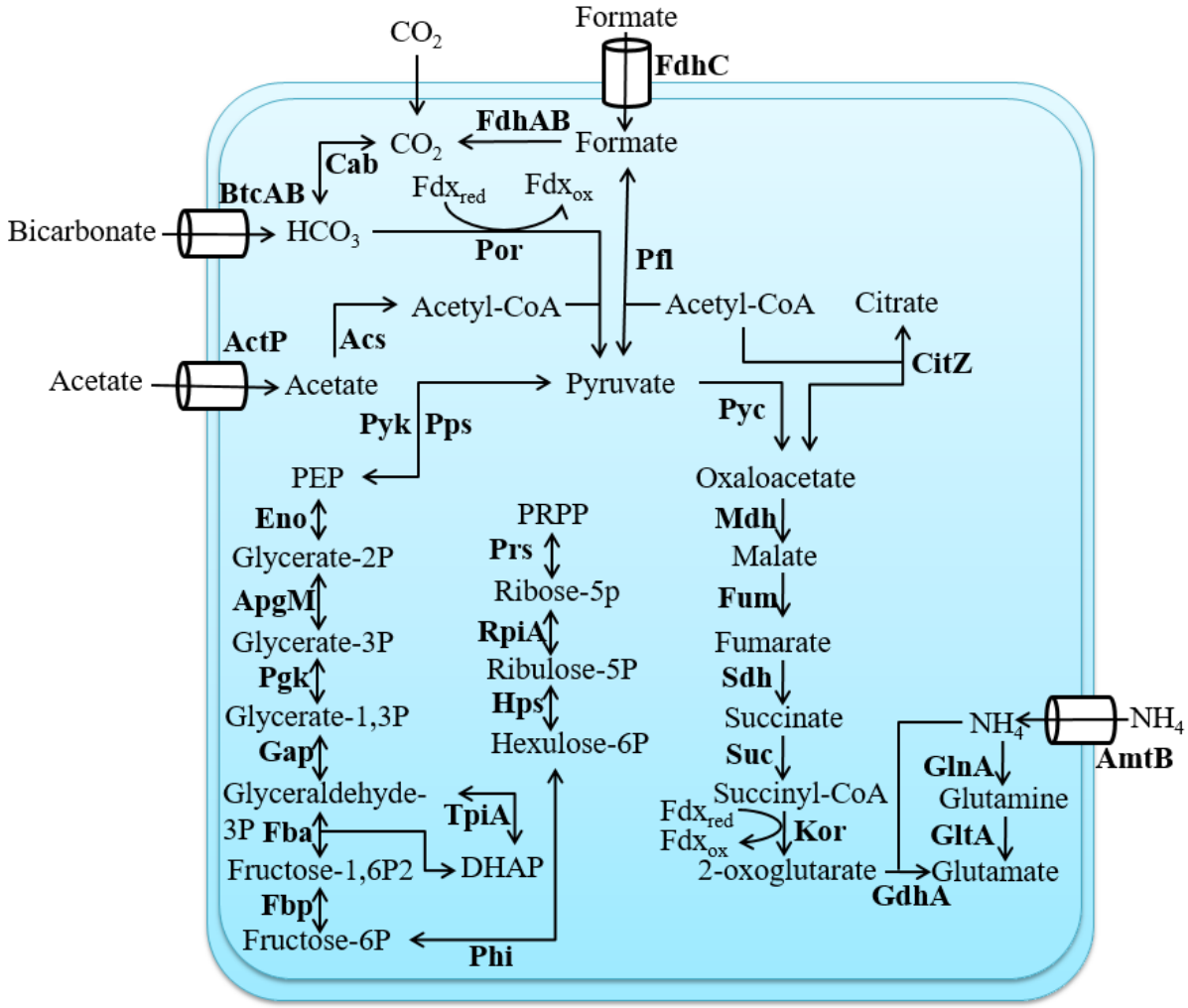


Figure 6.19. Central carbon metabolism of *Methanobrevibacter* sp. D5. Enzymes are displayed in bold. The enzymes involved are: bicarbonate ABC transporter (BtcAB), carbonic anhydrase (Cab), formate/nitrite transporter (FdhC), formate dehydrogenase (FdhAB), ferredoxin (Fdx), pyruvate Fdx oxidoreductase (Por), pyruvate formate-lyase (Pfl), cation/acetate symporter (ActP), acetyl-CoA synthetase (Acs), pyruvate kinase (Pyk), Phosphoenolpyruvate synthase (Pps), enolase (Eno), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (ApgM), phosphoglycerate kinase (Pgk), glyceraldehyde3-phophsate dehydrogenase (Gap), fructose-bisphosphate aldolase (Fba), triose-phosphate isomerase (TpiA), fructose 1,6-bisphosphatase (Fbp), 3-hexulose-6-phosphate isomerase (Phi), 3-hexulose-6-phosphate synthase (Hps), ribose 5-phosphate isomerase A (RpiA), ribose-phosphate diphosphokinase (Prs), pyruvate carboxylase (Pyc), malate dehydrogenase (Mdh), citrate synthase (CitZ), fumarate hydratase (Fum), succinate dehydrogenase (Sdh), succinyl-CoA synthetase (Suc), 2-oxoglutarate Fdx oxidoreductase (Kor), glutamate dehydrogenase (GdhA), glutamine synthase (GlnA), glutamate synthase (GltA) and ammonium transporter (AmtB). The compounds involved are phosphoeol pyruvate (PEP), dihydroxyacetone phosphate (DHAP) and phosphoribosyl pyrophosphate (PRPP).

Gluconeogenesis. D5 encodes all of the genes required for gluconeogenesis from pyruvate to fructose-6-phosphate (Figure 6.19., Table A.6.1.).

Pentose phosphate pathway. The D5 genome encodes all the genes required for a non-oxidative pentose phosphate pathway and phosphoribosyl pyrophosphate (PRPP) production (Figure 6.19., Table A.6.1.). Two homologues of the 3-hexulose-6-phosphate isomerase (Phi) were found.

Amino acid biosynthesis

The D5 is predicted to be able to make 18 amino acids based on the genes present in the genome and the genes involved are listed (Figure 6.20., Table A.6.1.). D5 is predicted to be unable to produce proline and threonine. D5 encodes an amino acid carrier protein (D5_1090) and an amino acid ABC transporter (D5_1294 – 1296) that may be able to facilitate uptake of these amino acids to overcome these auxotrophies.

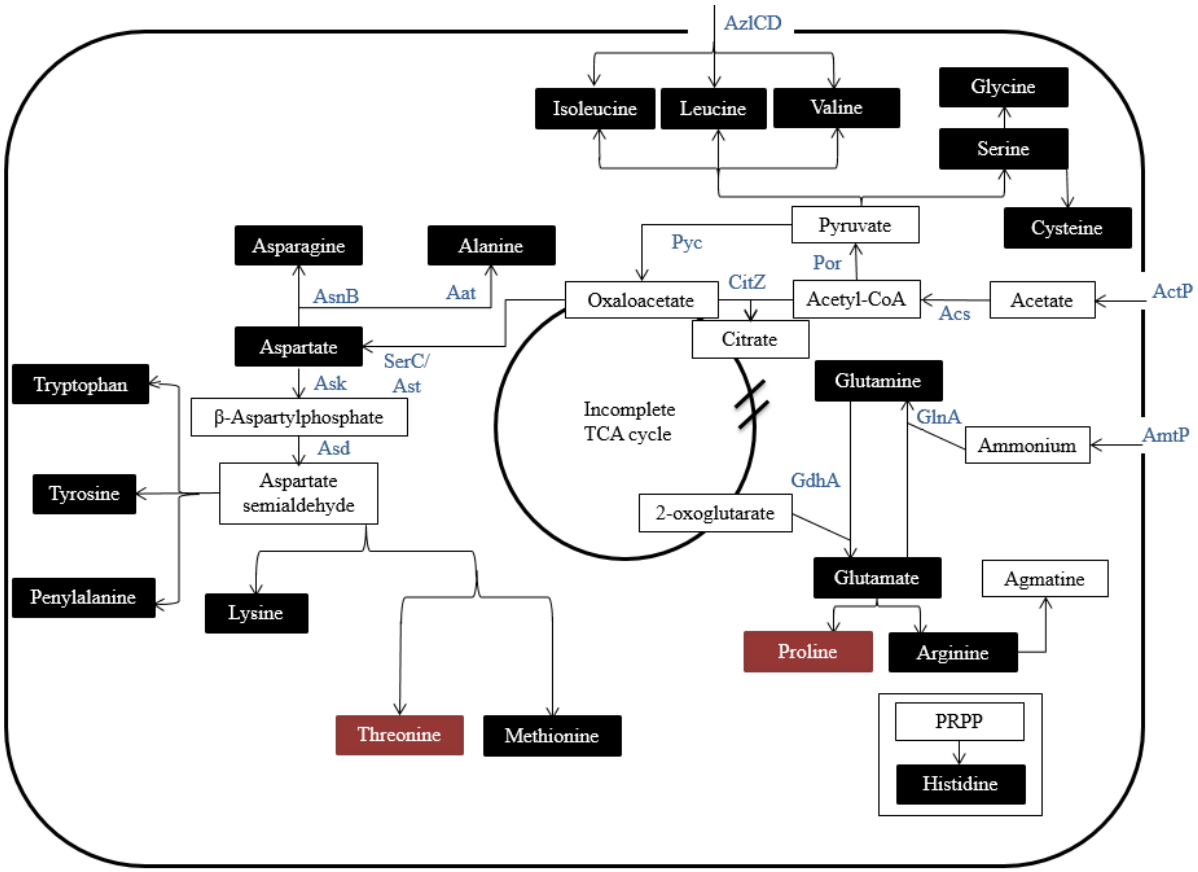


Figure 6.20. Amino acid biosynthesis of *Methanobrevibacter* sp. D5. Amino acids predicted to be synthesized are displayed in black text boxes. Amino acids predicted not to be synthesized are displayed in red text boxes. Intermediates and molecules are displayed in white text box. The enzymes involved are displayed in blue text, including citrate synthase (CitZ), pyruvate:Fdx oxidoreducta (Por), acetyl-CoA synthase (Acs), pyruvate carboxylase (Pyc), glutamine synthetase (GlnA), : glutamate dehydrogenase (GdhA), ammonium transporter (AmtP), cation/acetate symporter (ActP), branched chain amino acid transporter (AzlCD), asparagine synthase (AsnB), alanine aminotransferase (Aat), broad substrate specificity phosphoserine aminotransferase (SerC), aspartate amino transferase (Ast), aspartate kinase (Asp), aspartate semialdehyde dehydrogenase (Asd). Phosphoribosylpyrophosphate is abbreviated PRPP.

Proline. The genes required for proline biosynthesis, *proABC*, were not found in the D5 genome.

Threonine. Threonine is produced from aspartate via a homoserine intermediate. The D5 genome is predicted to lack the *thrB* gene, which encodes the homoserine kinase required to

convert homoserine to *O*-phospho-L-homoserine, the substrate of threonine synthase, ThrC (D5_1657). Therefore, D5 is predicted to be incapable of threonine biosynthesis.

Vitamins and cofactors

F_{430}. As described in Chapter 3, Section 3.2.4, uroporphyrinogen is the prerequisite to tetrapyrole cofactors and cobalamin. The D5 genome possesses the full complement of genes required to produce uroporphyrinogen (Table A.6.1.), including glutamyl-tRNA reductase, HemA (D5_0334), delta-aminolevulinic acid dehydratase, HemB (D5_0789), porphobilinogen deaminase, HemC (D5_0519), uroporphyrinogen III synthase, HemD (D5_0819) and glutamate-1-semialdehyde-2,1-aminomutase, HemL (D5_0045). The only identified enzyme in the pathway of F_{430} biosynthesis, uroporphyrin-III C-methyltransferase, CorA (D5_0865), is also present. No operon arrangements were observed.

Cobalamin/thiamine. The sirohydrochlorin intermediate produced from F_{430} biosynthesis can be used for cobalamin (vitamin B_{12}) biosynthesis (Table A.6.1.). The D5 genome is predicted to encode all but 2 genes required for the anaerobic pathway of cobalamin biosynthesis, the missing genes are *cobC* and cobirinic acid-a,c-diamide reductase. D5 is also predicted to encode genes required for thiamine biosynthesis, including two copies of *thiC* and *thiD* genes (Table A.6.1.).

Riboflavin/FMN/FAD. The D5 genome is predicted to encode the full complement of genes required to produce riboflavin (Table A.6.1.).

MF/H_4MPT/CoM/CoB/F_{420}/F_{390}. There are five important coenzymes involved in methanogenesis, MF, H_4MPT, CoM, CoB and F_{420}. The D5 genome is predicted to possess the full complement of genes required to produce CoM, CoB and F_{420} (Table A.6.1.). D5 is predicted to encode all except the 2-furaldehyde phosphate aminotransferase *mtnC* gene required for MF biosynthesis (Table A.6.1.). The D5 genome is also predicted to possess eight genes involved in the biosynthesis of H_4MPT, however the genes involved in the later part of the biosynthetic pathway remain to be identified. Cofactor F_{390} (F_{390}) is produced from F_{420} when the methanogen is exposed to O_2. The D5 genome is predicted to possess two *ftsA* genes encoding F_{390} synthetase (D5_1079, D5_1085).

Ubiquinone. The D5 genome encodes five genes predicted to be involved in the ubiquinone biosynthesis, including 3-polyprenyl-4-hydroxybenzoate decarboxylase UbiX, three copies of

ubiB genes encoding 2-polyprenylphenol 6-hydroxylase and a UbiD family decarboxylase gene, however the function of these genes remain to be elucidated.

Nicotinate. The D5 genome has all the genes encoding the enzymes required for nicotinate biosynthesis (Table A.6.1.).

Biotin. The D5 genome is predicted to contain three genes involved in biotin synthesis, two biotin synthases, BioB (D5_1332, D5_2133) and a biotin-acetyl-CoA-carboxylase ligase, BirA (D5_0677), no others genes required for biotin biosynthesis were identified.

Glutathione. The D5 genome is predicted to encode a gamma-glutamylcysteine synthetase *gshA* gene and a bifunctional glutamate-cysteine ligase/glutathione synthetase *gshF* gene required for glutathione biosynthesis, and two copies of glutathione peroxidase *gpxA* genes and glutahionie-disulfide reductase *gor* genes (Table A.6.1.).

Nitrogen metabolism

D5 is predicted to encode two copies of the ammonium transporter *amtB* gene (D5_0957, D5_1675). Ammonium uptake is linked to glutamine and glutamate biosynthesis by glutamine synthase *glnA* (D5_2378) and glutamate dehydrogenase *gdhA* (D5_0429) as depicted in Figure 6.19. The D5 genome also predicts two genes involved in regulation of nitrogen assimilation, a gene encoding the transcriptional repressor adjacent to *glnA*, *nrpR* (D5_2379), and two nitrogen regulatory protein P-II *glnK* genes (D5_0958, D5_1674). The D5 genome is predicted to lack genes encoding nitrogenase, but D5 has several other proteins predicted to be involved in nitrogen assimilation (Table A.6.1.), including a nitrogenase cofactor biosynthesis protein, NifB (D5_0636), a dinitrogenase iron-molybdenum cofactor biosynthesis protein (D5_2272), a cysteine desulfurase, NifS (D5_1698), and a hydroxylamine reductase, Hcp (D5_0749).

Lipids and cell envelope

The D5 genome has a full complement of genes required to produce archaeol via the mevalonate pathway (Table A.6.1.), the proposed *de novo* phospholipid biosynthesis pathway is displayed in Figure 6.21.

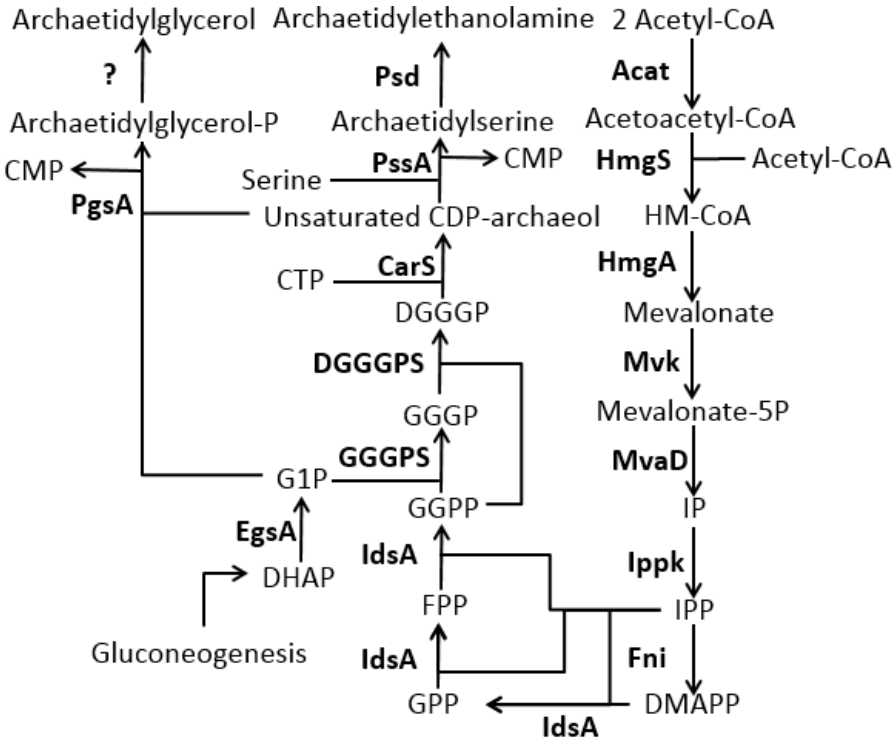


Figure 6.21. Predicted phospholipid biosynthesis pathway in *Methanobrevibacter* sp. D5. Enzymes are displayed in bold. The enzymes involved are: acetyl-CoA acetyltransferase (Acat), hydroxymethylglutaryl-CoA synthase (HmgS), hydroxymethylglutaryl-CoA reductase (HmgA), mevalonate kinase (Mvk), phosphomevalonate decarboxylase (MvaD), isopentenyl diphosphate kinase (Ippk), isopentenyl-diphosphate δ-isomerase (Fni), bifunctional short chain isoprenyl diphosphate synthase (IdsA), NAD(P)-dependent glycerol-1-phosphate dehydrogenase (EgsA), geranylgeranylglyceryl phosphate synthase (GGGPS), digeranylgeranylglyceryl phosphate synthase (DGGGPS), CDP-archaeol synthase (CarS), CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (PgsA), phosphatidylserine synthase (PssA), phosphatidylserine decarboxylase (Psd). The compounds involved are hydroxymethylglutaryl-CoA (HM-CoA), isopentenyl phosphate (IP), isopentenyl diphosphate (IPP), 3,3-dimethylallyldiphosphate (DMAPP), geranyl diphosphate (GPP), farnesyl diphosphate (FPP), geranylgeranyl diphosphate (GGPP), dihydroxyacetone phosphate (DHAP), glycerol-1-phosphate (G1P), geranylgeranylglyceryl phosphate (GGGP), digeranylgeranylglyceryl phosphate (DGGGP), cytidine triphosphate (CTP), cytidine monophosphate (CMP).

The biosynthesis of phospholipid utilises acetyl-CoA and D-glyceraldehyde 3-phosphate from central carbon metabolism. Based on the presence of the CDP-archaeol synthase (*carS*) gene, D5 is predicted to use unsaturated CDP-archaeol as an intermediate in phospholipid biosynthesis. The genes in D5 suggest that CDP-archaeol proceeds towards archaetidylethanolamine and archaetidylglycerol biosynthesis.

D5 has several genes required to produce the precursor of pseudomurein, UDP-*N*-acetylglucosamine (UDP-GlcNAc) and UDP-*N*-acetylgalactosamine (UDP-GalNAc) from fructose-6-phosphate (Table A.6.1.), including two copies each of *glmM* and *glmS*. In addition, several genes are predicted to be associated with pentapeptide formation from UDP-glutamic acid and crosslinking to alanine, lysine and glutamic acid, including six genes belonging to the cell wall biosynthesis protein of Mur ligase family (D5_0091, D5_0092, D5_0520, D5_0948,

D5_1536, D5_1822). The UDP-pentapeptide is linked to a disaccharide by the cell wall biosynthesis protein, phospho-*N*-acetylmuramoyl-pentapeptide-transferase, a cell wall biosynthesis protein UDP-glycosyltransferase and a UDP-*N*-acetylglucosamine diphosphorylase/glucosamine-1-phosphate *N*-acetyltransferase (GlmU) (D5_1823, D5_2237, D5_2238). D5 also has genes involved in teichoic acid and sialic acid biosynthesis (Table A.6.1.), which may be crosslinked to the pseudomurein.

Nucleic acid metabolism

The D5 genome is predicted to possess all the genes required for pyrimidine biosynthesis, but two genes are missing to complete the full complement of genes required for *de novo* purine biosynthesis. The proposed pathway of *de novo* biosynthesis is illustrated in Figure 6.22. The two absent genes from purine biosynthesis are phosphoribosylglycinamide formyltransferase (*purT*) gene whose corresponding enzyme converts 5-phospho-β-D-ribosyl) glycinamide (GAR) to formyl-(5-phospho-β-D-ribosyl)glycinamide (FGAR), and the gene (*gmk*) encoding guanylate kinase that phosphorylates GMP to GDP.

D5 is also predicted to possess genes involved in nucleotide salvage and interconversion (Table A.6.1.), including an adenine phosphoribosyltransferase (Apt) gene (D5_2430), two CMP/dCMP deaminase genes (D5_1176 and D5_2402) and a cytidylate kinase (Cmk) gene (D5_0708).

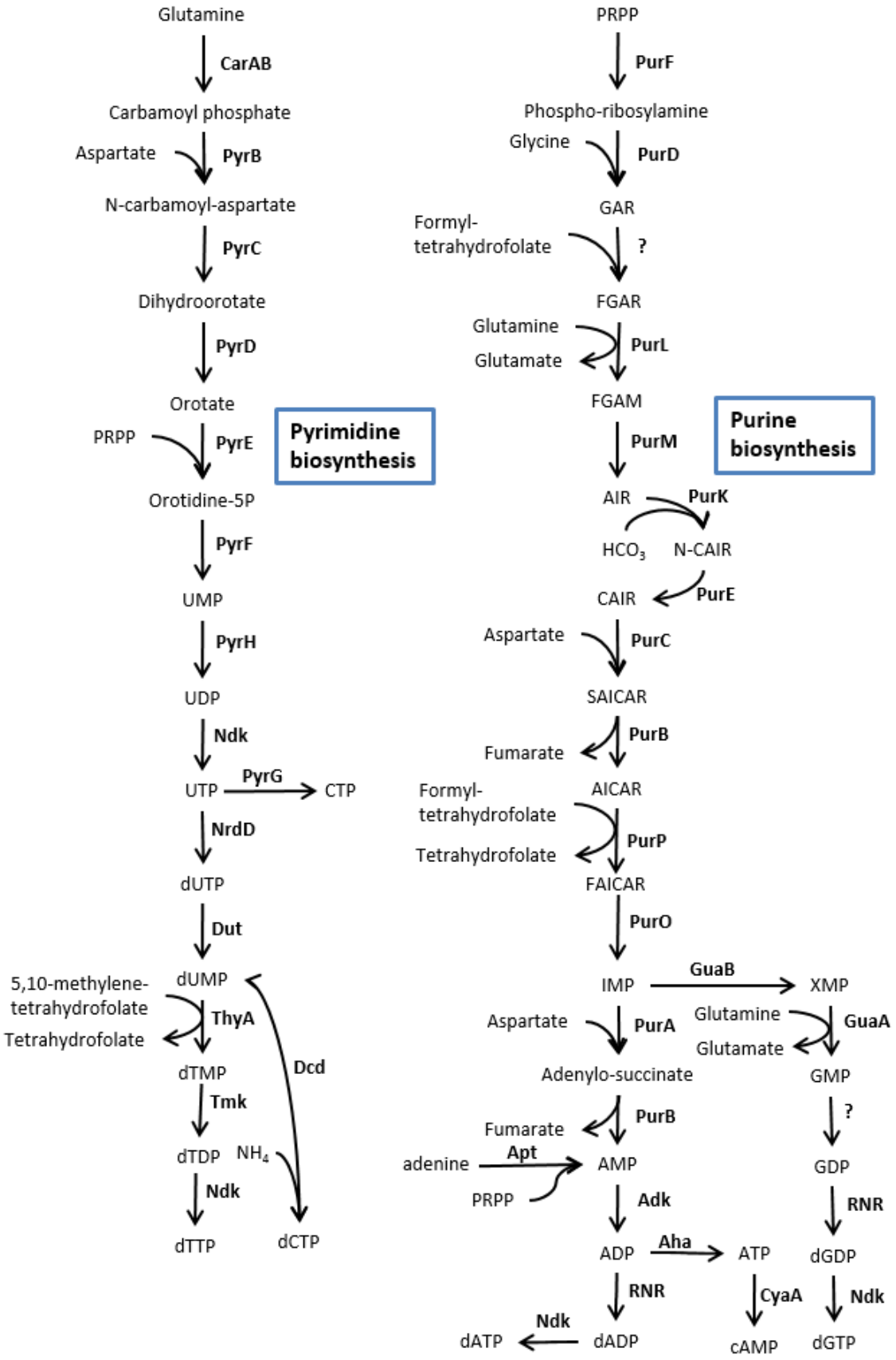

Glutamine
CarAB
Carbamoyl phosphate
Aspartate
PyrB
N-carbamoyl-aspartate
PyrC
Dihydroorotate
PyrD
Orotate
PRPP
PyrE
Orotidine-5P
PyrF
UMP
PyrH
UDP
Ndk
UTP
PyrG
CTP
NrdD
dUTP
Dut
dUMP
5,10-methylene-tetrahydrofolate
ThyA
Tetrahydrofolate
dTMP
Tmk
dTDP
NH4
Ndk
dTTP
Dcd
dCTP
Pyrimidine biosynthesis
PRPP
PurF
Phospho-ribosylamine
Glycine
PurD
GAR
Formyl-tetrahydrofolate
?
FGAR
Glutamine
Glutamate
PurL
FGAM
PurM
AIR
PurK
HCO3
N-CAIR
CAIR
PurE
Aspartate
PurC
SAICAR
Fumarate
PurB
AICAR
Formyl-tetrahydrofolate
Tetrahydrofolate
PurP
FAICAR
PurO
IMP
GuaB
XMP
Aspartate
PurA
Glutamine
Glutamate
GuaA
Adenylo-succinate
GMP
Fumarate
PurB
?
adenine
Apt
AMP
GDP
PRPP
Adk
RNR
ADP
Aha
ATP
dGDP
RNR
CyaA
Ndk
Ndk
dATP
dADP
cAMP
dGTP
Purine biosynthesis

Figure 6.22. *De novo* nucleotide biosynthesis in *Methanobrevibacter* sp. D5. Enzyme names are displayed in bold. Enzymes involved include carbamoyl phosphate synthase (CarAB), aspartate transcarbamylase (PyrB), dihydroorotase (PyrC), dihydroorotate dehydrogenase (PyrD), orotate phosphoribosyltransferase (PyrE), UMP kinase (PyrH), broad substrate specificity nucleoside diphosphate kinase (Ndk), anaerobic ribonucleoside-triphosphate reductase (NrdD), CTP synthase (PyrG), dUTP diphosphatase (Dut), thymidylate synthase (ThyA), thymidylate kinase (Tmk), CMP/dCMP deaminase (Dcd), amidophosphoribosyl transferase (PurF), phosphoribosylamineglycine ligase (PurD), N^5-carboxyaminoimidazole ribonucleotide synthetase (PurK), N^5-carboxyaminoimidazole ribonucleotide mutase (PurE), phosphoribosylformylglycinamidine synthetase (PurL), phosphoribosylformylglycinamide cyclo-ligase (PurM), phosphoribosylaminoimidazole-succinocarboxamide synthase (PurC), bifunctional enzyme adenosuccinate lyase (PurB), bifunctional AICAR transformylase/IMP cyclohydrolase (PurH), adenylosuccinate synthetase (PurA), adenylate kinase (Adk), A_1A_0 ATP synthase (Aha), GMP synthase (GuaAab), IMP dehydrogenase (GuaB), adenine phosphoribosyltransferase (Apt), adenylate kinase (Adk), adenylate cyclase (CyaA), ribonucleotide-diphosphate reductase (RNR). The compounds involved includes phosphoribose pyrophosphate (PRPP), (5-phospho-β-D-ribosyl)glycinamide (GAR), formyl-(5-phospho-β-D-ribosyl)glycinamide (FGAR), formamido-5-phospho-β-D-ribosyl)acetamidine (FGAM), amino-(5-phospho-ribosyl)imidazole (AIR), N^5-carboxyaminoimidazole ribonucleotide (*N*-CAIR), amino-(5-phospho-ribosyl)imidazole-4-carboxylate (CAIR), 5-phosphoribosyl-4-(*N*-succinocarboxamide)-5-aminoimidazole (SAICAR), amino-1-(5-phosphoribosyl)imidazole-4-carboxamide (AICAR). The nucleotides are abbreviated as dYZP: d: deoxy, Y-[A:adenine, C: cytidine, U: uridine, T: thymidine, G: guanine, I: inosine, X: xanthosine], ZP-[MP: monophosphate, DP: diphosphate, TP: triphosphate], cAMP: cyclic AMP.

Cell cycle

The archaeal cell replication process has been described in detail in Chapter 3, Section 3.2.4. The D5 genome is predicted to possess a full complement of genes required for cell replication, encoding two origin of replication *orc1/cdc6* genes (D5_0001, D5_2279) as described previously in Section 6.2.3.

The replication machinery of D5 includes MCM helicase (D5_2038) that unwinds DNA at the origin, replication factor A (D5_2470) and single-stranded DNA binding protein (D5_0815) which stabilizes single stranded DNA. An OB fold nucleic acid binding domain-containing protein (D5_2250) with unknown function may also take part in stabilization of single stranded DNA. Primase initiates DNA synthesis by producing short primers, and two DNA primase genes have been predicted in the D5 genome; DNA primases PriAB (D5_1490, D5_1495) and DnaG (D5_1934). Primers are extended by DNA polymerase family B, PolB (D5_0261, D5_0803) on the leading strand, and DNA polymerase family D, PolD (D5_2404, D5_2507) on the lagging strand. The replication requires a sliding clamp PCNA family protein, Pcn (D5_0164), with assistance of clamp loading replication factor C (D5_0103, D5_0104). The sliding clamp also assists other enzymes to process and join the Okazaki fragments, including flap endonuclease Fen (D5_2373), ribonuclease HII, RnhB (D5_0320), and ATP-dependent DNA ligase, DnlI (D5_2246). Ribonuclease HI (D5_1611) degrades the DNA/RNA hybrid. The DNA integrase/recombinase (D5_2195) homologous to XerD could be involved in replication termination.

The D5 genome is predicted to encode a gene containining a SMC (structural maintenance of chromosome) domain, and a DNA double-strand break repair protein (Rad50) gene (D5_1538) that may be involved in genome segregation. D5 is also predicted to encode DNA topoisomerases I (D5_2361) and VI (D5_0347, D5_0348) that are involved in removing DNA supercoiling from single and double stranded DNA, respectively.

The D5 genome is predicted to encode a tubulin-like cell division protein FtsZ (D5_2227), which carries out cell division by undergoing GTP-dependent polymerization into a Z-ring, with assistance of other proteins, such as the cell division ATPase, MinD (D5_0031) and ParA/MinD-like ATPase (D5_1415), and cell division protein pelota PelA (D5_2239), the Z-ring then drives the cell division. A DNA recombinase gene (D5_2189) homologous to XerD may be involved in termination of replication.

DNA Recombination and repair

The D5 genome contains several genes predicted to encode DNA recombination and repair systems (Table A.6.1.). The archaeal nucleosomes are an irregular mass of genomic DNA, and requires DNA-binding histones to bind and protect the chromosome, regulating gene expressions. Two archaeal histones are predicted in the D5 genome (D5_0005, D5_0593). Histone acetylation and methylation regulates gene expression by controlling the accessibility of the chromosome; a histone acetyltransferase of ELP3 family has been predicted in the D5 genome (D5_0595). In addition, the DNA binding protein, AlbA (D5_0027), and the associated silencing NAD-dependent deacetylase, Sir2 family protein (D5_0215) have been predicted in the D5 genome.

Protein synthesis

Transcription. The D5 genome has a full complement of genes encoding the transcriptional machinery (Table A.6.1.). This includes the TATA-box binding protein (D5_2364), transcription initiation factor, TFIIB (D5_1931, D5_2109), the TFIIE alpha subunit (D5_2230) involved in transcription initiation, the transcription factor S, Tfs (D5_2436), that enhances fidelity, and the transcription elongation factor, NusA-like protein (D5_0415), involved in transcriptional termination. D5 is predicted to possess 12 genes encoding RNA polymerase subunits responsible for transcribing the genetic information onto mRNA (Table A.6.1.).

Maturation of tRNA. The tRNA precursor molecules that are transcribed contain 5' and 3' extensions which need to be removed to produce functional tRNAs. The D5 genome encodes genes required for this function, including RNAse P subunits P29 (*rpp29*, D5_0690), P30 (*rpp30*, D5_1688), P14 (*rpp14*, D5_1687), and RPR2 (*rpr2*, D5_2348), ribonuclease Z (*rnz*, D5_2013), tRNA nucleotidyltransferase Cca (*cca*, D5_1466), tRNA intron endonuclease EndA (*endA*, D5_1660), 2'-5' RNA ligase (*ligT*, D5_1467), $tRNA^{His}$ guanylyltransferase ThgL (*thgL*, D5_1966) and RNA-splicing ligase RtcB (*rtcB*, D5_0920). The aminoacyl-tRNA synthetases are responsible for linking amino acids to corresponding tRNAs for amino acids and these are listed in Table A.6.1.

Maturation of rRNA. The endonuclease and ribonuclease involved in tRNA maturation are also involved in rRNA maturation. In addition, the D5 genome has six genes encoding exosomes subunits (D5_1685, D5_1748), RNA-binding protein Rrp4 (D5_1684), Rrp42 (D5_1682) and Csl4 (D5_2432), exonuclease Rrp41 (D5_1683).

RNA modification. D5 is predicted to possess genes involved in RNA post-transcriptional modifications, including tRNA modification to wyosine, quenosine and archaeosine as shown in Table A.6.1. Post-transcriptional modifications, such as methylation and pseudouridylation, are commonly observed on rRNA, tRNA and snRNAs. The ribonucleoprotein (RNP) H/ACA box RNP complex is responsible for pseudoridylation of nucleotides, and three protein subunits have been predicted in the D5 genome, Nop10p (D5_0169), Cbf5p (D5_0710) and Gar1p (D5_1932). The H/ACA box RNP interacts with ribosomal protein L7Ae (D5_1656). The snoRNA that associates with H/ACA box RNP remains undetected.

Ribosomal proteins. The constituent of the large (50*S*) subunit and small (30*S*) subunit of ribosome are listed within Table A.6.1.

Translation factors. Translation factors are essential in facilitation of translational initiation, peptide elongation and termination (Table A.6.1.). Two modified amino acids have been observed in the translation factors, including deoxyhypusine found in translation initiation factor 5A and diphthamide in elongation factor 2.

Degradation of mRNA. The aforementioned exosome complex not only mediates the maturation of rRNA, but it also facilitates in degradation of defective mRNAs. In addition, a NMD3 family protein (D5_2035) is involved in nonsense-mediated mRNA decay.

NRPS systems. The D5 genome is predicted to encode a large non-ribosomal peptide synthase (NRPS) of 2,605 amino acids (D5_1740), and an even larger non-ribosomal surfactin synthase (NRSS) of 2,966 aa (D5_0482). Adjacent to the NRPS gene is a MatE efflux family protein (D5_1739) that might be involved in export of the NRPS product. Domain analysis of the D5 NRPS reveals two modules, each containing an adenylation domain, a condensation domain and ending with a peptidyl carrier protein domain. The NRSS is predicted with two additional condensation domains near the 3' terminal (Figure 6.23.). The substrate specificity for adenylation domains were predicted from a database of adenylation domains with known substrate, and the predicted D5 NRPS adenylation substrate is phenylalanine for both of the adenylation domains. The predicted D5 NRSS adenylation substrates are proline and pipecolinic acid.

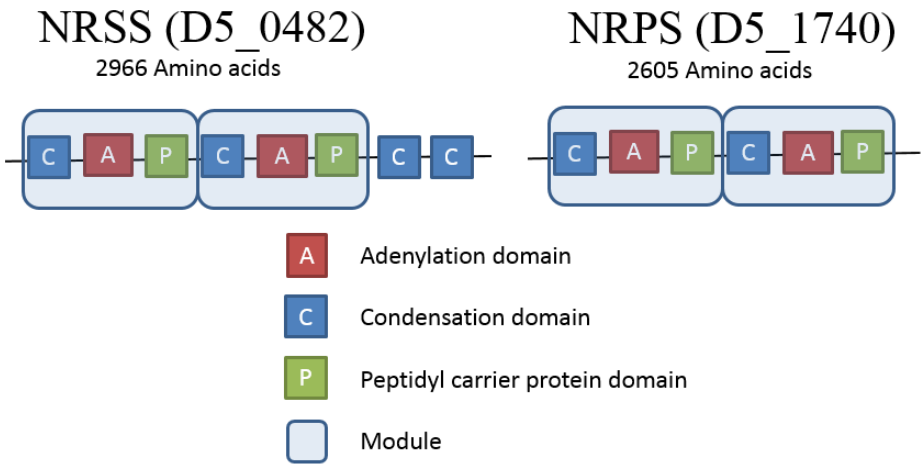


Figure 6.23. Non-ribosomal surfactin synthase and NRPS of D5. The diagram is not drawn to scale.

Protein fate

Secretion. The genes predicted to be involved in protein secretion are summarised in Table 6.17. The secretion mechanism utilised by D5 is similar to the other methanogenic archaea, depending on the *N*-terminus signal sequence and the universally conserved Sec pathway. A gene encoding a sortase family protein (D5_0315) has been predicted in the D5 genome. Sortase acts on secreted proteins with an LPxTG-like sorting motif at the C-terminal and fix the extracellular proteins to the pseudomeurein by cleaving between threonine and glycine, and catalyse amide bond formation between threonine and the cell wall. However, there are no LPxTG/NPxTG/LpxTA/LAxTG motifs predicted amongst the proteins encoded by the D5 genome. Three genes homologous to subunit E and F of type II secretion system have been predicted in the D5 genome.

Table 6.17. D5 genes predicted to be involved in secretion

Locus_tag	Predicted product
D5_0315	sortase family protein
Signal recognition particle	
D5_2338	signal recognition particle receptor FtsY
D5_0815	signal recognition particle SRP19 protein
D5_2438	signal recognition particle SRP54 protein
Sec pore	
D5_0704	preprotein translocase subunit SecY
D5_2226	preprotein translocase subunit SecE
D5_2426	preprotein translocase subunit SecG
D5_2360	oligosaccharyl transferase
Type II secretion system	
D5_1979	Type II secretion system protein F GspF
D5_2197	Type II secretion system protein E GspE
D5_2198	Type II secretion system protein F GspF
Signal peptidase	
D5_0046	signal peptidase I
D5_1565	Type IV leader peptidase family protein
D5_1673	signal peptidase I

Protein folding and degradation. The D5 genome is predicted to encode a variety of chaperones and peptidases (Table A.6.1.). The genes involved in assisting protein folding and denaturing of misfolded proteins are similar to what was described for ISO4-H5 in Chapter 3, Section 3.2.4.

Regulation

The D5 genome is predicted to encode 21 transcriptional regulators, most have unknown functions (Table A.6.1.).

Oxidative stress

Methanobrevibacter sp. D5 is an obligate anaerobic methanogen, and some of its enzymes are highly sensitive to O_2, such as the core enzyme complex of methanogenesis Mcr. Therefore *Methanobrevibacter* sp. D5 must carry genes allowing it to combat oxidative stress. The D5 genome is predicted to possess a full complement of genes allowing D5 to synthesise and utilise glutathione from glutamate, cysteine and glycine (Table 6.18.). Glutathione peroxidase detoxifies H_2O_2 using reduced glutathione as an electron donor. Glutathione also restores oxidised molecules by acting as a H_2 donor. No superoxide dismutases were predicted in the D5 genome. O_2 can be detoxified via $F_{420}H_2$ oxidase, FprA, to H_2O or detoxified to H_2O_2 by desulfoferrodoxin or NADH oxidase. The H_2O_2 is detoxified by rubrerythrin to H_2O.

Table 6.18. D5 genes involved in oxidative stress

Locus_tag	Predicted product
O_2 detoxification	
D5_1478	$F_{420}H_2$ oxidase FprA
D5_2452	$F_{420}H_2$ oxidase FprA
D5_0115	NADH oxidase Nox
D5_1416	NADH oxidase Nox
D5_1696	desulfoferrodoxin Dfx
H_2O_2 detoxification	
D5_1600	rubredoxin Rub1
D5_1601	rubredoxin Rub2
D5_1622	rubrerythrin Rbr1
D5_1623	rubrerythrin Rbr2
D5_1624	rubrerythrin Rbr3
D5_2257	rubredoxin Rub3
D5_2453	rubrerythrin Rbr4
Thioredoxin	
D5_0599	thioredoxin TrxA
D5_1807	thioredoxin-disulfide reductase TrxB
Glutathione	
D5_0781	gamma-glutamylcysteine synthetase GshA
D5_1281	glutathione peroxidase GpxA
D5_1347	glutathione-disulfide reductase Gor
D5_1806	glutathione peroxidase GpxA
D5_2236	glutathione-disulfide reductase Gor
D5_2270	bifunctional glutamate-cysteine ligase/glutathione synthetase GshF

6.2.7. Comparative analysis of methanogenesis pathway in *Methanobrevibacter* spp.

The complete comparison of all metabolic pathways between 18 *Methanobrevibacter* genomes was not possible during the timeframe of this thesis, but a detailed analysis of the hydrogenotrophic methanogenesis pathway was performed. Hydrogenotrophic methanogenesis is the sole energy generating mechanism of the *Methanobrevibacter* spp. analysed. The genes predicted to be involved in methanogenesis are summarized in Figure 6.21 and Table A.6.9.

The observations from culture based experiments has shown all *Methanobrevibacter* spp. utilise CO_2 and H_2 as substrates for methanogenesis (Table 6.1.), this is reflected by the full complement of genes required to carry out hydrogenotrophic methanogenesis in all of the genomes analysed. The H_2-dependent fixation of CO_2 is carried out by a formylmethanofuran dehydrogenase Fwd, which binds CO_2 to a MF and reduces it to a formyl-MF. There are two isozymes of formylmethanofuran dehydrogenase encoded by two gene clusters, the *fwdHFGDACB* operon that is transcribed in the presence of tungstate or molybdate and a *fmdECB* operon that is transcribed in the presence of molybdate. Only the *fwd* operon was identified in the *Methanobrevibacter* spp. genomes. The formyl group on formyl-MF is then

transferred to the H_4MPT by formylmethanofuran-H_4MPT formyl transferase Ftr, producing formyl-H_4MPT. The formyl-H_4MPT is reversibly catalysed to N^5,N^{10}-methenyl-H_4MPT by N^5,N^{10}-methenyl-H_4MPT cyclohydrolase Mch. The JMR01 genome is predicted to possess an additional copy of the *mch* gene. Methenyl-H_4MPT is reduced to methylene-H_4MPT by either F_{420}-dependent methylene-H_4MPT dehydrogenase, Mtd or H_2-dependent methylene-H_4MPT dehydrogenase, Hmd. All of the *Methanobrevibacter* genomes were predicted to possess the *mtd* gene, with an additional copy of the *mtd* gene predicted in the D5, ZA-10^T and SM9 genomes.

The *hmd* gene was not found in the RFM-1^T, RFM-2^T, RFM-3^T, ATM^T and ANOR1 genomes, whereas YE315, ZA-10^T and SM9 genomes are predicted to possess an additional copy of the *hmd* gene. Methylene-H_4MPT is subsequently reduced to methyl-H_4MPT by Methylene-H_4MPT reductase, Mer. Both Mer and Mtd acquire reducing potential from the F_{420} hydrogenase, Frh, encoded by *frh*ADGB, and all *Methanobrevibacter* genomes have the Frh-encoding genes. The methyl group is transferred from methyl-H_4MPT to CoM via H_4MPT S-methyltransferase complex, Mtr, which is coupled to the generation of Na^+ gradient across the cytoplasmic membrane. The methyl-CoM is reduced to CH_4 by methyl-CoM reductase, Mcr/Mrt, all genomes were predicted to possess genes encoding Mcr, whereas the genes encoding Mrt was only predicted in the D5, ZA-10^T, SM9, HO^T, PS^T, JMR01, KM1H5-$1P^T$ and YLM1 genomes.

Some members of the genus *Methanobrevibacter* are capable of utilising formate as a substrate for methanogenesis. Two sets of formate utilising genes has been found in the *Methanobrevibacter* genomes analysed, one set is the *fdhDCAB* found in all *Methanobrevibacter* spp. genomes, the other is the *flpABD* that was found in 11 of the 18 genomes analysed. The *flpBD* genes are present at end of the contig in the HO^T genome without a *flpA* gene (Table A.6.9.).

In the hydrogenotrophic methanogenesis pathway, three cofactors act as the C_1 carrier, the C_1 moiety is transferred sequentially from CO_2 to MF, H_4MPT and CoM, and reduced to CH_4. All *Methanobrevibacter* genomes are predicted to be capable of producing all the cofactors involved in methanogenesis except the $M1^T$ genome (Figure 6.24., Table A.6.9.), which lacks the *comADE* genes required for *de novo* CoM biosynthesis.

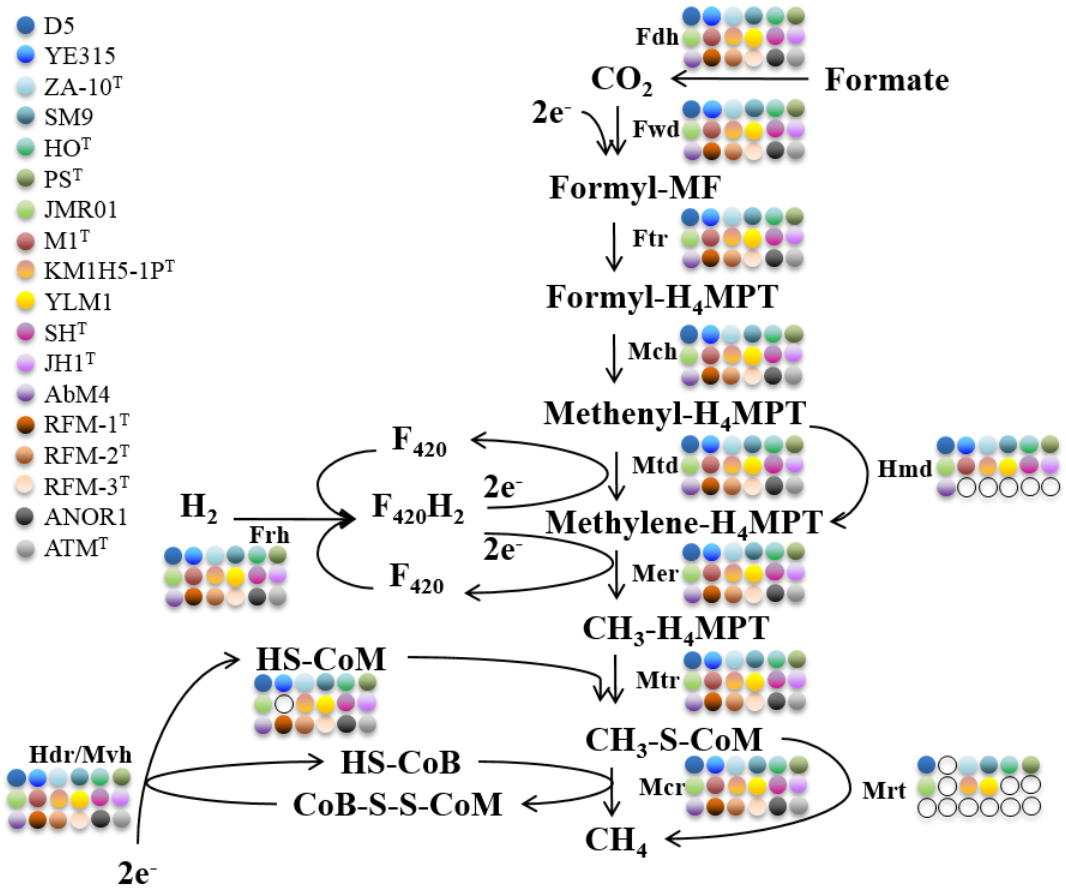


Figure 6.24. Hydrogenotrophic methanogenesis pathway of *Methanobrevibacter* spp. The presence or absence of each gene/set of genes is highlighted by colored circles. A white circle indicates absence of that gene in the corresponding genome. The enzymes involved are: formate dehydrogenase and transporter (Fdh), formylmethanofuran dehydrogenase (Fwd), formylmethanofuran-H_4MPT formyl transferase (Ftr), N^5, N^{10}-methenyl-H_4MPT cyclohydrolase (Mch), F_{420}-dependent methylene-H_4MPT dehydrogenase (Mtd), F_{420} reducing hydrogenase (Frh), methylene-H_4MPT reductase (Mer), H_4MPT S-methyltransferase (Mtr), H_2-dependent methylene-H_4MPT dehydrogenase (Hmd), Methyl-CoM reductase (Mcr/Mrt), heterodisulfide reductase (Hdr), methyl-viologen hydrogenase (Mvh). The cofactors involved are: methanofuran (MF), tetrahydromethanopterin (H_4MPT), coenzyme M (CoM), coenzyme B (CoB), cofactor F_{420} (F_{420}).

The reducing potential required for methanogenesis is supplied by Frh and $F_{420}H_2$ independently of membrane potential, all genomes analysed are predicted to possess a full set of genes encoding hydrogenases Frh, MvH and EhaA involved in supplying reducing potential for methanogenesis, and possess a full set of genes encoding A_1A_0 ATP synthase that utilise the membrane potential for energy generation. In addition, a gene encoding NADP-dependent F_{420} reductase, NpdG (D5_1461) has been predicted in the all *Methanobrevibacter* genomes, however, the NADP dependent alcohol dehydrogenase, *adh* gene, that allows *Methanobrevibacter* spp. to utilise alcohol as reducing potential, is only predicted in the YE315, $M1^T$, $KM1H5\text{-}1P^T$ and YLM1 genomes.

6.3. Discussion

There are 15 type strains of *Methanobrevibacter* spp, all utilise CO_2 and H_2 as a substrate for hydrogenotrophic methanogenesis. In addition, *Mbb. smithii* PS^T, *Mbb. ruminantium* M1^T, *Mbb. woesei* GS^T, *Mbb. olleyae* KM1H5-1P^T, *Mbb. millerae* ZA-10^T, *Mbb. boviskoreani* JH1^T, *Mbb. arboriphilus* DH1^T and *Mbb. cuticularis* RFM-1^T are capable of using formate. Other strain dependent nutrient requirements such as CoM and acetate have been detailed in Table 6.1. In culture studies with D5, a slightly higher growth yield has been observed when acetate (20 mM) was supplemented in the media. *Methanobrevibacter* sp. D5 is reported to have a H_2 threshold similar to *Mbb. ruminantium* M1^T, *Methanobrevibacter* sp. AbM4 and *Mbb. millerae* SM9 (Kim 2012), which suggests the growth conditions and niche occupied by these strains are highly similar.

This chapter describes the sequencing, gap closing and bioinformatics analysis of the D5 genome. D5 is a member of the *Mbb. gottschalkii* clade isolated from the ovine rumen (Seedorf *et al.* 2014), and its genome sequence represents an important contribution to the knowledge accumulating around this dominant group of rumen methanogens. Its 2.8 Mb genome has provided new insight into the lifestyle and cellular processes of this organism. At the beginning of this project, no genome sequence was available for methanogens of the *Mbb. gottschalkii* clade. Currently (July 2016), four other members of the *Mbb. gottschalkii* clade have been sequenced, and 17 genome sequences of *Methanobrevibacter* spp. in total are included in this study. This chapter also compares the D5 genome to the other available *Methanobrevibacter* spp. genomes and has identified the core genome that is important to the survival of *Methanobrevibacter* spp.

The assembly confirmation indicated a discrepancy predicted to be caused by an undigested *Apa*I site at 226,045 bp (Figure 6.5.). Base 226,045 is within scaffold one, contig 26, and was not within a gap region, therefore the *Apa*I site was not introduced by PCR sequencing error. The *Apa*I restriction endonuclease cleavage may be inhibited by DNA-cytosine methyltransferase (Dcm) catalysed methylation at CpG (Larimer 1987). D5 is predicted to possess two *dcm* genes (D5_1219, D5_1220), therefore it is plausible that the *Apa*I restriction recognition site at 226,045 bp may have been methylated and prevented the *Apa*I cleavage. The 5-methylcytosine can be spontaneously deaminated to thymidine (Ehrlich *et al.* 1986), the resulting TG mismatch base pairs can be repaired by a short patch endonuclease Vsr (D5_1221) (Lieb 1991), which is encoded adjacent to the *dcm* genes.

A 170 kb band was observed in the undigested lane of D5 genomic DNA (Figure 6.5.). This result would typically indicate an extrachromosomal element, however, no contig of 170 kb in size was observed within the D5 assembled genome. Within the assembled sequence derived from the Illumina sequencing data, a second genome was present. The contaminant was identified as a member of the genus *Bacillus*, therefore it is likely that the extrachromosomal element belongs to the *Bacillus* contaminant, as no *Methanobrevibacter* spp. have been reported of carrying extrachromosomal elements thus far (Samuel *et al.* 2007; Leahy *et al.* 2010; Leahy *et al.* 2013; Lee *et al.* 2013; Khelaifia *et al.* 2014; Khelaifia *et al.* 2014; Kelly *et al.* 2016).

Within the assembled sequence derived from the Illumina data, a contig of 10,943 bp was found to have a read coverage four times that of the D5 contigs and over 300 times higher read number than the contaminating *Bacillus* contigs (Figure 6.3.). This result is often suggestive of an extrachromosomal element. The difference in read number suggested that this contig may be more likely to be associated with the D5 rather than the contaminating *Bacillus*, although some plasmids such as pUC19 from *Escherichia coli* may have as high as 700 copies per cell (Casali and Preston 2003). This contig has a %G+C of 27.6%, which is more similar to D5 (33.1%) than the *Bacillus* contaminant (37.4%). Attempts to circularise the contig via PCR have been unsuccesful. Furthermore, this contig was identified within the enrichment culture H6 sequenced by 454 pyrosequencing (Figure 6.2.). PCR was able to identity the persistence of this contig within the D5 culture (result not shown). This contig encodes 16 genes; 14 are hypothetical proteins with no homology to known genes, and two genes are homologous to an rRNA methyltransferase, *rlmN* gene (Tsac_1724) and a stage II sporulation protein E gene (Tsac_1725) of *Thermoanaerobacterium saccharolyticum* JW/SL-YS485 (GenBank: CP003184.1). *T. saccharolyticum* harbors an active prophage, THSA-485A, but the prophage region does not encompass the two homologous genes (Currie *et al.* 2015). Prophage and transposable elements have terminal repeats that facilitate their integration and excision by recombination (Levis *et al.* 1980; Temin 1981; Dejean *et al.* 1984). A direct repeat of 127 bp (ATGGGTTTATCAAATGATATTAAATTTTCAATCAATGTCTTATGGAATAATACTT TAATATATAGAAAACATTTAATTAATTTAAATGACTTAATACAAAGATTTAGAGT TTATAATAAAAATCAG) with 100% nucleotide sequence identity is present at base 1 and base 8975 of the contig. This repeat sequence displayed a 68% pairwise identity to a region upstream to a transposase (D5_1474) in the D5 genome, and 64.6% pairwise identity to a region on Node_163 of the contaminating *Bacillus* contigs. The true identity and association of this

contig remains unknown. The presence of prophage has been shown in four *Methanobrevibacter* genomes, including M1^T, SM9, YLM1 and PS^T, while no *Methanobrevibacter* genomes have plasmids. Therefore it is more likely that this contig may belong to a stray prophage rather than a plasmid sequence.

Eighteen genomes of *Methanobrevibacter* spp. sourced from four different gut environments (human, bovine, ovine and termite) and from an anaerobic digester have been compared in this study, to identify the extent of gene conservation and to gain insight into the metabolism and physiology of this group of methanogens. Amongst the sequenced *Methanobrevibacter* spp., the genome size ranges between 1.66 Mbp (ATM^T) and 2.94 Mbp (M1^T), and the %G+C range is between 24.2% (SH^T) and 36.4% (ZA-10^T). The genomes all contain circular chromosomes with no extrachromosomal elements. Seven of the genomes are complete, including D5, YE315, SM9, M1^T, PS^T, YLM1 and AbM4.

The phylogenetic relationship of members of the *Methanobrevibacter* can be inferred from their 16*S* rRNA gene sequences (Seedorf *et al.* 2014) (Figure 6.6.). The inferred relationships suggests that the organisms D5, YE315, ZA-10^T, SM9 and HO^T are members of the *Mbb. gottschalkii* clade, PS^T is in the *Mbb. smithii* clade, JMR01 is in the *Mbb. oralis* clade, M1^T, KM1H5-1P^T and YLM1 are members of the *Mbb. ruminantium* clade, SH^T, AbM4 and JH1^T are members of the *Mbb. wolinii* clade, while ANOR1, RFM-1^T, RFM-2^T, RFM-3^T and ATM^T are placed in the *Mbb. arboriphilus* clade, *Mbb. cuticularis* clade, *Mbb. curvatus* clade, *Mbb. filiformis* clade and *Mbb. acididurans* clade, respectively. RFM-1^T, RFM-2^T, RFM-3^T and ANOR1 are the most distantly related to other *Methanobrevibacter* spp. (Figure 6.6.). Recent advancements in bioinformatics methods recommends that a 16*S* rRNA gene identity of ≤98.7% can be considered as a separate species (Stackebrandt and Ebers 2006; Petti 2007; Schlaberg *et al.* 2012). Using this recommendation, *Mbb. boviskoreani* JH1^T and *Methanobrevibacter* sp. AbM4 can be considered as different strains of the same species, *Methanobrevibacter* sp. D5, *Mbb. millerae* ZA-10^T and *Methanobrevibacter* sp. YE315 can be considered as different strains of the same species, *Mbb. olleyae* KM1H5-1P^T and *Mbb. olleyae* YLM1 can be considered as different strains of the same species (Table 6.4.). A FGD analysis and a pangenome tree reveals a different profile (Figures 6.7., 6.8.) where JMR01 and PS^T are clustered together with members of the *Mbb. gottschalkii* clade. This suggests JMR01 and PS^T are similar to members of the *Mbb. gottschalkii* clade at the ORFeome level, whereas for members of the *Mbb. ruminantium* clade and *Mbb. wolinii* clade, their ORFeome similarity is consistent with the inferred phylogenetic relationship.

A reported feature of genomes from the *Methanobrevibacter* spp. is the organisation of rRNA genes into an operon (Samuel *et al.* 2007; Leahy *et al.* 2010), which is typically ordered 16*S*, 23*S* and 5*S* rRNA genes (Brosius *et al.* 1981). Archaeal rRNA operons vary in both structure and frequency (Dennis 1997), with 50% of sequenced archaeal genomes containing only one rRNA operon, but with some having as many as four operons (Yip *et al.* 2013). The rRNA operons in the *Methanobrevibacter* genomes analysed in this study contain between one to three rRNA operons. Some rRNA operons may contain one or two tRNAs within spacer sequences, including $tRNA^{Glu}$, $tRNA^{Ile}$, $tRNA^{Ala}$ and $tRNA^{Cys}$ (Lund *et al.* 1976; Ikemura and Nomura 1977; Morgan *et al.* 1977; Keller *et al.* 1980; Chant and Dennis 1986). Some archaeal genomes are found with the 5*S* rRNA gene existing outside the rRNA operon, as observed in the *Desulfurococcus mobilis*, *Sulfolobus solfataricus* and *Methanococcus jannaschii* genomes (Kjems and Garrett 1987; Bult *et al.* 1996; She *et al.* 2001). A $tRNA^{Ala}$ has been identified in one of the rRNA operons in the D5 genome, between the 16*S* rRNA gene and the 23*S* rRNA gene. The arrangement of $tRNA^{Ala}$ between the 16*S* and the 23*S* rRNA genes have also been identified in other completed *Methanobrevibacter* genomes. YE315, SM9 and PS^{T} only have one $tRNA^{Ala}$ containing rRNA operon, the $tRNA^{Ala}$ is present in both rRNA operons of $M1^{T}$ and YLM1 genome, and in the AbM4 genome the $tRNA^{Ala}$ is found in two of the three rRNA operons. The arragenement of $tRNA^{Ala}$ in the rRNA operon may reflect the phylogenetic relationship of the *Methanobrevibacter* spp., as both YLM1 and $M1^{T}$ are members of the *Mbb. ruminantium* clade, and D5, YE315 and SM9 are members of the *Mbb. gottschalkii* clade (Seedorf *et al.* 2014).

The genome of D5 encodes 44 tRNAs encoding all amino acids. All except one of the *Methanobrevibacter* genomes analysed are predicted to encode a full set of tRNAs corresponding to 20 amino acids. The $KM1H5\text{-}1P^{T}$ genome is predicted to lack a $tRNA^{Ala}$, however the $KM1H5\text{-}1P^{T}$ genome is incomplete and thus the $tRNA^{Ala}$ may be contained within the missing region. Introns in tRNAs are found in all three kingdoms of life and are present on average in 15% of archaeal tRNAs (Yoshihisa 2014). Two of the D5 tRNAs, $tRNA^{Met}$ and $tRNA^{Trp}$, contain an intron. The tRNA intron endonuclease *endA* (D5_1660) required for tRNA intron removal (Thompson *et al.* 1989) has been identified within the D5 genome. All of the *Methanobrevibacter* genomes analysed are predicted to encode intron containing $tRNA^{Met}$ and $tRNA^{Trp}$, which suggests this may be a conserved trait to the *Methanobrevibacter* spp. Similar to D5, all the genomes analysed are predicted to encode an *endA* gene. Sixteen of the genomes analysed possess similar number of tRNAs, ranging from 33 (YE315, JMR01, $KM1H5\text{-}1P^{T}$,

YLM1, RFM-3^T and ATM^T) to 44 (D5). There are two genomes predicted with exceptionally high number of tRNAs, M1^T with 58 and ZA-10^T with 77, attributed partially by a high copy number of pseudo tRNAs and tRNA^{Leu}. M1^T and ZA-10^T also possess an additional intron-containing tRNA, a tRNA^{Lys} in M1^T and a tRNA^{Ser} in ZA-10^T.

The codon and amino acid usage of the genomes was compared and found to be similar between the 18 *Methanobrevibacter* spp. (Figure 6.11., Table A.6.2.). A higher usage of asparagine was observed in the termite sourced RFM-1^T, RFM-2^T and RFM-3^T. M1^T and ZA-10^T displayed a mildly different codon usage pattern, and incidentally, these two genomes also possess an unusually high number of tRNAs with many clustered together. It was found that laterally transferred genes may carry codons not used by the host, and thereby prompting the selective retention of foreign tRNAs in the host genome (McDonald *et al.* 2015). M1^T was found to possess genes that are likely to have been horizontally acquired, including a NRPS (Leahy *et al.* 2010). Therefore it is plausible that ZA-10^T may also possess horizontally transferred genes that has contributed to the difference in codon usage and the high copy number of tRNAs.

The CRISPR/Cas system facilitates acquired immunity (Marraffini and Sontheimer 2010) and has been classified into three major types and ten subtypes based on the signature genes present (Makarova *et al.* 2011). Both Type I and Type II contain the central *cas1* and *cas2* genes required for spacer sequence insertion, while the signature gene for Type I CRISPR/Cas system is a *cas3* gene that cleaves the invading DNA. For Type II the signature gene is the large *cas9* gene that utilises trans-encoded small RNA (tracrRNA) to guide the processing of pre-crRNA. The signature gene for Type III is *cas10* gene, and many Type III were found to lack *cas1* and *cas2* genes (Makarova *et al.* 2011; Makarova *et al.* 2011). While most of the CRISPR/Cas system can be readily classified into the three major types, some CRISPR/Cas system lacks signature genes to all three major types, and remain unclassified as Type U (Makarova *et al.* 2011). The CRISPR/Cas system of *Methanobrevibacter* sp. D5 lacks the signature genes to all three major types, and is therefore classified as type U. The presence of *cas10* and *csm3* genes in SM9, JMR01, M1^T, KM1H5-1P^T, YLM1, SH^T, RFM-2^T and RFM-3^T suggested these genomes possess a Type III-A CRISPR system. The presence of *cas3* gene in YE315, ZA-10^T, PS^T, AbM4, JH1, RFM-1^T and ATM^T suggested these genomes possess a type I CRISPR system, and the genes present are insufficient to further classify into subclasses. The *cas3* and *cas10* genes were not identified in the HO^T genome, which classifies the CRISPR elements in the HO^T genome as type U. Interestingly, despite the presence of CRISPR associated genes in the ZA-10^T and HO^T genomes, no CRISPR elements were identified in the ZA-10^T and HO^T

genomes. Since neither genomes were complete, and repeat elements such as CRISPR can often be difficult to sequence, it is possible the CRISPR elements are within the gap regions. No CRISPR associated genes were detected in the ANOR1 genome, which questions whether the ANOR1 genome CRISPR elements identified were genuine. The three predicted CRISPR elements in the ANOR1 genome consist of small number of spacers ranging from two to five, which suggest the identified elements may be other types of repeat elements.

The spacer sequences of the identified CRISPR element were used to identify foreign genetic elements that *Methanobrevibacter* sp. D5 may have encountered in the past (Biswas *et al.* 2013; Shariat *et al.* 2015). Out of 20 spacer sequences analysed, 11 matched known phage and plasmid sequences. Plasmid hosts included *Providencia rettgeri*, *Bacillus cereus* and *Clostridium perfringens*, which can be found in the soil (Yamagishi *et al.* 1964; Kitts *et al.* 1994; Arnesen *et al.* 2008), as well as the parasitic Periwinkle leaf yellowing phytoplasma found in plants (Nejat *et al.* 2013). It is conceivable that within the rumen environment *Methanobrevibacter* sp. D5 may have come into contact with similar organisms, as ruminants ingest soil while grazing. Furthermore, phytoplasma may have colonised the forage plant material, and exposed the grazing ruminant to their mobile genetic elements.

Horizontal gene transfer occurs commonly amongst prokaryotes, and continuous acquisition and removal of genes in the genome is an integral part of genome evolution (Ochman *et al.* 2000; Koonin *et al.* 2001), assisting with environmental adaptation (Gilbert and Cordaux 2013). The genes acquired from horizontal transfer often differ to other genes in the genome in base composition, codon usage and bias of di- and tri-nucleotide frequencies (Ochman *et al.* 2000). The regions that may contain horizontally transfered DNA have been predicted in the D5 genome. However, most of the regions have a low IVOM score as well as a high average CAI which suggests the likelihood of containing horizontally transferred DNA is low. Furthermore, some regions contain genes predicted to be involved in important metabolism; the two regions with highest IVOM score is actually where the rRNA operons are found, and they are unlikely to be horizontally transferred. Region 30 had the third highest IVOM score with a predicted transposase within and two hypothetical genes that are also predicted to be horizontally transferred.

IS elements are segments of DNA, less than 2.5 Kb in length, and capable of inserting into a target sequence (Frost *et al.* 2005). IS elements often encode a transposase flanked with short inverted repeats necessary for transposase recognition and cleavage (Mahillon and Chandler

1998). A region of DNA flanked by two IS elements may be mobilized as a composite transposon (Siguier *et al.* 2015). IS elements constitute 1.8% of the D5 genome, whereas IS elements were found to constitute only 0.2% of the $M1^T$ genome (Leahy *et al.* 2010). Higher number of IS elements may act as a driving force of genome evolution (Biemont and Vieira 2006), aiding D5 to persist and define its niche within the competitive habitat of the rumen environment.

The *Methanobrevibacter* cells are resistant to lysis by SDS (Miller and Lin 2002), and to various antibiotics that targets bacterial cell wall, including but not limited to vancomycin, penicillin and cephalothin (Dermoumi and Ansorg 2001). These properties are coneyed by possession of pseudomurein in their cell wall and the triple-layered structure of their cell envelopes (Zeikus and Henning 1975; Kandler and Konig 1978), with an electron dense, thin pseudomurein layer outside the cell membrane, followed by a thicker but less tightly packed pseudomurein layer surrounding it, and a cell wall glycopolymer at outermost layer (Leahy *et al.* 2010). The specific genes involved in pseudomurein biosynthesis have been proposed but still require experimental characterisation (Kandler and Konig 1998). In addition to pseudomurein, the sugar content of the cell wall includes galactose, rhamnose and low levels of glucose and mannose (Kandler and Konig 1978), which may consitute the glycopolymer outer layer. Several genes may be involved in the biosynthesis of teichoic acid, sialic acid and exopolysacharide, as listed in Table A.6.1.

Extracellular and surface-associated proteins are likely to play an important role in many essential interactions and adaptations to the environment. D5 is predicted to use the SRP, Sec61 translocation system for translocation of gene products (Zwieb and Bhuiyan 2010). Although D5 is predited to encode a sortase family protein (D5_0315), no LPxTG/NPxTG/LpxTA/LAxTG sorting motif harboring genes were detected (Ton-That *et al.* 2004), which suggests the sortase family protein may not be a functional sortase. Three homologs (D5_1979, D5_2197, D5_2198) of a Type II secretion system were predicted in the D5 genome, however, a functional Type II secretion system typically involves at least 12 different gene products (Sandkvist 2001), which suggests D5 does not carry a functional Type II secretion system. The three genes homologous to the Type II secretion system are conserved across the *Methanobrevibacter* genomes, with an average amino acid identity of 59.1%, 68.6% and 52.2% respectively for the corresponding D5_1979, D5_2197 and D5_2198 genes. The twin-arginine translocation system only requires two subunits to export folded gene products (Ellen *et al.* 2010), however no homologous genes were identified in the D5 genome. Another

type of secretion mechanism employed by archaea is the formation of a vesicle by ESCRT-III (endosomal sorting complex required for transport) proteins (Ellen *et al.* 2009), but no homologs of ESCRT gene were predicted in the D5 genome. There are 173 D5 genes predicted to encode an N-terminal signal peptide (Figure 6.12.), 33 proteins are predicted with two or more transmembrane domain, and an additional 103 proteins are predicted to be anchored through a N-terminal transmembrane domain, two proteins are predicted to be anchored through a C-terminal transmembrane domain. In the genomes analysed, an average of 4.9% of their ORFeomes are devoted to the process of protein export. D5 is predicted to have the highest proportion of exported gene products, 7% of its ORFeome, while SH^T is predicted to have the lowest proportion of exported gene products, 3.5% of its ORFeome (Table 6.12.). Proteins predicted to contain a signal peptide without a transmembrane helice or lipobox are thought to be secreted into the environment. The number of proteins predicted to be secreted into the environment range between 3 and 26 per genome for the *Methanobrevibacter* genomes analysed. Extracellular proteins with domain repeats are often involved in binding to the cell (Cabanes *et al.* 2002), and all of the genomes examined contain genes with predicted repeat domains (27, 19, 22, 26, 5, 11, 7, 37, 9, 10, 3, 2, 3, 12, 3, 9, 2 and 6 genes for D5, YE315, ZA-10^T, SM9, HO^T, PS^T, JMR01, $M1^T$, KM1H5-$1P^T$, YLM1, SH^T, AbM4, $JH1^T$, RFM-1^T, RFM-2^T, RFM-3^T, ATM^T and ANOR1 respectively). Amongst the secretome, there are four other Pfam domains conserved across all the genomes analysed, these domains are PF09972 DUF2207, PF13229 Right handed beta helix, PF02516 Oligosaccharyl transferase STT3 and PF04206 H_4MPT S-methyltransferase subunit E (MtrE) (Table A.6.4.). The function of DUF2207 is unkown, right handed beta helix is a solenoid protein domain formed by parallel beta strands in helical arrangement, its structure has been found within pectate lyase (Jenkins *et al.* 2004), surface adhesin PfbA of *Streptococcus pneumoniae* (Beulin *et al.* 2014) and periplasmic alginate epimerase AlgG (Wolfram *et al.* 2014). The oligosaccharyl transferase STT3 domain has been known to function in *N*-glycosylation of the S-layer glycoprotein in *Haloferax volcanii* (Abu-Qarn *et al.* 2007), this suggests all *Methanobrevibacter* spp. utilise a *N*-glycosylated pseudomurein. The Mtr is essential to hydrogenotrophic methanogenesis and contributes to generation of the Na^+ gradient, as all *of the Methanobrevibacter* spp. are hydrogenotrophic methanogens, it is conceivable for all genomes analysed to encode a transmembrane protein with signal peptide predicted with the MtrE domain.

Adhesins are large, repeat-rich cell surface proteins that mediate adhesion to other cells or surfaces and biofilm formation (Fuqua 2010). Adhesins are responsible for methanogen

adherence to H_2 producers such as protozoa (Finlay *et al.* 1994). Adhesin-like proteins have been identified in *Mbb. smithii* and *Mbb. ruminantium* $M1^T$, and *Mbb. smithii* has adhesin-like proteins homologous to pectin esterase and mediates in binding of human mucosal chondroitin (Samuel *et al.* 2007). *Mbb. ruminantium* $M1^T$ was found to co-aggregate with *Butyrivibrio proteoclasticus* in coculture, and six adhesin-like proteins were upregulated (Leahy *et al.* 2010). One of the upregulated adhesin-like proteins (mru_1499) was found to be a broad spectrum adhesin important in binding of $M1^T$ to a broad range of rumen protozoa (Ng *et al.* 2016). D5 is predicted to encode 82 adhesin-like proteins, which is less than $M1^T$ (105), SM9 (95) but more than PS^T (48) and AbM4 (29). The mru_1499 adhesin-like protein is predicted to harbor a N-terminal signal peptide, four Big_1 (PF02369: bacterial Ig-like (group 1)) domains and a C-terminal transglutaminase-like domain (PF01841) (Ng *et al.* 2016). Using phage display, the N-terminal Big_1 domain was discovered to have an important role in cell binding (Ng *et al.* 2016). The adhesin-like proteins predicted with transglutaminase domain was proposed to mediate cell adherence by cross-linking (Griffin *et al.* 2002), which was later revealed to be less important than big_1 domains in cell adherence (Ng *et al.* 2016), and has been proposed that it may be involved in modification of the cell surface (Leahy *et al.* 2010). Of the 82 predicted adhesin-like proteins in D5, 16 are predicted to harbor Big_1 domains, of which three are also predicted to harbor a transglutaminase-like domain, including a D5_1662 gene that is homologous to mru_1499. Therefore, it is conceivable that D5 may possess a similar cell binding ability to $M1^T$, capable of binding to H_2 producing protozoa and bacteria in the rumen. In addition to D5, adhesin-like proteins homologous to mru_1499 has been identified in KM1H5-$1P^T$, YLM1, PS^T, ZA-10^T, SM9, AbM4 and *Methanosphaera stadmanae* MCB-3 (Ng *et al.* 2016), which suggests conserved domains. The adhesins maybe a crucial component of how the *Methanobrevibacter* spp. interact with their environment, the difference in the microorganism or host gut it binds to, assists them in establishing their own niche in the presence of closely related species.

The conservation in genome structure and gene order between different genomes is known as synteny (Passarge *et al.* 1999). When two genomes have a high degree of synteny, a linear genome alignment can be observed across the entire length of the genome, and when an X-alignment is observed, as reported for *Mycobacterium leprae* and *Mybacterium tuberculosis*, this indicates that matching sequences occur at the same distance from the origin but not necessarily on the same side of the origin (Eisen *et al.* 2000). Using the D5 genome as a reference, it was shown that a high degree of synteny is shared with SM9 and YE315 genomes.

An inversion was observed in the YE315 analysis, this inversion encompasses approximately 1 Mb of the genome. A similar inversion was also observed in the SM9 genome that encompasses approximately 2.3 Mb of the genome. The D5 genome has a moderate degree of conservation to $M1^T$, AbM4 and YLM1 genome in regions approximately 0.9 Mb from the origin of replication *cdc6-1* gene. However the middle section of the genome lacks synteny, which suggest this region of the genome may harbor clade-specific genes, and the conserved genes between *Methanobrevibacter* spp. tend to lie closer to the origin of replication.

Comparison between *Methanobrevibacter* genomes can reveal conserved genes important for the organisms' growth and survival as well as those genes that might be important for niche adaptation. The core genome of the 18 *Methanobrevibacter* spp. examined consists of 764 gene families, while the order level pan-genome encompasses 9,382 gene families (Figure 6.14.). The sequenced *Methanobrevibacter* spp. analysed in this study originated from six environments, ovine (7) and bovine (4) rumen, termite gut (3), anaerobic digester (1), human faeces (2) and subgingival plague (1). There are two gene families conserved between the rumen strains, which might shed light on the genes involved in adaptation to the rumen environment. Each genome also contains its own unique gene families, from 31 gene families in AbM4 to 649 gene families in JMR01 (Figure 6.16.). These unique gene families can offer insight into how these members of the *Methanobrevibacter* may establish their own niche in the presence of similar organisms within the same environment.

The core genome of *Methanobrevibacter* spp. was classified according to the COG database (Table 6.14.). A total of 85 gene families are implicated in energy conservation, which includes the genes required for hydrogenotrophic methanogenesis, and the *fdh*BCD genes involved in formate utilisation, but not the *fdhA* gene, as the *fdhA* gene exists as a pseudogene in the YE315 genome, although it is conserved in 17 other *Methanobrevibacter* spp. genomes analysed. This suggests CO_2, H_2 and formate are the conserved substrates of *Methanobrevibacter* spp. While experimental findings support the ubiquitous usage of CO_2 and H_2 amongst known *Methanobrevibacter* spp., experimental evidence has shown some members of the genus *Methanobrevibacter* are incapable of utilising formate, including *Mbb. wolinii* SH^T, *Mbb. gottschalkii* HO^T, *Mbb. oralis* ZR^T, *Mbb. curvatus* RFM-2^T, *Mbb. filiformis* RFM-3^T and *Mbb. acididurans* ATM^T (Smith and Hungate 1958; Bryant *et al.* 1971; Balch *et al.* 1979; Ferrari *et al.* 1994; Leadbetter and Breznak 1996; Leadbetter *et al.* 1998; Miller 2001; Miller and Lin 2002; Savant *et al.* 2002; Rea *et al.* 2007; Lee *et al.* 2013). A similar connundrum has been encountered in the *Methanococcus* spp., where the presence of formate utilising genes in the

genomes does not correspond to actual utilisation of formate in the cultures (Wood *et al.* 2003). Two sets of formate utilising genes have been found in the *Methanobrevibacter* spp. genomes analysed, one set is the *fdhDCAB* found in all *Methanobrevibacter* spp. genomes analsyed, the other is the *flpABD* that was found in 11 of the 18 genomes analysed. The *flpBD* genes are present at end of the contig in the HO^T genome, therefore it is possible the *flpA* gene lies in the gap region, which would make it 12 of the 18 genomes possessing *flpABD* genes (Table A.6.9.). The *flpA* genes encodes approximately 384 aa and is homologous to the N-terminal domain of FdhA, whereas the *fdhA* genes encode approximately 712 aa in length. Furthermore, *flpAB* genes are annotated as molybdopterin oxidoreductase in the ZA-10^T, KM1H5-1P^T and JMR01 genomes, and since it is not present in the genome of the formate utilising RFM-1^T, it brings into questions whether the *flpABD* genes are genuinely involved in formate utilisation. The formate transporter gene *fdhC* forms a conserved operon preceding the formate dehydrogenase *fdhAB* genes in all *Methanobrevibacter* genomes analysed, with one or two accessory protein *fdhD* genes present elsewhere in the genome. Directly downstream of *fdhCAB* genes are molybdenum cofactor biosynthesis protein *moaA*, *mobB* genes, followed by the formylmethanofuran dehydrogenase *fwdHFGDBAC* operon in all genomes analysed except RFM-2^T, in which the *moaA* gene is located elsewhere. To identify whether differences within the genes contributed to the difference in formate utilisation, *fdhCAB* genes from all *Methanobrevibacter* spp. genomes were aligned (result not shown), and a high degree of gene conservation was observed between the *Methanobrevibacter* genomes. The *fdhA*, *fdhB* and *fdhC* genes have 76%, 72.2% and 67.1% aa identity respectively, but no conservation could be attributed solely to formate utilising genomes. To identify whether differences in regulation may contribute to the difference in formate utilisation, the 5' region of *fdhA* and *fdhC* genes were analysed. The formate utilising genes in *Methanococcus* spp. were upregulated by the absence of H_2 (Wood *et al.* 2003), and perhaps the *fdhCAB* operon in the strains that were found to not utilise formate were down-regulated strongly by the presence of H_2. The experimental conditions carried out by strains SH^T, HO^T, CW^T, RFM-2^T, RFM-3^T and ATM^T all used H_2:CO_2 (80:20, v/v, 202 kPa), therefore reducing the partial pressure of H_2 in culture might enable formate utilisation of these strains. The genes involved in proline, tryptophan and CoM biosynthesis are absent from the core genome. All genomes analysed lack the genes required to produce proline, therefore *Methanobrevibacter* spp. may require external proline to survive. Most strains other than PS^T and KM1H5-1P^T are reported to require amino acid containing supplements such as yeast extract, casamino acid, or fecal extract to enable good growth. Perhaps PS^T and KM1H5-1P^T possess an uknown pathway of proline biosynthesis.

JMR01 lacks the genes required to produce tryptophan, and may require external supply to survive, which is conceivably found in the yeast extract and human fecal extract supplemented to JMR01 under culture conditions (Ferrari *et al.* 1994). D5 lacks the *thrB* gene required to produce threonine, and may require external supply of threonine to survive. A similar conundrum to the formate utilisation is observed in the biosynthesis of CoM. $M1^T$ has long been known to require external CoM for growth (Taylor *et al.* 1974), and the genome sequence showed the lack of *comADE* genes required for CoM biosynthesis (Leahy *et al.* 2010). Two other *Methanobrevibacter* spp. were also shown to require external CoM supply in culture, $JH1^T$ and SH^T of the *Mbb. wolinii* clade. However, both genomes are predicted to encode the full set of genes required for CoM biosynthesis. The genes *comABCDE* were aligned between all genomes analysed, the genes were highly conserved, with an average of 67.7%, 53.3%, 66%, 69.2% and 62.4% aa identity respectively. There are minor difference in amino acid residues, but whether this affects the functionality of the enzyme would require experimental evidence. In ComE, the residue 25 is F and residue 33 is I in all genomes but SH^T, $JH1^T$ and AbM4 which used V for both residues, residue 89 is P in all genomes except SH^T, which has a C at residue 89. In ComC, the residue 42 is T in all genomes except SH^T, which has a S at residue 42. In ComB, the residue 97 is T in all genomes except SH^T, which has a S at residue 97, the residue 103 is I in all genomes except SH^T and $M1^T$, which has V and T respectively at residue 103, the ComB is also 10 to eight residues shorter at the C-terminal in $M1^T$, SH^T, $JH1^T$ and AbM4 genomes. Whether or not these sequence differences could lead to difference in functionality requires further investigation.

There are two gene families identified as being conserved amongst the *Methanobrevibacter* genomes of rumen origin (Table 6.15.). The glycosyltransferase GT2 family protein is classified under cell wall/membrane biogenesis in COG categories, it is likely involved in protein glycosylation and may be an adaptation to the rumen environment.

There are 15 gene families identified as conserved amongst the genomes of *Mbb. gottschalkii* clade (Table 6.16.). Although most of the gene families lack predicted function, the aforementioned putative formate dehydrogenase gene, *flpA,* and a MFS (Major facilitator superfamily) transporter gene family consisting of two genes are functionally characterised and conserved to all members of *Mbb. gottschalkii* clade. The MFS transporters are known to catalyse the transport of a broad range of solutes (Kaback *et al.* 2001), it may be of interest to identify the substrate of this particular MFS transporter gene family experimentally.

There are 171 gene families uniquely conserved in the genomes of the *Mbb. ruminantium* clade (Figure 6.16.). While 125 gene families are poorly characterised, there are 17 gene families involved in coenzyme transport and metabolism (Table 6.14.). It was found only members of the *Mbb. ruminantium* clade carry the full complement of genes *bioWFADB* required to produce biotin from pimelate (Bower *et al.* 1996), which suggests only members of *Mbb. ruminantium* clade can carry out *de novo* biosynthesis of biotin. All the genomes analysed are predicted to encode *pycAB* genes encoding a biotin dependent pyruvate carboxylase along with *birA* gene encoding a biotin-acetyl-CoA carboxylase ligase, which suggests *Methanobrevibacter* spp. other than YLM1, M1^T and KM1H5-1P^T would require an external supply of biotin to survive. The biotin transporter BioY that could facilitate biotin uptake has been predicted in ten of the genomes analysed, including D5, ZA-10^T, SM9, YE315 and HO^T of *Mbb. gottschalkii* clade, JMR01, PS^T, AbM4, SH^T, RFM-1^T, RFM-2^T, RFM-3^T, ANOR1 and ATM^T, which suggsets these organisms are capable of biotin uptake.

There are 168 gene families uniquely conserved in the genomes of the *Mbb. wolinii* clade (Figure 6.16.), while 120 gene families are poorly characterised, four gene families are classified under cellwall/membrane/envelope biogenesis (Table 6.14.). Most notable is a gene encoding aspartate racemase involved in conversion of L-aspartate to D-aspartate (Schleifer and Kandler 1972), this gene is not predicted outside of the *Mbb. wolinii* clade. Bacteria utilise D-aspartate ligase to add D-aspartate onto peptidoglycan pentapeptide (Bellais *et al.* 2006), no homolog of D-aspartate ligase gene were identified in all of the *Methanobrevibacter* spp. genomes analysed, which makes it uncertain whether the D-aspartate is attached to the pseudomurein. Possibly the enzymes involved differ strongly to the bacerial counterpart that targets peptidoglycan. The possible incorporation of D-aspartate into the pseudomurein contradicts the dogma that pseudomurein lacks D-amino acids and muramic acid (Konig and Kandler 1979). 2% (D/D+L) D-aspartate has been identified in the membrane fractions of *Pyrobaculum islandicum*, *Methanosarcina barkeri* and *Halobacterium salinarium* (Nagata *et al.* 1999), which suggests the cell membrane of members of the *Mbb. wolinii* clade may contain D-aspartate.

There are 444 gene families identified as unique to D5 (Table A.6.4.), including six genes encoding glycosyl transferases that may be involved in exopolysaccharide formation, and 75 adhesin-like proteins. The adhesin-like proteins may play a role in adherence to solid substrates, host cells or other microorganisms within the rumen. A set of genes unique to D5 were the endonuclease (D5_2164) and methylase (D5_2165) of a Type III restriction system,

while Type I and Type II restriction systems are found in other *Methanobrevibacter* genomes. Restriction systems are considered a defence system against phage and plasmids, the host DNA is normally methylated by methylases post-replication, and restriction endonuclease targets non-methylated DNA, such as phage and plasmids (Tock and Dryden 2005). This suggests D5 may have an additional line of defence against foreign genetic elements.

Other noteworthy genes unique to D5 includes a non-ribosomal surfactin synthetase (D5_0482). As mentioned in Chapter 4, Section 4.3, NRPSs are responsible for biosynthesis of small peptides and antibiotics (Witting and Sussmuth 2011), that are produced by consecutive condensation of substrates (Caboche *et al.* 2008). Surfactin refers to bacterial cyclic lipopeptide consisting of hydrophobic fatty acids and hydrophilic peptide chains (Roongsawang *et al.* 2010). The first surfactin was discovered in *Bacillus subtilis* (Arima *et al.* 1968), and it was found to have antibacterial and antitumor activities (Tsukagoshi *et al.* 1970; Kameda *et al.* 1974). Keeping a gene as large as a NRPS in the genome is energetically expensive, therefore it seems likely that the gene provides certain advantages for D5 survival within the rumen. D5 is not the only member of *Methanobrevibacter* that possess NRPS genes, $M1^T$ and SM9 have also been predicted to contain NRPSs (Leahy *et al.* 2013; Kelly *et al.* 2016). Additionally, YE315, ZA-10^T, RFM-1^T and ANOR1 are also predicted to possess NRPSs.

All *Methanobrevibacter* genomes analysed are predicted to possess a full complement of genes required to carry out hydrogenotrophic methanogenesis from CO_2 and H_2 (Table A.6.9.). The C_1 group is transferred from CO_2 to MF by Fwd, which acquires electrons from Fdx_{red} generated by the Eha complex via a Na^+ potential (de Poorter *et al.* 2003; Thauer *et al.* 2010). The Ehb complex is also involved in the supply of reducing potential, however the reducing potential is supplied to the autotrophic CO_2 assimilation in central carbon metabolism rather than the methanogenesis pathway (Major *et al.* 2010). The C_1 group is then transferred to H_4MPT by Ftr, and reduced sequentially from formyl to methenyl by Mch, from methenyl to methylene by Mtd or Hmd, from methylene to methyl by Mer (Thauer 1998; Shima *et al.* 2002). There are two enzymes that can reduce methenyl-H_4MPT to methylene-H_4MPT, the F_{420} dependent Mtd and the F_{420} independent Hmd. 13 *Methanobrevibacter* genomes are predicted to encode both enzymes. ANOR1, RFM-1^T, RFM-2^T, RFM-3^T and ATM^T are predicted to only encode Mtd, however hydrogenotrophic methanogenesis can be carried out by either Mtd or Hmd (Hendrickson and Leigh 2008). While most of the genes encoding hydrogenotrophic methanogenesis exist in single copy, additional copies of the *mtd* gene are predicted in the D5, ZA-10^T and SM9 genomes, an additional copy of *hmd* gene is predicted in the YE315, ZA-10^T

and SM9 genomes, and an additional copy of *mch* gene is predicted in the JMR01 genome. The multiple copies of these genes suggests their functional importance to the methanogen hosts. The H_2 and $F_{420}H_2$ can be interconverted by the Hmd-Mtd cycle (Hendrickson and Leigh 2008), which suggests ANOR1, RFM-1^T, RFM-2^T, RFM-3^T and ATM^T would not be able utilise the Hmd-Mtd cycle to interconvert $F_{420}H_2$ and H_2. It is also possible that the *hmd* gene may be present in the gap region of ANOR1, RFM-1^T, RFM-2^T, RFM-3^T and ATM^T, as these genomes are currently in draft form, only completion of the genome can rule out this possibility. Both Mtd and Mer require reducing potential supplied from $F_{420}H_2$ (Thauer 1998; Shima *et al.* 2002). All genomes are predicted to encode the genes required for F_{420} biosynthesis and the Frh hydrogenase (Table A.6.4.). Alcohol may act as an alternative source of reducing potential for F_{420} in the presence of the F_{420}-dependent NADP reductase NdpG and NADP-dependent alcohol dehydrogenase (Berk and Thauer 1997), as demonstrated in M1^T (Leahy *et al.* 2010). While all genomes are predicted to encode the *npdG* gene, the homologous gene to the NADP-dependent alcohol dehydrogenase is only predicted in YE315 and three members of the *Mbb. ruminantium* clade. The conservation of the gene encoding NADP-dependent alcohol dehydrogenase in the *Mbb. ruminantium* clade suggests it could be a conserved characteristic of this group. The conservation of the *ndpG* gene suggests other *Methanobrevibacter* spp. that lack the NADP-dependent alcohol dehydrogenase may utilise an unknown source of reducing potential other than alcohol.

The methyl group is transferred to CoM by Mtr, which is the main Na^+ potential generating enzyme complex (Harms *et al.* 1995). The Na^+ potential is utilised by the A_1A_0 ATP to generate ATP, however the A_1A_0 ATP synthase can also use a H^+ gradient at low pH and low Na^+ condition (McMillan *et al.* 2011). The methyl-CoM is then reduced to CH_4 by the Mcr/Mrt. complex. Isozymes Mcr and Mrt display different operon structures within a genome, *mcrBDCGA* and *mrtBDGA* respectively (Friedrich 2005). The two isozymes were found to be differentially regulated due to availability of CO_2 and H_2. Mrt is favoured in high H_2 availability, whereas Mcr is favoured in low H_2 availability (Bonacker *et al.* 1993). M1^T and AbM4 were found to only contain the *mcr* genes, and thus may favour a low H_2 environment (Leahy *et al.* 2010; Leahy *et al.* 2013). In addition, ovine rumen metatranscriptomics has revealed Mcr encoding genes to have a significantly higher expression in high CH_4 emitting animals, while no significant difference was found for the expression of Mrt encoding genes between high and low CH_4 emitting animals (Shi *et al.* 2014). In this study, YE315, ANOR1, SH^T and JH1^T are predicted to only carry the *mcr* genes, similar to M1^T and AbM4, which

suggests these *Methanobrevibacter* spp. may be similar to M1^T and AbM4 and favour a low H_2 environment. All three members of the *Mbb. wolinii* clade were found to only possess *mcr*, which suggests it could be a conserved feature of the *Mbb. wolinii* clade. Conversely, other *Methanobrevibacter* spp. genomes analysed encodes both *mcr* and *mrt*, which suggests they may be more versatile and capable of growing in a wider range of H_2 partial pressures.

Methanobrevibacter spp. are the predominant members of rumen methanogens, but it is not the only genus of the order Methanobacteriales in the rumen (Janssen and Kirs 2008; Seedorf *et al.* 2014; Henderson *et al.* 2015; Seedorf *et al.* 2015). There are several difference between *Methanobrevibacter* spp. and other methanogens that may have contributed to its predominance. Both *Methanobrevibacter* spp. and *Methanosphaera* spp. belong to the Methanobacteriales order, but they differ morphologically and metabolically (Fricke *et al.* 2006). *Methanobrevibacter* spp. are short rods while *Methanosphaera* spp. are spherical shaped, *Methanobrevibacter* spp. rely on hydrogenotrophic methanogenesis to live, while *Methanosphaera* rely on methylotrophic methanogenesis to survive (Fricke *et al.* 2006). Another order of methanogens that possesses the hydrogenotrophic methanogenesis pathway are the Methanosarcinales, but because Methanosarcinales contain cytochromes (Heiden *et al.* 1994; Deppenmeier 2004; Maeder *et al.* 2006), the hydrogenotrophic methanogenesis would require much higher concentration of H_2 in the environment for the reaction to be thermodynamically favorable (Thauer *et al.* 2008). Therefore, Methanosarcinales primarily rely on the dispropotionating methylotrophic methanogenesis to survive in the rumen (Hippe *et al.* 1979; Maeder *et al.* 2006; Liu and Whitman 2008). The conservation of hydrogenotrophic methanogenesis genes across all *Methanobrevibacter* genomes suggests *Methanobrevibacter.* spp. occupy an important niche in the rumen. Hydrogenotrophic methanogenesis requires four moles of H_2 to produce one mole of CH_4, and it is able to effectively deplete H_2 and prevent its accumulation in the rumen, which is beneficial to the fermentative microbes that form a syntrophic relationship with the *Methanobrevibacter* spp.

6.4. Conclusions

Members of the *Methanobrevibacter* genus are the predominant rumen methanogen, and the *Mbb. gottschalkii* clade is the most dominant within the *Methanobrevibacter* (Janssen and Kirs 2008; Henderson *et al.* 2015). *Methanobrevibacter* sp. D5 is the fifth member of the *Mbb. gottschalkii* clade to be genome sequenced, alongside the genomes of *Mbb. millerae* ZA-10^T, *Mbb. millerae* SM9, *Methanobrevibacter* sp. YE315 and *Mbb. gottschalkii* HO^T. The genes

present in the D5 genome predicts it utilises CO_2, H_2 and formate for hydrogenotrophic methanogenesis, acetate as a carbon source, ammonium as a nitrogen source, and potentially requires an external supply of threonine and proline to survive. The utilisation of formate would require experimental evidence to prove, as all 18 *Methanobrevibacter* genomes analysed possess the full set of genes required to utilise formate, yet only eight type strains have been observed to be capable of doing so.

This study revealed the core genome conserved between 18 *Methanobrevibacter* spp and also clade-specific features. The core genome includes the full complement of genes required for methanogenesis, with only the Mcr, but not Mrt, complex conserved across *Methanobrevibacter* spp. This study also revealed clade specific genes, and members of the *Mbb. ruminantium* clade may be specialised in biotin biosynthesis and the use of alcohol as a source of reducing potential. Members of the *Mbb. wolinii* clade may have a specialised cell envelope that employs D-aspartate, as well as a lack of an Mrt complex that might make them less versatile than other *Methanobrevibacter* spp. Members of the *Mbb. gottschalkii* clade carries the lowest amount of clade specific genes, mostly with unknown function. The D5 genome was found to possess unique genes including large number of adhesins genes and a large NRPS gene predicted to produce surfactin, which may be beneficial to the survival of D5 in the rumen.

The *Mbb. gottschalkii* clade and *Mbb. ruminantium* clade are the two most dominant clades of rumen methanogens, this study finds the methanogenesis pathway of all *Methanobrevibacter* spp. genomes analysed to be highly conserved with only minor differences, which is unlikely to explain the co-existence of *Methanobrevibacter* spp. within the rumen. The most probable explanation to the co-existence of *Methanobrevibacter* spp. in the rumen is the large number of genes involved in adhesion. By binding to different microorganisms, digesta or host surface, the *Methanobrevibacter* spp. may be able to establish their own specialised niche despite their similarity in energy metabolism.

Chapter 7
Summary, conclusions and future directions

7.1. Rationale

NZ has a large number of ruminant livestock relative to its human population, and the CH_4 produced by these animals via enteric fermentation accounts for 79.9% of NZ's total CH_4 emissions (Ministry for the Environment, 2015). In order to formulate strategies for enteric CH_4 mitigation, it is necessary to understand the source of CH_4 and its formation in the rumen environment.

Rumen fermentation produces H_2 as a by-product, which is mainly utilised by rumen methanogens (Hungate 1966). Because most rumen methanogens use H_2, methanogen activity and growth are closely linked to the activity of H_2-producing microbes, such as rumen protozoa, fungi and bacteria (Janssen 2010). Understanding the characteristics of rumen methanogens and their complex relationships with other rumen microbes, is therefore essential to formulate strategies to mitigate enteric CH_4 emissions. Mitigation strategies such as ruminant vaccination and chemical inhibitors of methanogens requires selective targets, which ideally are functions that are conserved across all rumen methanogens and absent from the animal host and other members of the rumen microbiome (Attwood *et al.* 2008). The genome sequences of rumen methanogens can effectively identify the genes encoding these conserved functions and also improves our understanding of rumen methanogen biology. Before this PhD project began, very little was known about the genomics of rumen methanogens and only one rumen methanogen - *Mbb. ruminantium* $M1^T$, had been genome sequenced (Leahy et al 2010). The order Methanomassiliicoccales had only recently been defined, and although there were enrichment cultures available for RCC organisms, there were no genome sequences available for rumen representatives of this order. The aim of the thesis was to i) to determine the genome composition of the methanogenic archaeon RCC ISO4-H5, and how this differs from other sequenced methanogens, and ii) determine the metabolic scheme of RCC ISO4-H5, how it grows in the rumen, and use these features to attempt isolation of a pure culture of this organism, iii) sequence the genome of an additional representative of the genus *Methanobrevibacter* (*Methanobrevibacter* sp. D5) to look at genome composition variation

within this genus and to determine how these differences allow co-existence of multiple *Methanobrevibacter* species in the rumen.

7.2. Advancements since the start of this thesis

The improvement of high throughput sequencing technologies and the corresponding decrease in the cost of genome sequencing, has led to a dramatic increase in the amount of microbial genome sequence data. Since my PhD project began, 10 additional Methanomassiliicoccales genomes have become available, including *"Candidatus* Methanomassiliicoccus intestinalis Mx1-Issoire", "*Candidatus* Methanomethylophilus alvus Mx1201" and *Methanomassiliicoccus luminyensis* B10 sequenced from human sources (Borrel *et al.* 2012; Dridi *et al.* 2012; Borrel *et al.* 2013), "*Candidatus* Methanoplasma termitum MpT1" sequenced from the termite gut (Lang *et al.* 2015), Thermoplasmatales archaeon BRNA1 (NCBI Reference Sequence NC_020892.1), "Candidatus Methanomethylophilus" sp. 1R26, Methanomassiliicoccales archaeon RumEn M1 and RumEn M2 sequenced from the bovine rumen (Noel *et al.* 2016; Sollinger *et al.* 2016), and methanogenic archaeon ISO4-G1 and ISO4-G11 sequenced from the ovine rumen (Kelly *et al*. 2016; Leahy *et al*. 2016).

Similar genome sequencing advancements amongst *Methanobrevibacter* spp. have been made as part of the Hungate1000 project (Creevey *et al.* 2014), including *Methanobrevibacter* sp. YE315, *Mbb. millerae* ZA-10^T from the bovine rumen, *Mbb. gottschalkii* HO^T from horse faeces, *Mbb. wolinii* SH^T from sheep faeces, and *Mbb. olleyae* KM1H5-1P^T from the ovine rumen. Additionally, genomes of *Methanobrevibacter* sp. AbM4 (ovine rumen; Leahy *et al.* 2013) and *Mbb. millerae* SM9 (ovine rumen; Kelly *et al.* 2016), *Mbb. olleyae* YLM1 (ovine rumen; Kelly *et al.* 2016), *Mbb. boviskoreani* JH1^T (bovine rumen; Lee *et al.* 2013), *Mbb. oralis* JMR01 (human faeces; Khelaifia *et al.* 2014) and *Mbb. arboriphilus* ANOR1 (human faeces; Khelaifia *et al.* 2014; Khelaifia *et al.* 2014), *Mbb. cuticularis* RFM-1^T (termite gut; NCBI whole genome shotgun sequencing project LWMW00000000), *Mbb. curvatus* RFM-2^T (termite gut; NCBI whole genome shotgun sequencing project LWMV00000000), *Mbb. filiformis* RFM-3^T (termite gut; NCBI whole genome shotgun sequencing project LWMT00000000) and *Mbb. acididurans* ATM^T (anaerobic fermenter, Rosewarne *et al.*, 2016) are available. These new sequences have enhanced the comparative analyses conducted in this thesis, and have allowed conserved traits and important metabolic characteristics of the genus *Methanobrevibacter* and the order Methanomassiliicoccales to be identified with greater reliability.

7.3. Summary of thesis results

A large part of this thesis has been the determination of the genome composition of methanogenic archaeon ISO4-H5 and its comparison to available genomes of members of the order Methanomassiliicoccales. The ISO4-H5 genome sequence revealed it as a H_2-dependent methylotrophic methanogen. Its predicted substrates for methanogenesis include mono-, di-, and tri-methylamine, methanol and the methylated thiol compound, methyl-3-methylthiopropionate (M3MSP). Experimental verification of these substrates confirmed the utilisation of these substrates. The comparative genomic analyses of members of the Methanomassiliicoccales revealed the absence of the hydrogenotrophic methanogenesis pathway prior to the methyl CoM reduction step catalysed by the Mcr/Mrt enzyme. This reaffirms the inability of Methanomassiliicoccales to disproportionate the methyl-substrates to CO_2 and their reliance on methyl-compound reduction only. All the genomes analysed are predicted to contain the genes required to allow utilisation of methylamine as a substrate for methylotrophic methanogenesis. All of the Methanomassiliicoccales genomes contain the genes which are predicted to encode autotrophic CO_2 fixation through AMP recycling, catalysed by a RubisCo operon. The alternative coding of the amber stop codon to encode for pyrrolysine incorporation is a conserved feature found in all the Methanomassiliicoccales genomes analysed. However, the Mx1, B10 and RumEn M1 organisms possess large numbers of genes with genuine amber stop codons, and appear to regulate the incorporation of pyrrolysine only where the amber codons are associated with methylamine methyltransferase genes. The remaining Methanomassiliicoccales organisms appear to have fewer amber codons in their genomes, but appear not to regulate the incorporation of pyrrolysine and are predicted to incorporate pyrrolysine at a higher frequency. The capability of CoM biosynthesis within the Methanomassiliicoccales appeared to be environment-dependent, as all three members of Methanomassiliicoccales from human sources are capable of CoM biosynthesis, whereas none of the rumen- or termite-derived strains possess genes required for CoM biosynthesis.

This study successfully isolated ISO4-H5 from its original enrichment culture, and is the first pure culture of a ruminal member of the order Methanomassiliicoccales. ISO4-H5 was identified as consuming ethanol, methanol and nicotinamide via NMR analysis of spent culture supernatants. Transcriptomic analyses were also conducted and identified specific set of ISO4-H5 genes that were differentially regulated in response to high and low H_2 environments. In the low H_2 conditions, several genes involved in methanogenesis and supply of reducing

potential were up-regulated, including *mrtA*, *mvhAGD* and *hdrABC* genes. All genomes analysed possess a small hypothetical protein (AR505_1068, and corresponding homologues) adjacent to the *mrtA* gene, which in ISO4-H5 was regulated in a manner similar to *mrtA*, which suggests it could be associated with the Mrt complex, or the process of methyl CoM reduction. This study also revealed that the genes encoding methylamine and M3MSP utilisation were up-regulated by the presence of methylamine.

The analyses of *Methanobrevibacter* genomes revealed that the D5 genome is the second largest amongst the 18 available *Methanobrevibacter* spp. genomes analysed. D5 is predicted to utilise CO_2, H_2 and formate for hydrogenotrophic methanogenesis. The comparison between *Methanobrevibacter* spp. genomes revealed their genomes are highly conserved, with a "core" genome of 764 gene families, including all the genes involved in hydrogenotrophic methanogenesis and general housekeeping functions. A large number of genes encoding adhesin-like proteins and a NRPS were unique to D5. Therefore, it is hypothesised that differences in microbial interactions and the type of microbial partner may be an important clue to explain the coexistence of many different *Methanobrevibacter* spp. in the rumen. Furthermore, a number of *Methanobrevibacter* clade-specific genetic traits were identified in this study. Only members of the *Mbb. ruminantium* clade encode biotin biosynthesis, which suggests other *Methanobrevibacter* spp. acquire exogenous biotin via cross-feeding. Only members of the *Mbb. wolinii* clade have an aspartate racemase gene to produce D-aspartate, which suggests the cell envelope may contain D-aspartate. *Methanobrevibacter* strains $M1^T$, AbM4, YE315, SH^T, $JH1^T$, RFM-1^T, RFM-2^T, RFM-3^T, ATM^T and ANOR1 are predicted to only encode Mcr, and not Mrt complex, which may indicate a difference in regard to how they respond to different levels of H_2 and their effective H_2 thresholds.

7.4. Relevance of thesis findings to CH_4 formation in the rumen

The genome sequences generated and analysed in this thesis provided insights into the lifestyle of two very different rumen methanogens, a member of the Methanomassiliicoccales, methanogenic archaeon ISO4-H5 and a member of the *Methanobrevibacter* genus, *Methanobrevibacter* sp. D5. These methanogens have very different modes of methanogenesis which impacts on how they live and survive in the rumen. ISO4-H5 depends strictly on methylotrophic methanogenesis and is unable to disproportionate methyl-substrates. Using methanol as an example of a methyl-substrate, the theoretical standard free energy change ($\Delta G^{o\prime}$) from the reaction:

$$CH_3\text{-}OH + H_2 \rightarrow CH_4 + H_2O$$

is -112.5 kJ/mole (Thauer *et al.* 1977). On the other hand, D5 relies on hydrogenotrophic methanogenesis according to the following reaction:

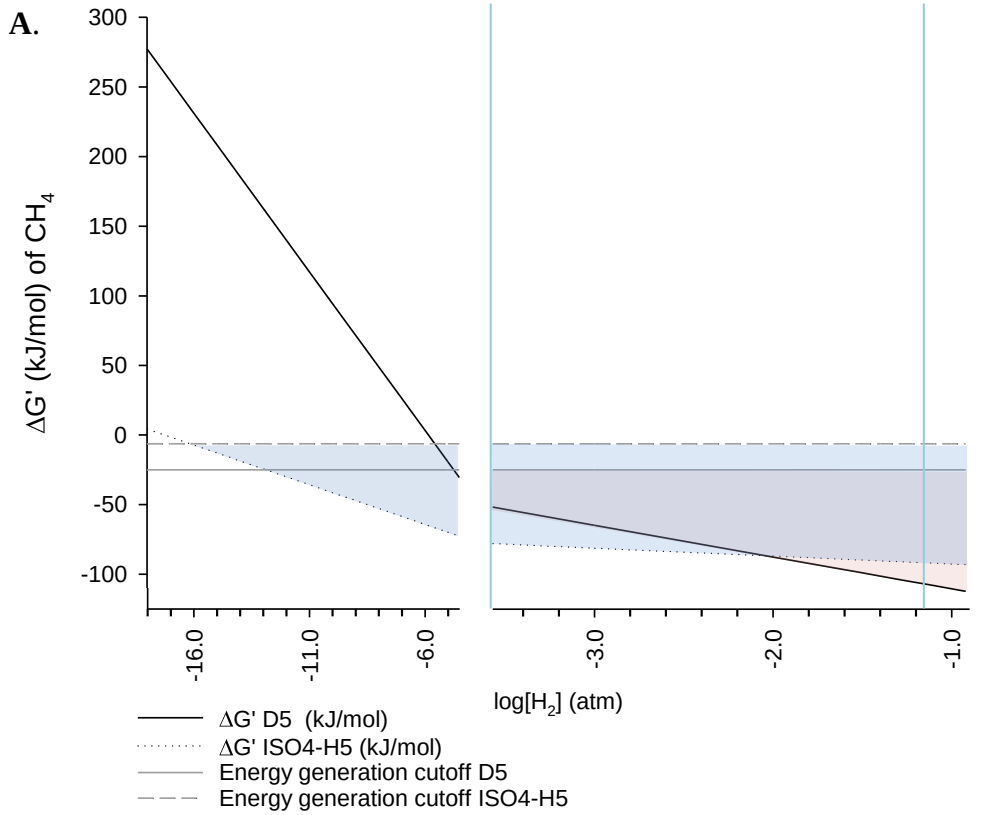
$$CO_2 + 4\ H_2 \rightarrow CH_4 + 2\ H_2O$$

The $\Delta G^{o\prime}$ from hydrogenotrophic methanogenesis is -131 kJ/mole (Thauer *et al.* 2008). Based on these stoichiometries, D5 and other hydrogenotrophic methanogens, require four moles of H_2 to produce one mole of CH_4, whereas ISO4-H5-like Methanomassiliicoccales only require one mole of H_2 to produce one mole of CH_4. This suggests that the Methanomassiliicoccales have a lower threshold of dissolved H_2 than the hydrogenotrophic methanogens. For methanogens to grow, the difference in free energy generated from methanogenesis needs to exceed the phosphorylation potential of ADP + P_i → ATP (Thauer *et al.* 2008). The free energy differential for methanogenesis also depends on the dissolved H_2 concentration (Janssen 2010). The H_2 in the rumen is produced by microbial fermentation of feed, it is dynamic and influenced by multiple factors that are interconnected, such as feed type, feed passage rate, ruminal pH, and the relative ratio of acetate and propionate production (Janssen 2010). Efficient removal of H_2 increases the rate of fermentation by eliminating the negative feedback inhibition imposed by H_2 (McAllister and Newbold 2008; Janssen 2010). The free energy change ($\Delta G'$) for ISO4-H5 and D5 under varying H_2 concentrations can be calculated using the following equations:

$$\text{D5: } \Delta G' = \Delta G^{o\prime} + 2.3RT\log [CH_4]/[H_2]^4\times[CO_2]$$

$$\text{ISO4-H5: } \Delta G' = \Delta G^{o\prime} + 2.3RT\log [CH_4]/[H_2]\times[CH_3OH]$$

Where R is the gas constant $8.314 JK^{-1}mol^{-1}$, T is absolute temperature 298.15°K.


A.
300
250
200
150
100
50
0
-50
-100
ΔG' (kJ/mol) of CH4
-16.0
-11.0
-6.0
-3.0
-2.0
-1.0
log[H2] (atm)
ΔG' D5 (kJ/mol)
ΔG' ISO4-H5 (kJ/mol)
Energy generation cutoff D5
Energy generation cutoff ISO4-H5

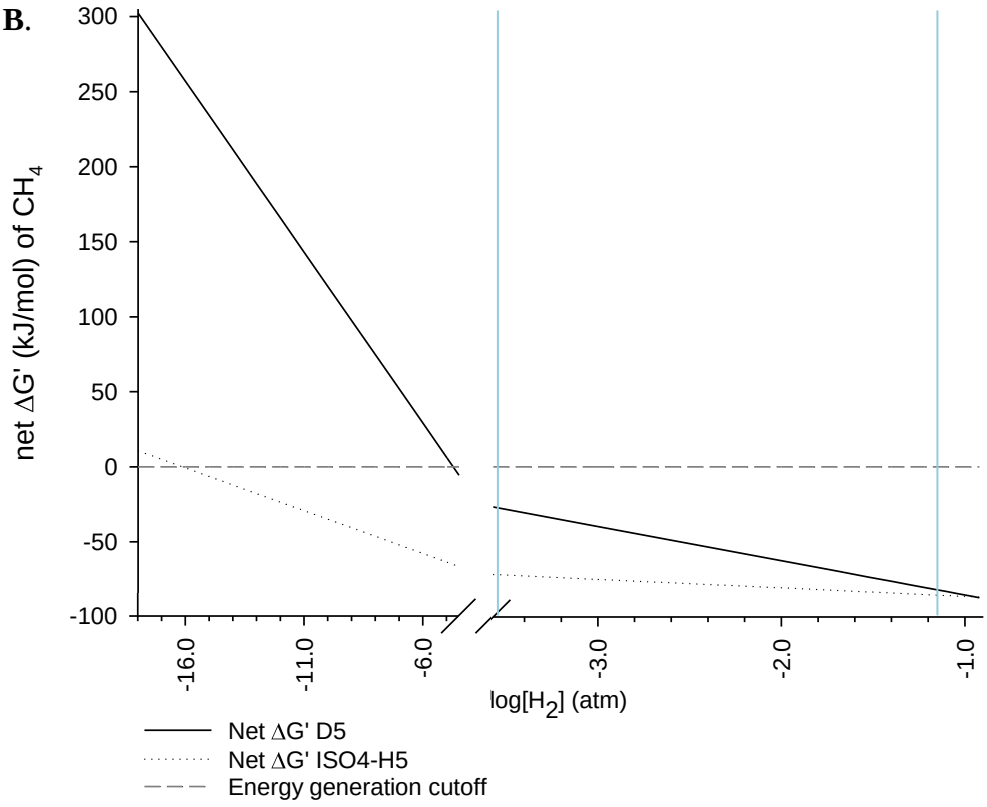

B.
300
250
200
150
100
50
0
-50
-100
net ΔG' (kJ/mol) of CH4
-16.0
-11.0
-6.0
-3.0
-2.0
-1.0
log[H2] (atm)
Net ΔG' D5
Net ΔG' ISO4-H5
Energy generation cutoff

Figure 7.1. Theoretical $\Delta G'$ of methanogenesis in relation to ATP synthesis in the rumen. $\Delta G^{o\prime}$ of $CO_2 + 4\ H_2 \rightarrow CH_4 + 2\ H_2O$ is -131 kJ/mole (Thauer *et al.* 2008), $\Delta G^{o\prime}$ of $CH_3\text{-}OH + H_2 \rightarrow CH_4 + H_2O$ is -112.5 kJ/mole (Thauer *et al.* 1977) under standard condition. Assuming the total gas pressure in the rumen is 1.01325×10^5 pa, comprising of CO_2:CH_4 of 65.35%:26.76%, [CO_2] would be 0.6535 atm, [CH_4] would be 0.2676 atm (McArthur and Miltimore 1961), [H_2] observed in the rumen is from 0.1 µM (0.000268atm, log [H_2] = -3.57) to 50 µM (0.067843 atm, log [H_2] = -1.17) as indicated between cyan lines (Janssen 2010). [CH_3OH] is 8.74E-4 mol/L in the rumen fluid of cows fed 40% hay 60% grain (Vantcheva *et al.* 1970). $\Delta G'$ of $ADP + P_i \rightarrow ATP$ is considered to be -50 kJ/mole, D5 generates approximately 0.5 mole of ATP per mole of CH_4, and ISO4-H5 generates 0.125 mole of ATP per mole of CH_4. **A**. Free energy change of ISO4-H5 and D5, blue and pink shaded area represents the $\Delta G'$ region in which ATP synthesis from ADP + Pi is possible for ISO4-H5 and D5 respectively. **B**. Net free energy change of ISO4-H5 and D5 after ATP biosynthesis is considered.

Under standard conditions, the methylotrophic methanogenesis in ISO4-H5 only needs one mole of H_2 to produce one mole of CH_4, and pumps one mole of ions every two moles of CH_4 due to electron bifurcation. The hydrogenotrophic methanogenesis in D5 needs four moles of H_2 to produce one mole of CH_4, and pumps two moles of ions protons for every mole of CH_4 produced (Thauer *et al.* 2008). Four moles of ions protons are required to drive one mole of ATP production. Assuming the energy required to make one mole of ATP is 50 kJ, ISO4-H5 diverts 6.25 kJ of free energy change to ATP synthesis for every mole of CH_4 produced, while D5 diverts 25 kJ of free energy change to ATP synthesis for every mole of CH_4 produced. If the free energies are plotted against the log H_2 concentration (Figure 7.1.), it shows that the concentration of dissolved H_2 typically present in the rumen is sufficient to support ATP generation from methanogenesis in both ISO4-H5 and D5. The higher the dissolved [H_2] becomes in the rumen, the more energetically favourable (more negative $\Delta G'$) the methanogenesis reaction becomes (Janssen 2010). This diagram also shows that compared to ISO4-H5, D5 has a narrower range of H_2 concentrations in which it can generate energy (Figure 7.1.). Dissolved H_2 can achieve a theoretical maximum of 737 µM, but inside the rumen H_2 is usually between 50 µM (0.067843 atm) and 0.1 µM (0.000268 atm) or lower (Janssen 2010). This suggests that under rumen conditions both ISO4-H5 and D5 are constantly exposed to sufficient concentrations of dissolved H_2 to permit methanogenesis. While D5 is predicted to generate more free energy change per mole of CH_4 than ISO4-H5 at H_2 concentration greater than 0.01 atm (approximately log[H_2] of -2, Figure 7.1A), when the energy requirements of ATP biosynthesis are considered, ISO4-H5 always has a greater net free energy change than D5 under rumen conditions (Figure 7.1B.). However, *Methanobrevibacter* spp. are the predominant rumen methanogens (Janssen and Kirs 2008; Henderson *et al.* 2015; Seedorf *et al.* 2015) and thus must be able to grow well under the conditions that prevail in the rumen. The fact that members of the Methanomassiliicoccales are not the dominate rumen methanogens suggests that their growth is limited by factors other than dissolved H_2 and is most likely governed by the availability of methyl substrates.

Despite the prediction that H_2 concentration is unlikely to limit Methanomassiliicoccales growth, transcriptomic analysis revealed up-regulation of ISO4-H5 *hdrABC* and *mvhAGD* genes in the low H_2 environment, presumably as a response to scavenge more H_2. This suggests that the decrease of ΔG' for ATP generation by low H_2 concentration is significant to the ISO4-H5 cells, and rather than produce less ATP, ISO4-H5 appears to adapt to the low H_2 concentration by up-regulating genes to scavenge H_2. This shows that ISO4-H5 maximises the gain of energy from the rumen environment under changing conditions and likely reflects the competitive nature of the rumen.

While ISO4-H5 and D5 showed clear differences in their predicted H_2 thresholds and their different ruminal niches, the difference between methanogens of the same order or genus is less apparent. Most of the differences between Methanomassiliicoccales appear to be associated with genes encoding cell envelope surface features rather than central metabolism. ISO4-H5 is predicted to encode a range of adhesin-like proteins that may mediate binding to a variety of surfaces. It also encodes a cluster of genes involved in exopolysaccharide biosynthesis that is not found in other members of the Methanomassiliicoccales, which may also be involved in cell surface modifications. Similarly, D5 is predicted to encode 50 adhesin-like proteins that are not found in other *Methanobrevibacter* spp. The interaction and binding of methanogens to other microorganisms in the rumen or to particular tissues in the ruminant digestive tract, could be important in the establishment of unique methanogen niches. D5 and a few other *Methanobrevibacter* spp. also encode NRPSs that may produce antimicrobial peptides; if so, these antimicrobial peptides may play a part in competition between similar methanogens, and thus influence the success of the producing species.

7.5. Future perspectives

The genome sequence of members of the order Methanomassiliicoccales and the genus *Methanobrevibacter* have offered insights into their lifestyles in the rumen. However, the information inferred from genomes alone is not the full picture and many factors can influence the behaviour of genes, such as transcriptional and translational regulation, post-translational modification, therefore the genomic information serve as a starting point to investigate genes or metabolic traits of interest. Exactly how these genes bring about the metabolic pathway *in vivo*, may only be verified by experimental investigation. For example, the genomes of *Methanobrevibacter* spp. revealed confounding results inconsistent to published literature in the utilisation of formate and requirement of CoM. All genomes analysed possess the necessary

genes to utilise formate, yet the SH^T, HO^T, RFM-2^T, RFM-3^T and ATM^T are observed to not utilise formate (Leadbetter and Breznak 1996; Leadbetter *et al.* 1998; Miller and Lin 2002; Savant *et al.* 2002). The $M1^T$ genome is the only genome to lack the full complement of genes required to produce CoM and requires external CoM to grow (Taylor *et al.* 1974), however SH^T also requires external CoM for growth and the growth of $JH1^T$ is stimulated by CoM supplementation (Miller and Lin 2002; Lee *et al.* 2013). These results highlights the importance of culture based experimental validation, it is clear that findings from bioinformatics approaches require biological validation to be meaningful.

Methanogen growth requirements should be conducted in pure cultures, without the uncertainties posed by other microorganisms within the enrichment culture. To date, ISO4-H5 and ISO4-G1 are the only members of the rumen Methanomassiliicoccales for which pure cultures have been obtained. This suggests there should be a renewed attempt to isolate pure cultures of new rumen methanogens, as well as enrichment of previously unculturable members of the Methanomassiliicoccales. Furthermore, many available genome sequences remain in draft status, and completing the genomes by gap closure would reduce the uncertainty posed by a draft genome. Furthermore, the available genome sequences only represent a small proportion of the rumen Methanomassiliicoccales population (Seedorf *et al.* 2014), and in the future, a high resolution metagenome sequencing of the rumen microbiome may enable genome assembly of previously uncultivated rumen methanogens.

While the theoretical calculations of $\Delta G'$ is useful for defining the thermodynamic boundaries of methanogen growth in the rumen, they are limited to the assumptions applied, and cannot fully portray the actual exchanges that are likely to occur between rumen methanogens and their co-resident microbes. For example, it does not account for factors such as high local H_2 concentrations resulting from methanogens adherence to H_2 producers or the possibility of interspecies electron transfer. Within the range of dissolved H_2 observed in the rumen, thermodynamics favours CH_4 formation from methyl compounds and ISO4-H5 is predicted to be better able to carry out ATP generation than D5 (Figure 7.1.). However, *Methanobrevibacter* spp., mainly the *Mbb. gottschalkii* clade, are the predominant methanogens and not members of the order Methanomassiliicoccales. The ruminal level of methyl substrates, such as methanol and methylamines, may limit the proliferation of Methanomassiliicoccales in the rumen. If this is true, feed supplementation with methyl-substrates may elevate the rumen Methanomassiliicoccales population to become the predominant rumen methanogen. Methanol arises in the rumen from cleavage of methylated compounds such as pectin (Siragusa *et al.*

1988), The *in vitro* supplementation of low esterified citrus pectin reduced the CH_4 production by 19%, while high esterified citrus pectin increased the CH_4 production by 7% (Geerkens *et al.* 2013). Methylamines originate from breakdown of methylated nitrogen compounds such as betaine and choline (Neill *et al.* 1978). Radiolabeled betaine supplementation *in vivo* gave rise to trimethylamine, CH_4 and CO_2 (Mitchell *et al.* 1979), but the methanogen populations were not assessed in these studies. The Tibetan yak was found to have 80.9% of its rumen methanogen population as members of the Methanomassiliicoccales (Huang *et al.* 2012), and while this was mainly attributed from the high tannin level in the alpine *Kobresia* pasture, the level of methyl-substrate from this forage has not been determined and may also contribute to the high Methanomassiliicoccales population. The logical next step to assess whether additional methyl-substrates stimulate Methanomassiliicoccales would be an *in vivo* trial involving prolonged feeding of methylamine compounds followed by assessment of methanogen populations. Assuming all other growth requirements are fulfilled, the hypothesis would be that the rumen Methanomassiliicoccales should proliferate to higher levels than the *Methanobrevibacter* spp.

The *in vitro* growth of ISO4-H5, ISO4-G1 and ISO4-G11 are all very slow, even with excess methanol, and they all produce comparatively low amounts of CH_4 compared to the *Methanobrevibacter* spp. Although this study showed ISO4-H5 utilised external ethanol and nicotinamide, it is obvious that there are additional nutritional requirements of rumen Methanomassiliicoccales that remain to be identified, therefore a defined medium which allows the growth requirements of Methanomassiliicoccales to be determined, would be useful in this regard.

A strong association between Succinivibrionaceae and Methanomassiliicoccaceae was found in a global census of the rumen microbiome (Henderson *et al.* 2015). This association may be explained, at least partially, by the release of methanol by *Succinivibrio* spp. from the degradation of pectin (Dehority 1969), but this study has demonstrated that methanol is not the limiting condition *in vitro*. Therefore, the association of Succinivibrionaceae with Methanomassiliicoccaceae deserves further investigation, to better understand the nature of this association and define the metabolites and/or interactions that underlie this relationship. Transcriptomic analyses of *Succinivibrio dextrinosolvens* H5 in pure and co-culture studies with Methanomassiliicoccales would be useful in this regard.

Despite the large differences between members of Methanomassiliicoccales and *Methanobrevibacter* spp., the central Mcr/Mrt and CoM-dependent methyl group reduction remains strongly conserved, as is found in all known methanogens to date. The rumen members of Methanomassiliicoccales depend on exogenous supply of CoM to survive as they are incapable of *de novo* CoM biosynthesis, therefore structural analogues of CoM would still be inhibitory to these methanogens. Previously, the CoM analogue, BES did not show lasting inhibition of methanogens *in vivo* (Immig *et al.* 1996). It is possible that BES caused the inhibition of CoM auxotrophic methanogens only and led to the drop in CH_4 production, but methanogens capable of *de novo* CoM biosynthesis were less susceptible and filled the vacant niche after a short time. It is possible that the biosynthetic pathway of CoM may be a better target for inhibition, especially the ComDE that are found in the Methanomassiliicoccales Mx1, Mx1201 and B10. If the *de novo* biosynthesis of CoM is inhibited, the Methanomassiliicoccales would initially remain unaffected, but as inhibition persisted and exogenous CoM is diluted by flow feed and water through the rumen, methanogens may eventually be inhibited. Persistant inhibition of CoM biosynthesis, if applied together with CoM analogue such as BES, may be synergistic and a more effective inhibition strategy.

While the transcriptome of ISO4-H5 illustrated the adaptation of ISO4-H5 to a low H_2 environment by up-regulation of *mrtA*, *hdr* and *mvh* genes, it also raised more interesting questions. For example, why are the other subunits of the Mrt complex not up-regulated in the same manner (Poulsen *et al.* 2013) and what is the function of the small hypothetical protein encoded adjacent to the *mrtA* gene? Previously the *mcr/mrt* operons from the rumen of sheep have been classified into three distinct groups (Shi *et al.* 2014). The *mrt* operons of ISO4-H5 and members of order Methanomassiliicoccales are distinct to the *mcr* and *mrt* of the Methanobacteriales. These observations suggests that the molecular mechanism of the Mrt complex in the order Methanomassiliicoccales behaves differently and warrants further investigation. The transcriptomic analyses of ISO4-H5 in the co-culture studies also hinted a possibility of direct interaction with *Ruminococcus flavefaciens* FD1. The interspecies interactions and cross-feeding between methanogens and H_2 producers is an important process that strongly influences CH_4 formation and deserves further investigation.

References

Abken, H. J., M. Tietze, J. Brodersen, S. Baumer, U. Beifuss and U. Deppenmeier (1998). Isolation and characterization of methanophenazine and function of phenazines in membrane-bound electron transport of *Methanosarcina mazei* Go1. J Bacteriol 180(8): 2027-2032.

Abu-Qarn, M., S. Yurist-Doutsch, A. Giordano, A. Trauner, H. R. Morris, P. Hitchen, O. Medalia, A. Dell and J. Eichler (2007). *Haloferax volcanii* AglB and AglD are involved in *N*-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. J Mol Biol 374(5): 1224-1236.

Albers, S. V. and B. H. Meyer (2011). The archaeal cell envelope. Nat Rev Microbiol 9(6): 414-426.

Alikhan, N. F., N. K. Petty, N. L. Ben Zakour and S. A. Beatson (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12: 402.

Alkalaeva, E., B. Eliseev, A. Ambrogelly, P. Vlasov, F. A. Kondrashov, S. Gundllapalli, L. Frolova, D. Soll and L. Kisselev (2009). Translation termination in pyrrolysine-utilizing archaea. FEBS Lett 583(21): 3455-3460.

Allison, C. and G. T. Macfarlane (1989). Influence of pH, nutrient availability, and growth rate on amine production by *Bacteroides fragilis* and *Clostridium perfringens*. Appl Environ Microbiol 55(11): 2894-2898.

Allocati, N., M. Masulli, C. Di Ilio and V. De Laurenzi (2015). Die for the community: an overview of programmed cell death in bacteria. Cell Death Dis 6: e1609. Retrieved from http://www.nature.com/cddis/journal/v6/n1/full/cddis2014570a.html.

Altermann, E. (2012). Tracing lifestyle adaptation in prokaryotic genomes. Front Microbiol 3: 48.

Altermann, E. and T. R. Klaenhammer (2003). GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. OMICS 7(2): 161-169.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic Local Alignment Search Tool. J Mol Biol 215(3): 403-410.

Ameri, A., D. K. Machiah, T. T. Tran, C. Channell, V. Crenshaw, K. Fernstrom, M. Khachidze, A. Duncan, S. Fuchs and T. E. Howard (2007). A nonstop mutation in the factor (F)X gene of a severely haemorrhagic patient with complete absence of coagulation FX. Thromb Haemost 98(6): 1165-1169.

Anders, S. and W. Huber (2010). Differential expression analysis for sequence count data. Genome Biol 11(10). Retrieved from http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-10-r106.

Andreesen, J. R. and L. G. Ljungdahl (1974). Nicotinamide adenine dinucleotide phosphate-dependent formate dehydrogenase from *Clostridium thermoaceticum*: purification and properties. J Bacteriol 120(1): 6-14.

Anderson, M. J., K. E. Ellingsen and B. H. McArdle (2006). Multivariate dispersion as a measure of beta diversity. Ecol Lett 9(6): 683-693.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arbing, M. A., S. Chan, A. Shin, T. Phan, C. J. Ahn, L. Rohlin and R. P. Gunsalus (2012). Structure of the surface layer of the methanogenic archaean *Methanosarcina acetivorans*. Proc Natl Acad Sci U S A 109(29): 11812-11817.

Arima, K., A. Kakinuma and G. Tamura (1968). Surfactin, a crystalline peptidelipid surfactant produced by *Bacillus subtilis*: isolation, characterization and its inhibition of fibrin clot formation. Biochem Biophys Res Commun 31(3): 488-494.

Arnesen, L. P. S., A. Fagerlund and P. E. Granum (2008). From soil to gut: *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol Rev 32(4): 579-606.

Attwood, G. T., W. J. Kelly, E. H. Altermann and S. C. Leahy (2008). Analysis of the *Methanobrevibacter ruminantium* draft genome: Understanding methanogen biology to inhibit their action in the rumen. Aust J Exp Agric 48(1-2): 83-88.

Attwood, G. T., W. J. Kelly, E. H. Altermann, C. D. Moon, S. Leahy and A. L. Cookson (2008). Application of rumen microbial genome information to livestock systems in the postgenomic era. Aust J Exp Agric 48(7): 695-700.

Aung, H. L., Dey, P. H. Janssen, R. S. Ronimus and G. M. Cook (2015). A high-throughput screening assay for identification of inhibitors of the A_1A_0-ATP synthase of the rumen methanogen *Methanobrevibacter ruminantium* M1. J Microbiol Meth 110: 15-17.

Avissar, Y. J. and S. I. Beale (1989). Identification of the enzymatic basis for Δ-aminolevulinic acid auxotrophy in a *hemA* mutant of *Escherichia coli*. J Bacteriol 171(6): 2919-2924.

Azad, R. K. and J. G. Lawrence (2007). Detecting laterally transferred genes: use of entropic clustering methods and genome position. Nucleic Acids Res 35(14): 4629-4639.

Bachmann, B. O. and J. Ravel (2009). Chapter 8. Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Meth Enzymol 458: 181-217.

Bailey, C. B. and C. C. Balch (1961). Saliva secretion and its relation to feeding in cattle. 1. The composition and rate of secretion of parotid saliva in a small steer. Br J Nutr 15: 371-382.

Balch, W. E. and R. S. Wolfe (1976). New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl Environ Microbiol 32(6): 781-791.

Balch, W. E. and R. S. Wolfe (1979). Transport of coenzyme M (2-mercaptoethanesulfonic acid) in *Methanobacterium ruminantium*. J Bacteriol 137(1): 264-273.

Balch, W. E., G. E. Fox, L. J. Magrum, C. R. Woese and R. S. Wolfe (1979). Methanogens: reevaluation of a unique biological group. Microbiol Rev 43(2): 260-296.

Balint, M., S. Domisch, C. H. M. Engelhardt, P. Haase, S. Lehrian, J. Sauer, K. Theissinger, S. U. Pauls and C. Nowak (2011). Cryptic biodiversity loss linked to global climate change. Nat Clim Change 1(6): 313-318.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev and P. A. Pevzner (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5): 455-477.

Bapteste, E., C. Brochier and Y. Boucher (2005). Higher-level classification of the Archaea: Evolution of methanogenesis and methanogens. Archaea 1(5): 353-363.

Barenkamp, S. J. (1996). Immunization with high-molecular-weight adhesion proteins of nontypable *Haemophilus influenzae* modifies experimental otitis media in chinchillas. Infect Immun 64(4): 1246-1251.

Barenkamp, S. J. and J. W. St Geme, 3rd (1996). Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable *Haemophilus influenzae*. Mol Microbiol 19(6): 1215-1223.

Baumer, S., T. Ide, C. Jacobi, A. Johann, G. Gottschalk and U. Deppenmeier (2000). The $F_{420}H_2$ dehydrogenase from *Methanosarcina mazei* is a redox-driven proton pump closely related to NADH dehydrogenases. J Biol Chem 275(24): 17968-17973.

Baysse, C., S. Matthijs, T. Pattery and P. Cornelis (2001). Impact of mutations in *hemA* and *hemH* genes on pyoverdine production by *Pseudomonas fluorescens* ATCC17400. FEMS Microbiol Lett 205(1): 57-63.

Beauchemin, K. A., M. Kreuzer, F. O'mara and T. A. McAllister (2008). Nutritional management for enteric methane abatement: a review. Anim Prod Sci 48(2): 21-27.

Beauchemin, K. A. and S. M. McGinn (2005). Methane emissions from feedlot cattle fed barley or corn diets. J Anim Sci 83(3): 653-661.

Beauchemin, K. A. and S. M. McGinn (2006). Methane emissions from beef cattle: effects of fumaric acid, essential oil, and canola oil. J Anim Sci 84(6): 1489-1496.

Beauchemin, K. A., S. M. McGinn, T. F. Martinez and T. A. McAllister (2007). Use of condensed tannin extract from quebracho trees to reduce methane emissions from cattle. J Anim Sci 85(8): 1990-1996.

Begley, M., P. D. Cotter, C. Hill and R. P. Ross (2009). Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. Appl Environ Microbiol 75(17): 5451-5460.

Belfaiza, J., A. Fazel, M. Klaus and G. N. Cohen (1984). *E. coli* aspartokinase II-homoserine dehydrogenase II polypeptide chain has a triglobular structure. Biochem Biophys Res Commun 123(1): 16-20.

Bellais, S., M. Arthur, L. Dubost, J. E. Hugonnet, L. Gutmann, J. van Heijenoort, R. Legrand, J. P. Brouard, L. Rice and J. L. Mainardi (2006). Aslfm, the D-aspartate ligase responsible for

the addition of D-aspartic acid onto the peptidoglycan precursor of *Enterococcus faecium*. J Biol Chem 281(17): 11586-11594.

Bender, D. A. (2012). The Aromatic Amino Acids: Phenylalanine, Tyrosine and Tryptophan. Amino Acid Metabolism, *Third Edition*, Online, John Wiley & Sons, Ltd: 323-376. Retrieved from http://onlinelibrary.wiley.com/book/10.1002/9781118357514.

Benjamini, Y. (2010). Discovering the false discovery rate. J Royal Stat Soc: Ser B (Stat Meth) 72(4): 405-416.

Berg, I. A., D. Kockelkorn, W. H. Ramos-Vera, R. F. Say, J. Zarzycki, M. Hügler, B. E. Alber and G. Fuchs (2010). Autotrophic carbon fixation in archaea. Nat Rev Micro 8(6): 447-460.

Berge, O., T. Heulin, W. Achouak, C. Richard, R. Bally and J. Balandreau (1991). *Rahnella aquatilis*, a nitrogen-fixing enteric bacterium associated with the rhizosphere of wheat and maize. Can J Microbiol 37(3): 195-203.

Berk, H. and R. K. Thauer (1997). Function of coenzyme F_{420}-dependent NADP reductase in methanogenic archaea containing an NADP-dependent alcohol dehydrogenase. Arch Microbiol 168(5): 396-402.

Bernstein, M., F. Kepes and R. Schekman (1989). Sec59 encodes a membrane protein required for core glycosylation in *Saccharomyces cerevisiae*. Mol Cell Biol 9(3): 1191-1199.

Berven, F. S., O. A. Karlsen, A. H. Straume, K. Flikka, J. C. Murrell, A. Fjellbirkeland, J. R. Lillehaug, I. Eidhammer and H. B. Jensen (2006). Analysing the outer membrane subproteome of *Methylococcus capsulatus* (Bath) using proteomics and novel biocomputing tools. Arch Microbiol 184(6): 362-377.

Betley, J. N., M. C. Frith, J. H. Graber, S. Choo and J. O. Deshler (2002). A ubiquitous and conserved signal for RNA localization in chordates. Curr Biol 12(20): 1756-1761.

Beulin, D. S., M. Yamaguchi, S. Kawabata and K. Ponnuraj (2014). Crystal structure of PfbA, a surface adhesin of *Streptococcus pneumoniae*, provides hints into its interaction with fibronectin. Int J Biol Macromol 64: 168-173.

Biemont, C. and C. Vieira (2006). Genetics: Junk DNA as an evolutionary force. Nature 443(7111): 521-524.

Bird, S. H., R. S. Hegarty and R. Woodgate (2008). Persistence of defaunation effects on digestion and methane production in ewes. Aust J Exp Agric 48(1-2): 152-155.

Biswas, A., J. N. Gagnon, S. J. Brouns, P. C. Fineran and C. M. Brown (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. RNA Biol 10(5): 817-827.

Blanche, F., C. Robin, M. Couder, D. Faucher, L. Cauchois, B. Cameron and J. Crouzet (1991). Purification, characterization, and molecular cloning of *S*-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from *Methanobacterium ivanovii*. J Bacteriol 173(15): 4637-4645.

Bray, J. R. and J. T. Curtis (1957). An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27(4): 325-349.

Bodegom, P. M., J. C. Scholten and A. J. Stams (2004). Direct inhibition of methanogenesis by ferric iron. FEMS Microbiol Ecol 49(2): 261-268.

Bonacker, L. G., S. Baudner, E. Morschel, R. Bocher and R. K. Thauer (1993). Properties of the two isoenzymes of methyl-coenzyme M reductase in *Methanobacterium thermoautotrophicum*. Eur J Biochem 217(2): 587-595.

Boomker, E. A (2000). Ruminant physiology: digestion, metabolism, growth, and reproduction. P. Cronjé, New York, USA, CABI Publishing.

Borrel, G., N. Gaci, P. Peyret, P. W. O'Toole, S. Gribaldo and J. F. Brugere (2014). Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. Archaea 2014: 374146. Retrieved from http://downloads.hindawi.com/journals/archaea/2014/374146.pdf.

Borrel, G., H. M. Harris, W. Tottey, A. Mihajlovski, N. Parisot, E. Peyretaillade, P. Peyret, S. Gribaldo, P. W. O'Toole and J. F. Brugere (2013). Genome sequence of "*Candidatus* Methanomassiliicoccus intestinalis" Issoire-Mx1, a third thermoplasmatales-related methanogenic archaeon from human feces. Genome Announc 1(4): e00453-13. Retrieved from http://genomea.asm.org/content/1/4/e00453-13.full.pdfJul.

Borrel, G., H. M. Harris, W. Tottey, A. Mihajlovski, N. Parisot, E. Peyretaillade, P. Peyret, S. Gribaldo, P. W. O'Toole and J. F. Brugere (2012). Genome sequence of "*Candidatus*

Methanomethylophilus alvus" Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. J Bacteriol 194(24): 6944-6945.

Borrel, G., P. W. O'Toole, H. M. Harris, P. Peyret, J. F. Brugere and S. Gribaldo (2013). Phylogenomic data support a seventh order of Methylotrophic methanogens and provide insights into the evolution of Methanogenesis. Genome Biol Evol 5(10): 1769-1780.

Borrel, G., N. Parisot, H. M. Harris, E. Peyretaillade, N. Gaci, W. Tottey, O. Bardot, K. Raymann, S. Gribaldo, P. Peyret, P. W. O'Toole and J. F. Brugere (2014). Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic Archaea that encodes pyrrolysine. BMC Genomics 15(1): 679. Retrieved from http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-679.

Bose, A., M. A. Pritchett, M. Rother and W. W. Metcalf (2006). Differential regulation of the three methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A. J Bacteriol 188(20): 7274-7283.

Bose, A., M. A. Pritchett and W. W. Metcalf (2008). Genetic analysis of the methanol- and methylamine-specific methyltransferase 2 genes of *Methanosarcina acetivorans* C2A. J Bacteriol 190(11): 4017-4026.

Boudsocq, F., S. Iwai, F. Hanaoka and R. Woodgate (2001). *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. Nucleic Acids Res 29(22): 4607-4616.

Bower, S., J. B. Perkins, R. R. Yocum, C. L. Howitt, P. Rahaim and J. Pero (1996). Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. J Bacteriol 178(14): 4122-4130.

Bray, J. R. and J. T. Curtis (1957). An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr 27(4): 325-349.

Bremer, H. and P. P. Dennis (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. EcoSal Plus 3(1). Retrieved from http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosal.5.2.3.

Brodsky, L., A. Moussaieff, N. Shahaf, A. Aharoni and I. Rogachev (2010). Evaluation of peak picking quality in LC-MS metabolomics data. Anal Chem 82(22): 9177-9187.

Brosius, J., T. J. Dull, D. D. Sleeter and H. F. Noller (1981). Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J Mol Biol 148(2): 107-127.

Brügger, K., P. Redder, Q. She, F. Confalonieri, Y. Zivanovic and R. A. Garrett (2002). Mobile elements in archaeal genomes. FEMS Microbiol Lett 206(2): 131-141.

Brune, A. (1998). Termite guts: the world's smallest bioreactors. Trends Biotechnol 16(1): 16-21.

Bryant, M. P. (1965). Rumen methanogenic bacteria. Physiology of digestion of the ruminant. R. W. Dehority, R. S. Allen, W. Burroughs, N. L. Jacobsen and A. D. McGilliard, Washington, D.C, Butterworths Publishing Inc.

Bryant, M. P. and N. Small (1956). Characteristics of two new genera of anaerobic curved rods isolated from the rumen of cattle. J Bacteriol 72(1): 22-26.

Bryant, M. P., L. L. Campbell, C. A. Reddy and M. R. Crabill (1977). Growth of *Desulfovibrio* in lactate or ethanol media low in sulfate in association with H_2-utilizing methanogenic bacteria. Appl Environ Microbiol 33(5): 1162-1169.

Bryant, M., S. Tzeng, I. Robinson and A. Joyner Jr (1971). Nutrient requirements of methanogenic bacteria. Urbana 101: 61801.

Buan, N. R. and W. W. Metcalf (2010). Methanogenesis by *Methanosarcina acetivorans* involves two structurally and functionally distinct classes of heterodisulfide reductase. Mol Microbiol 75(4): 843-853.

Buckel, W. and R. K. Thauer (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Biochim Biophys Acta 1827(2): 94-113.

Buddle, B. M., M. Denis, G. T. Attwood, E. Altermann, P. H. Janssen, R. S. Ronimus, C. S. Pinares-Patiño, S. Muetzel and D. N. Wedlock (2011). Strategies to reduce methane emissions from farmed ruminants grazing on pasture. Vet J 188(1): 11-17.

Buhk, H. J., B. Behrens, R. Tailor, K. Wilke, J. J. Prada, U. Gunthert, M. Noyer-Weidner, S. Jentsch and T. A. Trautner (1984). Restriction and modification in *Bacillus subtilis*: nucleotide sequence, functional organization and product of the DNA methyltransferase gene of bacteriophage SPR. Gene 29(1-2): 51-61.

Bullard, J. H., E. Purdom, K. D. Hansen and S. Dudoit (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94. Retrieved from http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-94.

Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghagen and J. C. Venter (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273(5278): 1058-1073.

Burke, S. A. and J. A. Krzycki (1997). Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. J Biol Chem 272(26): 16570-16577.

Burrows, M. and D. J. Wheeler (1994). A block-sorting lossless data compression algorithm. SRC 124. Retrieved from http://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-124.pdf.

Bygrave, F. L. and R. M. Dawson (1976). Phosphatidylcholine biosynthesis and choline transport in the anaerobic protozoon *Entodinium caudatum*. Biochem J 160(3): 481-490.

Cabanes, D., P. Dehoux, O. Dussurget, L. Frangeul and P. Cossart (2002). Surface proteins and the pathogenic potential of *Listeria monocytogenes*. Trends Microbiol 10(5): 238-245.

Caboche, S., M. Pupin, V. Leclere, A. Fontaine, P. Jacques and G. Kucherov (2008). NORINE: a database of nonribosomal peptides. Nucleic Acids Res 36(Database issue): D326-331. Retrieved from http://nar.oxfordjournals.org/content/36/suppl_1/D326.full.

Canchaya, C., C. Proux, G. Fournous, A. Bruttin and H. Brussow (2003). Prophage genomics. Microbiol Mol Biol Rev 67(2): 238-276.

Candlish, E., T. J. Devlin and L. J. LaCroix (1970). Tryptophan utilization by rumen microorganisms *in vitro*. Can J Anim Sci 50(2): 331-335.

Cappenberg, T. E. (1974). Interrelations between sulfate-reducing and methane-producing bacteria in bottom deposits of a fresh-water lake. II. Inhibition experiments. Antonie van Leeuwenhoek 40(2): 297-306.

Carulla, J., M. Kreuzer, A. Machmüller and H. D. Hess (2005). Supplementation of *Acacia mearnsii* tannins decreases methanogenesis and urinary nitrogen in forage-fed sheep. Crop Pasture Sci 56(9): 961-970.

Carver, T., N. Thomson, A. Bleasby, M. Berriman and J. Parkhill (2009). DNAPlotter: circular and linear interactive genome visualization. Bioinformatics (Oxford, England) 25(1): 119-120.

Casali, N. and A. Preston (2003). *E. coli* plasmid vectors: Methods and applications 235. Totowa, Humana Press.

Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 42(Database issue): D459-471. Retrieved from http://nar.oxfordjournals.org/content/42/D1/D459.long.

Chagan, I., M. Tokura, J. P. Jouany and K. Ushida (1999). Detection of methanogenic archaea associated with rumen ciliate protozoa. J Gen Appl Microbiol 45(6): 305-308.

Chant, J. and P. Dennis (1986). Archaebacteria: transcription and processing of ribosomal RNA sequences in *Halobacterium cutirubrum*. EMBO J 5(5): 1091-1097.

Chaudhary, P. P., S. K. Sirohi, D. Singh and J. Saxena (2011). Methyl coenzyme M reductase (*mcrA*) gene based phylogenetic analysis of methanogens population in Murrah buffaloes (*Bubalus bubalis*). J Microbiol 49(4): 558-561.

Cheeseman, P., A. Toms-Wood and R. S. Wolfe (1972). Isolation and properties of a fluorescent compound, $Factor_{420}$, from *Methanobacterium* strain M.o.H. J Bacteriol 112(1): 527-531.

Chen, H., N. Huang and Z. Sun (2006). SubLoc: a server/client suite for protein subcellular location based on SOAP. Bioinformatics 22(3): 376-377.

Chen, S., L. Wang and Z. Deng (2010). Twenty years hunting for sulfur in DNA. Protein Cell 1(1): 14-21.

Christensen, S. K. and K. Gerdes (2003). RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. Mol Microbiol 48(5): 1389-1400.

Church, J. A. and N. J. White (2011). Sea-level rise from the late 19th to the early 21st century. Surv Geophys 32(4): 585-602.

Clark, H., F. Kelliher and C. Pinares-Patino (2011). Reducing CH_4 emissions from grazing ruminants in New Zealand: challenges and opportunities. Asian Aust J Anim Sci 24(2): 295-302.

Conrad, R., C. Erkel and W. Liesack (2006). Rice Cluster I methanogens, an important group of Archaea producing greenhouse gas in soil. Curr Opin Biotechnol 17(3): 262-267.

Corcelli, A. (2009). The cardiolipin analogues of Archaea. Biochim Biophys Acta 1788(10): 2101-2106.

Craciun, S. and E. P. Balskus (2012). Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. Proc Natl Acad Sci U S A 109(52): 21307-21312.

Creevey, C. J., W. J. Kelly, G. Henderson and S. C. Leahy (2014). Determining the culturability of the rumen bacterial microbiome. Microb Biotechnol 7(5): 467-479.

Cubasch, U., D. Wuebbles, D. Chen, M. C. Facchini, D. Frame, N. Mahowald and J. G. Winther (2013). Introduction. Climate Change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P. M. Midgley. Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press: 119–158.

Cunin, R., N. Glansdorff, A. Pierard and V. Stalon (1986). Biosynthesis and metabolism of arginine in bacteria. Microbiol Rev 50(3): 314-352.

Currie, D. H., B. Raman, C. M. Gowen, T. J. Tschaplinski, M. L. Land, S. D. Brown, S. F. Covalla, D. M. Klingeman, Z. K. Yang, N. L. Engle, C. M. Johnson, M. Rodriguez, A. J. Shaw, W. R. Kenealy, L. R. Lynd, S. S. Fong, J. R. Mielenz, B. H. Davison, D. A. Hogsett and C. D. Herring (2015). Genome-scale resources for *Thermoanaerobacterium saccharolyticum*. BMC Syst Biol 9: 30. Retrieved from http://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-015-0159-x.

D'Orval, B. C., M. L. Bortolin, C. Gaspin and J. P. Bachellerie (2001). Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The $tRNA^{Trp}$ intron guides the formation of two ribose-methylated nucleosides in the mature $tRNA^{Trp}$. Nucleic Acids Res 29(22): 4518-4529.

D'Souza, G., S. Waschina, S. Pande, K. Bohl, C. Kaleta and C. Kost (2014). Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution 68(9): 2559-2570.

Dabard, J., C. Bridonneau, C. Phillipe, P. Anglade, D. Molle, M. Nardi, M. Ladire, H. Girardin, F. Marcille, A. Gomez and M. Fons (2001). Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. Appl Environ Microbiol 67(9): 4111-4118.

De Poorter, L. M., W. G. Geerts, A. P. Theuvenet and J. T. Keltjens (2003). Bioenergetics of the formyl-methanofuran dehydrogenase and heterodisulfide reductase reactions in *Methanothermobacter thermautotrophicus*. Eur J Biochem 270(1): 66-75.

De Vrije, T. and P. A. M. Claassen (2003). Dark hydrogen fermentations. Bio-methane & Bio-hydrogen. J. H. Reith, R. H. Wijffels and H. Barten. Netherlands, Dutch Biological Hydrogen Foundation: 103-123.

Dehority, B. A. (1969). Pectin-fermenting bacteria isolated from the bovine rumen. J Bacteriol 99(1): 189-196.

Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. Cell 89(7): 1007-1010.

Dejean, A., P. Sonigo, S. Wain-Hobson and P. Tiollais (1984). Specific hepatitis B virus integration in hepatocellular carcinoma DNA through a viral 11-base-pair direct repeat. Proc Natl Acad Sci U S A 81(17): 5350-5354.

Dejean, S., I. Gonzalez and M. K. Le Cao (2011). Package 'mixOmics'. Image 9: 1. Retrieved from http://camoruco.ing.uc.edu.ve/cran/web/packages/mixOmics/mixOmics.pdf.

Delcher, A. L., D. Harmon, S. Kasif, O. White and S. L. Salzberg (1999). Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27(23): 4636-4641.

Delcher, A. L., S. L. Salzberg and A. M. Phillippy (2003). Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics 10-3. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1003s00/full.

Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. Cell 89(7): 1007-1010.

Deppenmeier, U. (2002). The unique biochemistry of methanogenesis. Prog Nucleic Acid Res Mol Biol 73: 223-283.

Deppenmeier, U. (2004). The membrane-bound electron transport system of *Methanosarcina* species. J Bioenerg Biomembr 36(1): 55-64.

Deppenmeier, U., L. Tanja and G. Gottschalk (1999). Novel reactions involved in energy conservation by methanogenic archaea. FEBS Lett 457(3): 291-297.

Dermoumi, H. L. and R. A. Ansorg (2001). Isolation and antimicrobial susceptibility testing of fecal strains of the archaeon *Methanobrevibacter smithii*. Chemotherapy 47(3): 177-183.

Devenish, S. R. A., J. W. Blunt and J. A. Gerrard (2010). NMR studies uncover alternate substrates for dihydrodipicolinate synthase and suggest that dihydrodipicolinate reductase is also a dehydratase. J Med Chem 53(12): 4808-4812.

Diekert, G., U. Konheiser, K. Piechulla and R. K. Thauer (1981). Nickel requirement and factor F_{430} content of methanogenic bacteria. J Bacteriol 148(2): 459-464.

Ding, S. Y., M. T. Rincon, R. Lamed, J. C. Martin, S. I. McCrae, V. Aurilia, Y. Shoham, E. A. Bayer and H. J. Flint (2001). Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. J Bacteriol 183(6): 1945-1953.

Doddema, H. J. and G. D. Vogels (1978). Improved identification of methanogenic bacteria by fluorescence microscopy. Appl Environ Microbiol 36(5): 752-754.

Dodt, M., J. T. Roehr, R. Ahmed and C. Dieterich (2012). FLEXBAR-Flexible barcode and adapter processing for next-generation sequencing platforms. Biology (Basel) 1(3): 895-905.

Dridi, B., M. L. Fardeau, B. Ollivier, D. Raoult and M. Drancourt (2012). *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. Int J Syst Evol Microbiol 62(8): 1902-1907.

Dridi, B., M. Henry, H. Richet, D. Raoult and M. Drancourt (2012). Age-related prevalence of *Methanomassiliicoccus luminyensis* in the human gut microbiome. APMIS 120(10): 773-777.

Dumora, C., A. M. Lacoste and A. Cassaigne (1983). Purification and properties of 2-aminoethylphosphonate:pyruvate aminotransferase from *Pseudomonas aeruginosa*. Eur J Biochem 133(1): 119-125.

Dykhuizen, D. (1978). Selection for tryptophan auxotrophs of *Escherichia coli* in glucose-limited chemostats as a test of the energy conservation hypothesis of evolution. Evolution 32(1): 125-150.

Eckard, R. J., C. Grainger and C. A. M. de Klein (2010). Options for the abatement of methane and nitrous oxide from ruminant production: A review. Livest Sci 130(1-3): 47-56.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson and D. A. Relman (2005). Diversity of the human intestinal microbial flora. Science 308(5728): 1635-1638.

Eddy, S. R. (1998). Profile hidden Markov models. Bioinformatics 14(9): 755-763.

Eddy, S. R. and R. Durbin (1994). RNA sequence analysis using covariance models. Nucleic Acids Res 22(11): 2079-2088.

Egert, M., B. Wagner, T. Lemke, A. Brune and M. W. Friedrich (2003). Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). Appl Environ Microbiol 69(11): 6659-6668.

Ehrlich, M., K. F. Norris, R. Y. Wang, K. C. Kuo and C. W. Gehrke (1986). DNA cytosine methylation and heat-induced deamination. Biosci Rep 6(4): 387-393.

Eichler, J. and R. Moll (2001). The signal recognition particle of Archaea. Trends Microbiol 9(3): 130-136.

Eisen, J. A., J. F. Heidelberg, O. White and S. L. Salzberg (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol 1(6). Retrieved from http://genomebiology.biomedcentral.com/articles/10.1186/gb-2000-1-6-research0011.

Ekiel, I., G. D. Sprott and G. B. Patel (1985). Acetate and CO_2 assimilation by *Methanothrix concilii*. J Bacteriol 162(3): 905-908.

Ellefson, W. L. and R. S. Wolfe (1981). Component C of the methylreductase system of *Methanobacterium*. J Biol Chem 256(9): 4259-4262.

Ellefson, W. L., W. B. Whitman and R. S. Wolfe (1982). Nickel-containing factor F_{430}: chromophore of the methylreductase of *Methanobacterium*. Proc Natl Acad Sci U S A 79(12): 3707-3710.

Ellen, A. F., B. Zolghadr, A. M. Driessen and S. V. Albers (2010). Shaping the archaeal cell envelope. Archaea 2010: 608243. Retrieved from http://downloads.hindawi.com/journals/arch/2010/608243.pdf.

Ellen, A. F., S. V. Albers, W. Huibers, A. Pitcher, C. F. Hobel, H. Schwarz, M. Folea, S. Schouten, E. J. Boekema, B. Poolman and A. J. Driessen (2009). Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. Extremophiles 13(1): 67-79.

Ellen, A. F., S. V. Albers and A. J. Driessen (2010). Comparative study of the extracellular proteome of *Sulfolobus* species reveals limited secretion. Extremophiles 14(1): 87-98.

Ermler, U., W. Grabarse, S. Shima, M. Goubeaud and R. K. Thauer (1997). Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. Science 278(5342): 1457-1462.

Evans, P. N., D. H. Parks, G. L. Chadwick, S. J. Robbins, V. J. Orphan, S. D. Golding and G. W. Tyson (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. Science 350(6259): 434-438.

Ezaki, S., N. Maeda, T. Kishimoto, H. Atomi and T. Imanaka (1999). Presence of a structurally novel type ribulose-bisphosphate carboxylase/oxygenase in the hyperthermophilic archaeon, *Pyrococcus kodakaraensis* KOD1. J Biol Chem 274(8): 5078-5082.

Fabry, V. J., B. A. Seibel, R. A. Feely and J. C. Orr (2008). Impacts of ocean acidification on marine fauna and ecosystem processes. ICES J Marine Sci 65(3): 414-432.

Fay, M. and M. M. Fay (2010). Package 'perm'. Retrieved from http://www.idg.pl/mirrors/CRAN/web/packages/perm/perm.pdf.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the Bootstrap. Evolution 39(4): 783-791.

Fenchel, T. and B. J. Finlay (1991). The biology of free-living anaerobic ciliates. Europ J Protistol 26(3): 201-215.

Ferguson, D. J., N. Gorlatova, D. A. Grahame and J. A. Krzycki (2000). Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. J Biol Chem 275(37): 29053-29060.

Ferguson, D. J. and J. A. Krzycki (1997). Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from *Methanosarcina barkeri*. J Bacteriol 179(3): 846-852.

Ferrari, A., T. Brusa, A. Rutili, E. Canzi and B. Biavati (1994). Isolation and characterization of *Methanobrevibacter oralis* sp. nov. Curr Microbiol 29(1): 7-12.

Fiebig, K. and B. Friedrich (1989). Purification of the F_{420}-reducing hydrogenase from *Methanosarcina barkeri* (strain Fusaro). Eur J Biochem 184(1): 79-88.

Fiebig, K. and G. Gottschalk (1983). Methanogenesis from choline by a coculture of *Desulfovibrio* sp. and *Methanosarcina barkeri*. Appl Environ Microbiol 45(1): 161-168.

Fimereli, D. K., K. D. Tsirigos, Z. I. Litou, T. D. Liakopoulos, P. G. Bagos and S. J. Hamodrakas (2012). CW-PRED: a HMM-based method for the classification of cell wall-anchored proteins of Gram-positive bacteria. Artificial Intelligence: Theories and Applications, Berlin, Heidelberg, Springer: 285-290.

Finlay, B. J., G. Esteban, K. J. Clarke, A. G. Williams, T. M. Embley and R. P. Hirt (1994). Some rumen ciliates have endosymbiotic methanogens. FEMS Microbiol Lett 117(2): 157-161.

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate and A. Bateman (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44(D1): D279-285. Retrieved from http://nar.oxfordjournals.org/content/44/D1/D279.long.

Fischer, M., W. Römisch, S. Schiffmann, M. Kelly, H. Oschkinat, S. Steinbacher, R. Huber, W. Eisenreich, G. Richter and A. Bacher (2002). Biosynthesis of riboflavin in archaea studies on the mechanism of 3, 4-dihydroxy-2-butanone-4-phosphate synthase of *Methanococcus jannaschii*. J Biol Chem 277(44): 41410-41416.

Fischer, M., A. K. Schott, W. Romisch, A. Ramsperger, M. Augustin, A. Fidler, A. Bacher, G. Richter, R. Huber and W. Eisenreich (2004). Evolution of vitamin B_2 biosynthesis. A novel class of riboflavin synthase in Archaea. J Mol Biol 343(1): 267-278.

Fischer, R. and R. Jensen (1987). Prephenate dehydratase (monofunctional). Methods Enzymol 142: 507-512.

Fischer, R. J., J. Helms and P. Durre (1993). Cloning, sequencing, and molecular analysis of the sol operon of *Clostridium acetobutylicum*, a chromosomal locus involved in solventogenesis. J Bacteriol 175(21): 6959-6969.

Flannigan, K. A., S. H. Hennigan, H. H. Vogelbacker, J. S. Gots and J. M. Smith (1990). Purine biosynthesis in *Escherichia coli* K12: structure and DNA sequence studies of the *purHD* locus. Mol Microbiol 4(3): 381-392.

Flint, D. H., J. F. Tuminello and T. J. Miller (1996). Studies on the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase in *Escherichia coli* crude extract. Isolation of *O*-acetylserine sulfhydrylases A and B and β-cystathionase based on their ability to mobilize sulfur from cysteine and to participate in Fe-S cluster synthesis. J Biol Chem 271(27): 16053-16067.

Foley, P. A., D. A. Kenny, J. J. Callan, T. M. Boland and F. P. O'Mara (2009). Effect of DL-malic acid supplementation on feed intake, methane emission, and rumen fermentation in beef cattle. J Anim Sci 87(3): 1048-1057.

Fournier, G. P., A. A. Dick, D. Williams and J. P. Gogarten (2011). Evolution of the Archaea: emerging views on origins and phylogeny. Res Microbiol 162(1): 92-98.

Fournier, G. P. and J. P. Gogarten (2008). Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*. J Bacteriol 190(3): 1124-1127.

Franzosa, E. A., X. C. Morgan, N. Segata, L. Waldron, J. Reyes, A. M. Earl, G. Giannoukos, M. R. Boylan, D. Ciulla, D. Gevers, J. Izard, W. S. Garrett, A. T. Chan and C. Huttenhower (2014). Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A 111(22): E2329-E2338.

Fraser, H. I., M. Kvaratskhelia and M. F. White (1999). The two analogous phosphoglycerate mutases of *Escherichia coli*. FEBS Lett 455(3): 344-348.

Fricke, W. F., H. Seedorf, A. Henne, M. Kruer, H. Liesegang, R. Hedderich, G. Gottschalk and R. K. Thauer (2006). The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H_2 for methane formation and ATP synthesis. J Bacteriol 188(2): 642-658.

Friedrich, M. W. (2005). Methyl‐coenzyme M reductase genes: unique functional markers for methanogenic and anaerobic methane‐oxidizing Archaea. Methods Enzymol. R. L. Jared. Academic Press. 397: 428-442.

Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005). Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3(9): 722-732.

Fukuda, W., N. Morimoto, T. Imanaka and S. Fujiwara (2008). Agmatine is essential for the cell growth of *Thermococcus kodakaraensis*. FEMS Microbiol Lett 287(1): 113-120.

Fuqua, C. (2010). Passing the baton between laps: adhesion and cohesion in *Pseudomonas putida* biofilms. Mol Microbiol 77(3): 533-536.

Furukawa, N., A. Miyanaga, M. Togawa, M. Nakajima and H. Taguchi (2014). Diverse allosteric and catalytic functions of tetrameric D-lactate dehydrogenases from three Gram-negative bacteria. AMB Express 4(1): 76.

Gao, B. and R. S. Gupta (2007). Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. BMC Genomics 8(1): 86. Retrieved from http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-8-86.

Garrity, G. M., T. G. Lilburn, J. R. Cole, S. H. Harrison, J. Euzeby and B. J. Tindall (2007). Part 1-The Archaea: Phyla "Crenarchaeota" and "Euryarchaeota". TOBA 7(7): 6-31.

Gaston, M. A., R. Jiang and J. A. Krzycki (2011a). Functional context, biosynthesis, and genetic encoding of pyrrolysine. Curr Opin Microbiol 14(3): 342-349.

Gaston, M. A., L. Zhang, K. i. B. Green-Church and J. A. Krzycki (2011b). The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature 471(7340): 647-650.

Geerkens, C. H., R. M. Schweiggert, H. Steingass, J. Boguhn, M. Rodehutscord and R. Carle (2013). Influence of apple and citrus pectins, processed mango peels, a phenolic mango peel extract, and gallic acid as potential feed supplements on *in vitro* total gas production and rumen methanogenesis. J Agric Food Chem 61(24): 5727-5737.

Gerdes, K., S. K. Christensen and A. Lobner-Olesen (2005). Prokaryotic toxin-antitoxin stress response loci. Nat Rev Microbiol 3(5): 371-382.

Ghali, M. B., P. T. Scott and R. A. Al Jassim (2004). Characterization of *Streptococcus bovis* from the rumen of the dromedary camel and Rusa deer. Lett Appl Microbiol 39(4): 341-346.

Gilbert, C. and R. Cordaux (2013). Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. Genome Biol Evol 5(5): 822-832.

Gish, W. and D. J. States (1993). Identification of protein coding regions by database similarity search. Nat Genet 3(3): 266-272.

Goel, G. and H. P. Makkar (2012). Methane mitigation from ruminants using tannins and saponins. Trop Anim Health Prod 44(4): 729-739.

Goldberg, T., M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansorge, K. Balasz, M. Bernhofer, A. Betz, L. Cizmadija, K. T. Do, J. Gerke, R. Greil, V. Joerdens, M. Hastreiter, K. Hembach, M. Herzog, M. Kalemanov, M. Kluge, A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldraff, J. Zierer, H. Nielsen and B. Rost (2014). LocTree3 prediction of localization. Nucleic Acids Res 42(Web Server issue): W350-W355. Retrieved from http://nar.oxfordjournals.org/content/42/W1/W350.full.

Gomez-Alarcon, R. A., C. O'Dowd, J. A. Leedle and M. P. Bryant (1982). 1,4-Naphthoquinone and other nutrient requirements of *Succinivibrio dextrinosolvens*. Appl Environ Microbiol 44(2): 346-350.

Gorlas, A., C. Robert, G. Gimenez, M. Drancourt and D. Raoult (2012). Complete genome sequence of *Methanomassiliicoccus luminyensis*, the largest genome of a human-associated Archaea species. J Bacteriol 194(17): 4745-4745.

Gottschalk, G. and R. K. Thauer (2001). The Na^+-translocating methyltransferase complex from methanogenic archaea. Biochim Biophys Acta 1505(1): 28-36.

Gower, J. C. and P. Legendre (1986). Metric and Euclidean properties of dissimilarity coefficients. J Classification 3(1): 5-48.

Grabarse, W., F. Mahlert, S. Shima, R. K. Thauer and U. Ermler (2000). Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: unusual amino acid modification, conservation and adaptation. J Mol Biol 303(2): 329-344.

Graber, J. R. and J. A. Breznak (2005). Folate cross-feeding supports symbiotic homoacetogenic spirochetes. Appl Environ Microbiol 71(4): 1883-1889.

Grada, A. and K. Weinbrecht (2013). Next-generation sequencing: methodology and application. J Invest Dermatol 133(8): 1-4.

Graham, D. E., S. M. Taylor, R. Z. Wolf and S. C. Namboori (2009). Convergent evolution of coenzyme M biosynthesis in the *Methanosarcinales*: cysteate synthase evolved from an ancestral threonine synthase. Biochem J 424(3): 467-478.

Green, S. M., T. Malik, I. G. Giles and W. T. Drabble (1996). The *purB* gene of *Escherichia coli* K-12 is located in an operon. Microbiology 142(11): 3219-3230.

Griffin, M., R. Casadio and C. M. Bergamini (2002). Transglutaminases: nature's biological glues. Biochem J 368(2): 377-396.

Griffiths-Jones, S., S. Moxon, M. Marshall, A. Khanna, S. R. Eddy and A. Bateman (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res 33(Database issue): D121-D124. Retrieved from http://nar.oxfordjournals.org/content/33/suppl_1/D121.long.

Grissa, I., G. Vergnaud and C. Pourcel (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35(Web Server issue): W52-W57. Retrieved from http://nar.oxfordjournals.org/content/35/suppl_2/W52.full.

Grochowski, L. L., H. Xu and R. H. White (2006). *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. J Bacteriol 188(9): 3192-3198.

Großkopf, R., S. Stubner and W. Liesack (1998). Novel euryarchaeotal lineages detected on rice roots and in the anoxic bulk soil of flooded rice microcosms. Appl Environ Microbiol 64(12): 4983-4989.

Gruer, M. J. and J. R. Guest (1994). Two genetically-distinct and differentially-regulated aconitases (AcnA and AcnB) in *Escherichia coli*. Microbiology 140(10): 2531-2541.

Guan, H., K. M. Wittenberg, K. H. Ominski and D. O. Krause (2006). Efficacy of ionophores in cattle diets for mitigation of enteric methane. J Anim Sci 84(7): 1896-1906.

Gunsalus, R. P. and R. S. Wolfe (1980). Methyl coenzyme M reductase from *Methanobacterium thermoautotrophicum*. Resolution and properties of the components. J Biol Chem 255(5): 1891-1895.

Guo, Y. Q., J. X. Liu, Y. Lu, W. Y. Zhu, S. E. Denman and C. S. McSweeney (2008). Effect of tea saponin on methanogenesis, microbial community structure and expression of *mcrA* gene, in cultures of rumen micro-organisms. Lett Appl Microbiol 47(5): 421-426.

Gupta, A. and U. B. Chaudhary (2010). Isolation and characterization of *Methanobrevibacter smithii* GMS-01 from rumen of goats. Indian Vet J 87(10): 1009-1012.

Guss, A. M., G. Kulkarni and W. W. Metcalf (2009). Differences in hydrogenase gene expression between *Methanosarcina acetivorans* and *Methanosarcina barkeri*. J Bacteriol 191(8): 2826-2833.

Haase, I., S. Mortl, P. Kohler, A. Bacher and M. Fischer (2003). Biosynthesis of riboflavin in archaea. 6,7-dimethyl-8-ribityllumazine synthase of *Methanococcus jannaschii*. Eur J Biochem 270(5): 1025-1032.

Haft, D. H., J. D. Selengut and O. White (2003). The TIGRFAMs database of protein families. Nucleic Acids Res 31(1): 371-373.

Haines, T. H. and N. A. Dencher (2002). Cardiolipin: a proton trap for oxidative phosphorylation. FEBS Lett 528(1-3): 35-39.

Hamby, S. E., N. S. Thomas, D. N. Cooper and N. Chuzhanova (2011). A meta-analysis of single base-pair substitutions in translational termination codons ('nonstop' mutations) that cause human inherited disease. Hum Genomics 5(4): 241-264.

Hansen, T., D. Wendorff and P. Schönheit (2004). Bifunctional phosphoglucose/phosphomannose isomerases from the Archaea *Aeropyrum pernix* and *Thermoplasma acidophilum* constitute a novel enzyme family within the phosphoglucose isomerase superfamily. J Biol Chem 279(3): 2262-2272.

Hao, B., W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki and M. K. Chan (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 296(5572): 1462-1466.

Hartmann, D. L., A. M. G. Klein Tank, M. Rusticucci, L. V. Alexander, S. Brönnimann, Y. Charabi, F. J. Dentener, E. J. Dlugokencky, D. R. Easterling, A. Kaplan, B. J. Soden, P. W. Thorne, M. Wild and P. M. Zhai (2013). Observations: Atmosphere and Surface. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley. Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press: 159–254.

Harms, U., D. S. Weiss, P. Gartner, D. Linder and R. K. Thauer (1995). The energy conserving N^5-methyltetrahydromethanopterin:coenzyme M methyltransferase complex from *Methanobacterium thermoautotrophicum* is composed of eight different subunits. Eur J Biochem 228(3): 640-648.

Haruyama, K., T. Nakai, I. Miyahara, K. Hirotsu, H. Mizuguchi, H. Hayashi and H. Kagamiyama (2001). Structures of *Escherichia coli* histidinol-phosphate aminotransferase and its complexes with histidinol-phosphate and *N*-(5'-phosphopyridoxyl)-L-glutamate: double substrate recognition of the enzyme. Biochem 40(15): 4633-4644.

Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science 301(5639): 1496-1499.

Hedderich, R. (2004). Energy-converting [NiFe] hydrogenases from archaea and extremophiles: ancestors of complex I. J Bioenerg Biomembr 36(1): 65-75.

Hedderich, R., A. Berkessel and R. K. Thauer (1990). Purification and properties of heterodisulfide reductase from *Methanobacterium thermoautotrophicum* (strain Marburg). Eur J Biochem 193(1): 255-261.

Hedderich, R., N. Hamann and M. Bennati (2005). Heterodisulfide reductase from methanogenic archaea: a new catalytic role for an iron-sulfur cluster. Biol Chem 386(10): 961-970.

Heiden, S., R. Hedderich, E. Setzke and R. K. Thauer (1994). Purification of a two-subunit cytochrome-*b*-containing heterodisulfide reductase from methanol-grown *Methanosarcina barkeri*. Eur J Biochem 221(2): 855-861.

Helgadóttir, S., G. Rosas-Sandoval, D. Söll and D. E. Graham (2007). Biosynthesis of phosphoserine in the *Methanococcales*. J Bacteriol 189(2): 575-582.

Helling, R. B. (1998). Pathway choice in glutamate synthesis in *Escherichia coli*. J Bacteriol 180(17): 4571-4575.

Hemmi, H., K. Shibuya, Y. Takahashi, T. Nakayama and T. Nishino (2004). (*S*)-2, 3-Di-*O*-geranylgeranylglyceryl phosphate synthase from the thermoacidophilic archaeon *Sulfolobus solfataricus* molecular cloning and characterization of a membrane-intrinsic prenyl transferase involved in the biosynthesis of archaeal ether-linked membrane lipids. J Biol Chem 279(48): 50197-50203.

Henikoff, S. and J. G. Henikoff (1992). Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89(22): 10915-10919.

Henderson, G., F. Cox, S. Ganesh, A. Jonker, W. Young and P. H. Janssen (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Sci Rep 5: 14567. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4598811/.

Hendrickson, E. L. and J. A. Leigh (2008). Roles of coenzyme F_{420}-reducing hydrogenases and hydrogen- and F_{420}-dependent methylenetetrahydromethanopterin dehydrogenases in

reduction of F_{420} and production of hydrogen during methanogenesis. J Bacteriol 190(14): 4818-4821.

Herring, S., A. Ambrogelly, C. R. Polycarpo and D. Soll (2007). Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. Nucleic Acids Res 35(4): 1270-1278.

Hertz, A. F. and A. Newton (1913). The normal movements of the colon in man. J Physiol 47(1-2): 57-65.

Hino, T., K. Takeshi, M. Kanda and S. Kumazawa (1993). Effects of aibellin, a novel peptide antibiotic, on rumen fermentation *in vitro*. J Dairy Sci 76(8): 2213-2221.

Hippe, H., D. Caspari, K. Fiebig and G. Gottschalk (1979). Utilization of trimethylamine and other *N*-methyl compounds for growth and methane formation by *Methanosarcina barkeri*. Proc Natl Acad Sci U S A 76(1): 494-498.

Hirabayashi, T., T. J. Larson and W. Dowhan (1976). Membrane-associated phosphatidylglycerophosphate synthetase from *Escherichia coli*: purification by substrate affinity chromatography on cytidine 5'-diphospho-1,2-diacyl-*sn*-glycerol sepharose. Biochem 15(24): 5205-5211.

Ho, C. L. and K. Saito (2001). Molecular biology of the plastidic phosphorylated serine biosynthetic pathway in *Arabidopsis thaliana*. Amino Acids 20(3): 243-259.

Hobson, P. N. and C. S. Stewart (1997). The rumen microbial ecosystem, *Second Edition*, London, Blackie Academic & Professional.

Hoch, J. A. L., R. Sonenshein and L. Abraham (1993). *Bacillus subtilis* and other Gram-positive bacteria: biochemistry, physiology, and molecular genetics. Washington, D.C, ASM.

Hoedt, E. C., P. Ó. Cuív, P. N. Evans, W. J. Smith, C. S. McSweeney, S. E. Denman, & M. Morrison (2016). Differences down-under: alcohol-fueled methanogenesis by archaea present in Australian macropodids. ISME J 10(10): 2376-2388.

Hollander, M., D. A. Wolfe and E. Chicken (2013). Nonparametric statistical methods, *Third Edition*, Hoboken, New Jersey, John Wiley & Sons.

Honzatko, R. B. and H. J. Fromm (1999). Structure-function studies of adenylosuccinate synthetase from *Escherichia coli*. Arch Biochem Biophys 370(1): 1-8.

Hor, L., R. C. J. Dobson, C. Dogovski, C. A. Hutton and M. A. Perugini (2010). Crystallization and preliminary X-ray diffraction analysis of diaminopimelate epimerase from *Escherichia coli*. Struct Biol and Crystal Commun 66(1): 37-40.

Horz, H. P., I. Seyfarth and G. Conrads (2012). McrA and 16*S* rRNA gene analysis suggests a novel lineage of *Archaea* phylogenetically affiliated with *Thermoplasmatales* in human subgingival plaque. Anaerobe 18(3): 373-377.

Hoskins, A. A., R. Anand, S. E. Ealick and J. Stubbe (2004). The formylglycinamide ribonucleotide amidotransferase complex from *Bacillus subtilis*: metabolite-mediated complex formation. Biochemistry 43(32): 10314-10327.

Hottes, A. K., P. L. Freddolino, A. Khare, Z. N. Donnell, J. C. Liu and S. Tavazoie (2013). Bacterial adaptation through loss of function. PLoS Genet 9(7): e1003617. Retrieved from http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003617.

Hove-Jensen, B., K. W. Harlow, C. J. King and R. L. Switzer (1986). Phosphoribosylpyrophosphate synthetase of *Escherichia coli*. Properties of the purified enzyme and primary structure of the *prs* gene. J Biol Chem 261(15): 6765-6771.

Højsgaard, S., M. S. Højsgaard and D. Hmisc (2006). The doBy package. The Newsletter of the R Project 6: 1. Retrieved from http://people.math.aau.dk/~sorenh/software/doBy/doBy/doBy.pdf.

Hsu, S. T., E. Breukink, E. Tischenko, M. A. Lutters, B. de Kruijff, R. Kaptein, A. M. Bonvin and N. A. van Nuland (2004). The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. Nat Struct Mol Biol 11(10): 963-967.

Huang, L. N., H. Zhou, Y. Q. Chen, S. Luo, C. Y. Lan and L. H. Qu (2002). Diversity and structure of the archaeal community in the leachate of a full-scale recirculating landfill as examined by direct 16*S* rRNA gene sequence retrieval. FEMS Microbiol Lett 214(2): 235-240.

Huang, L. N., Y. Q. Chen, H. Zhou, S. Luo, C. Y. Lan and L. H. Qu (2003). Characterization of methanogenic Archaea in the leachate of a closed municipal solid waste landfill. FEMS Microbiol Ecol 46(2): 171-177.

Huang, X. D., H. Y. Tan, R. Long, J. B. Liang and A. D. Wright (2012). Comparison of methanogen diversity of yak (*Bos grunniens*) and cattle (*Bos taurus*) from the Qinghai-Tibetan plateau, China. BMC Microbiol 12(1): 237. Retrieved from https://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-12-237.

Huisingh, J., J. J. McNeill and G. Matrone (1974). Sulfate reduction by a *Desulfovibrio* species isolated from sheep rumen. Appl Microbiol 28(3): 489-497.

Hungate, R. E. (1966). The rumen and its microbes. Michigan, Academic Press.

Hungate, R. E. (1967). Hydrogen as an intermediate in the rumen fermentation. Arch Mikrobiol 59(1-3): 158-164.

Huntington, G. B. (1997). Starch utilization by ruminants: from basics to the bunk. J Anim Sci 75(3): 852-867.

Huss, R. J. and M. Glaser (1983). Identification and purification of an adenylate kinase-associated protein that influences the thermolability of adenylate kinase from a temperature-sensitive *adk* mutant of *Escherichia coli*. J Biol Chem 258(21): 13370-13376.

Iannotti, E. L., D. Kafkewitz, M. J. Wolin and M. P. Bryant (1973). Glucose fermentation products in *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes*: changes caused by interspecies transfer of H_2. J Bacteriol 114(3): 1231-1240.

Iino, T., H. Tamaki, S. Tamazawa, Y. Ueno, M. Ohkuma, K. Suzuki, Y. Igarashi and S. Haruta (2013). *Candidatus* Methanogranum caenicola: a novel methanogen from the anaerobic digested sludge, and proposal of *Methanomassiliicoccaceae* fam. nov. and *Methanomassiliicoccales* ord. nov., for a methanogenic lineage of the class *Thermoplasmata*. Microbes Environ 28(2): 244-250.

Ikemura, T. and M. Nomura (1977). Expression of spacer tRNA genes in ribosomal RNA transcription units carried by hybrid Col E1 plasmids in *E. coli*. Cell 11(4): 779-793.

Ilag, L. L., D. Jahn, G. Eggertsson and D. Söll (1991). The *Escherichia coli hemL* gene encodes glutamate 1-semialdehyde aminotransferase. J Bacteriol 173(11): 3408-3413.

Imachi, H., S. Sakai, H. Nagai, T. Yamaguchi and K. Takai (2009). *Methanofollis ethanolicus* sp. nov., an ethanol-utilizing methanogen isolated from a lotus field. Int J Syst Evol Microbiol 59(4): 800-805.

Immig, I., D. Demeyer, D. Fiedler, C. Van Nevel and L. Mbanzamihigo (1996). Attempts to induce reductive acetogenesis into a sheep rumen. Arch Tierernahr 49(4): 363-370.

Irbis, C. and K. Ushida (2004). Detection of methanogens and proteobacteria from a single cell of rumen ciliate protozoa. J Gen Appl Microbiol 50(4): 203-212.

Iverson, V., R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales and E. V. Armbrust (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science 335(6068): 587-590.

Jansen, K., E. Stupperich and G. Fuchs (1982). Carbohydrate synthesis from acetyl CoA in the autotroph *Methanobacterium thermoautotrophicum*. Arch Microbiol 132(4): 355-364.

Janssen, P. H. and M. Kirs (2008). Structure of the archaeal community of the rumen. Appl Environ Microbiol 74(12): 3619-3625.

Janssen, P. H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. Anim Feed Sci Technol 160(1): 1-22.

Janssen, P. H. and M. Kirs (2008). Structure of the archaeal community of the rumen. Appl Environ Microbiol 74(12): 3619-3625.

Jarrell, K. F., G. M. Jones, L. Kandiba, D. B. Nair and J. Eichler (2010). S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications. Archaea 2010: 918-922. Retrieved from http://downloads.hindawi.com/journals/arch/2010/612948.pdf.

Jarvis, G. N., C. Strompl, D. M. Burgess, L. C. Skillman, E. R. Moore and K. N. Joblin (2000). Isolation and identification of ruminal methanogens from grazing cattle. Curr Microbiol 40(5): 327-332.

Jayanegara, A., E. Wina and J. Takahashi (2014). Meta-analysis on methane mitigating properties of saponin-rich sources in the rumen: Influence of addition levels and plant sources. Asian-Australas J Anim Sci 27(10): 1426-1435.

Jenkins, J., V. E. Shevchik, N. Hugouvieux-Cotte-Pattat and R. W. Pickersgill (2004). The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*. J Biol Chem 279(10): 9139-9145.

Jeyanathan, J. (2010). Investigation of rumen methanogens in New Zealand livestock Philosophy in Animal Science Thesis, Massey University.

Jeyanathan, J., M. Kirs, R. S. Ronimus, S. O. Hoskin and P. H. Janssen (2011). Methanogen community structure in the rumens of farmed sheep, cattle and red deer fed different diets. FEMS Microbiol Ecol 76(2): 311-326.

Jin, W., Y. F. Cheng, S. Y. Mao and W. Y. Zhu (2014). Discovery of a novel rumen methanogen in the anaerobic fungal culture and its distribution in the rumen as revealed by real-time PCR. BMC Microbiol 14(1): 104. Retrieved from http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-14-104.

Jindou, S., J. M. Brulc, M. Levy-Assaraf, M. T. Rincon, H. J. Flint, M. E. Berg, M. K. Wilson, B. A. White, E. A. Bayer, R. Lamed and I. Borovok (2008). Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*. FEMS Microbiol Lett 285(2): 188-194.

Joblin, K. N. (1999). Ruminal acetogens and their potential to lower ruminant methane emissions. Aust J Agric Res 50(8): 1307-1314.

Jofre, E., A. Lagares and G. Mori (2004). Disruption of dTDP-rhamnose biosynthesis modifies lipopolysaccharide core, exopolysaccharide production, and root colonization in *Azospirillum brasilense*. FEMS Microbiol Lett 231(2): 267-275.

Johansson, L., G. Gafvelin and E. S. Arner (2005). Selenocysteine in proteins-properties and biotechnological use. Biochim Biophys Acta 1726(1): 1-13.

Johnson, K. A. and D. E. Johnson (1995). Methane emissions from cattle. J Anim Sci 73(8): 2483-2492.

Johnson, P. F. and J. Abelson (1983). The yeast $tRNA^{Tyr}$ gene intron is essential for correct modification of its tRNA product. Nature 302: 681-687.

Jonah, M. and J. A. Erwin (1971). The lipids of membraneous cell organelles isolated from the ciliate, *Tetrahymena pyriformis*. Biochim Biophys Acta 231(1): 80-92.

Jones, D. T., W. R. Taylor and J. M. Thornton (1992). The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8(3): 275-282.

Jongsareejit, B., R. N. Rahman, S. Fujiwara and T. Imanaka (1997). Gene cloning, sequencing and enzymatic properties of glutamate synthase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1. Mol Gen Genet 254(6): 635-642.

Jordan, P. M., S. D. Thomas and M. J. Warren (1988). Purification, crystallization and properties of porphobilinogen deaminase from a recombinant strain of *Escherichia coli* K12. Biochem J 254(2): 427-435.

Kaback, H. R., M. Sahin-Toth and A. B. Weinglass (2001). The kamikaze approach to membrane transport. Nat Rev Mol Cell Biol 2(8): 610-620.

Kaesler, B. and P. Schonheit (1989). The role of sodium ions in methanogenesis. Formaldehyde oxidation to CO_2 and $2H_2$ in methanogenic bacteria is coupled with primary electrogenic Na^+ translocation at a stoichiometry of 2-3 Na^+/CO_2. Eur J Biochem 184(1): 223-232.

Kameda, Y., S. Oira, K. Matsui, S. Kanatomo and T. Hase (1974). Antitumor activity of bacillus natto. V. Isolation and characterization of surfactin in the culture medium of *Bacillus natto* KMD 2311. Chem Pharm Bull (Tokyo) 22(4): 938-944.

Kaminski, L. and J. Eichler (2014). *Haloferax volcanii N*-glycosylation: delineating the pathway of dTDP-rhamnose biosynthesis. PloS one 9(5): e97441. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0097441.

Kandler, O. and H. Konig (1978). Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria. Arch Microbiol 118(2): 141-152.

Kandler, O. and H. Konig (1998). Cell wall polymers in Archaea (Archaebacteria). Cell Mol Life Sci 54(4): 305-308.

Kanehisa, M. and S. Goto (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28(1): 27-30.

Karl, T. R. and K. E. Trenberth (2003). Modern global climate change. Science 302(5651): 1719-1723.

Kaster, A. K., J. Moll, K. Parey and R. K. Thauer (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proc Natl Acad Sci U S A 108(7): 2981-2986.

Kato, N., H. Yurimoto and R. K. Thauer (2006). The physiological role of the ribulose monophosphate pathway in bacteria and archaea. Biosci Biotechnol Biochem 70(1): 10-21.

Kawamukai, M., R. Utsumi, K. Takeda, A. Higashi, H. Matsuda, Y. L. Choi and T. Komano (1991). Nucleotide sequence and characterization of the *sfs*1 gene: *sfs1* is involved in CRP*-dependent mal gene expression in *Escherichia coli*. J Bacteriol 173(8): 2644-2648.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647-1649.

Keller, M., G. Burkard, H. J. Bohnert, M. Mubumbila, K. Gordon, A. Steinmetz, D. Heiser, E. J. Crouse and J. H. Weil (1980). Transfer RNA genes associated with the 16*S* and 23*S* rRNA genes of *Euglena* chloroplast DNA. Biochem Biophys Res Commun 95(1): 47-54.

Kelly, W. J., S. C. Leahy, D. Li, R. Perry, S. C. Lambie, G. T. Attwood and E. Altermann (2014). The complete genome sequence of the rumen methanogen *Methanobacterium formicicum* BRM9. Stand Genomic Sci 9(1): 15. Retrieved from https://standardsingenomics.biomedcentral.com/articles/10.1186/1944-3277-9-15.

Kelly, W. J., D. M. Pacheco, D. Li, G. T. Attwood, E. Altermann and S. C. Leahy (2016). The complete genome sequence of the rumen methanogen *Methanobrevibacter millerae* SM9. Stand Genomic Sci. Unpublished manuscript.

Kelly, W. J., D. Li, S. C. Lambie, F. Cox, G. T. Attwood, E. Altermann and S. C. Leahy (2016). Draft genome sequence of the rumen methanogen *Methanobrevibacter olleyae* YLM1. Genome Announc 4(2): e00232-16. Retrieved from http://genomea.asm.org/content/4/2/e00232-16.full.

Kelly, W. J., D. Li, S. C. Lambie, J. Jeyanathan, F. Cox, Y. Li, G. T. Attwood, E. Altermann and S. C. Leahy (2016). Complete genome sequence of methanogenic archaeon ISO4-G1, a member of the Methanomassiliicoccales, isolated from a sheep rumen. Genome Announc 4(2): e00221-16. Retrieved from http://genomea.asm.org/content/4/2/e00221-16.full.

Keltjens, J. T. and G. D. Vogels (1993). Conversion of methanol and methylamines to methane and carbon dioxide. Methanogenesis. J. G. Ferry. Boca Raton, Florida, CRC Press: 253-303.

Kenklies, J., R. Ziehn, K. Fritsche, A. Pich and J. R. Andreesen (1999). Proline biosynthesis from L-ornithine in *Clostridium sticklandii*: purification of Δ1-pyrroline-5-carboxylate reductase, and sequence and expression of the encoding gene, *proC*. Microbiology 145(4): 819-826.

Khakh, B. S. and G. Burnstock (2009). The double life of ATP. Sci Am 301(6): 84-92.

Khan, H., S. H. Flint and P. L. Yu (2013). Determination of the mode of action of enterolysin A, produced by *Enterococcus faecalis* B9510. J Appl Microbiol 115(2): 484-494.

Khelaifia, S., M. Garibal, C. Robert, D. Raoult and M. Drancourt (2014). Draft genome sequence of a human-associated isolate of *Methanobrevibacter arboriphilicus*, the lowest-G+C-content archaeon. Genome Announc 2(1): e01181-13. Retrieved from http://genomea.asm.org/content/2/1/e01181-13.full.

Khelaifia, S., M. Garibal, C. Robert, D. Raoult and M. Drancourt (2014). Draft genome sequencing of *Methanobrevibacter oralis* strain JMR01, isolated from the human intestinal microbiota. Genome Announc 2(1): e00073-14. Retrieved from http://genomea.asm.org/content/2/1/e00073-14.full.

Khomyakova, M., Ö. Bükmez, L. K. Thomas, T. J. Erb and I. A. Berg (2011). A methylaspartate cycle in haloarchaea. Science 331(6015): 334-337.

Kiehl, J. T. and K. E. Trenberth (1997). Earth's annual global mean energy budget. Bull Am Meteorol Soc 78(2): 197-208.

Kim, C. (2012). Identification of rumen methanogens, characterization of substrate requirements and measurement of hydrogen thresholds. Master in Microbiology, Massey University.

Kim, S. Y., K. S. Ju, W. W. Metcalf, B. S. Evans, T. Kuzuyama and W. A. van der Donk (2012). Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces* species. Antimicrob Agents Chemother 56(8): 4175-4183.

Kimura, A., K. T. Mountzouros, D. A. Relman, S. Falkow and J. L. Cowell (1990). *Bordetella pertussis* filamentous hemagglutinin: evaluation as a protective antigen and colonization factor in a mouse respiratory infection model. Infect Immun 58(1): 7-16.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2): 111-120.

Kirby, J., J. C. Martin, A. S. Daniel and H. J. Flint (1997). Dockerin-like sequences in cellulases and xylanases from the rumen cellulolytic bacterium *Ruminococcus flavefaciens*. FEMS Microbiol Lett 149(2): 213-219.

Kitts, C. L., D. P. Cunningham and P. J. Unkefer (1994). Isolation of three hexahydro-1,3,5-trinitro-1,3,5-triazine-degrading species of the family Enterobacteriaceae from nitramine explosive-contaminated soil. Appl Environ Microbiol 60(12): 4608-4611.

Kjems, J. and R. A. Garrett (1987). Novel expression of the ribosomal RNA genes in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. EMBO J 6(11): 3521.

Kloda, A. and B. Martinac (2001). Mechanosensitive channel of *Thermoplasma*, the cell wall-less archaea: cloning and molecular characterization. Cell Biochem Biophys 34(3): 321-347.

Kneidinger, B., M. Graninger, M. Puchberger, P. Kosma and P. Messner (2001). Biosynthesis of nucleotide-activated D-glycero-D-manno-heptose. J Biol Chem 276(24): 20935-20944.

Kolberg, M., K. R. Strand, P. Graff and K. K. Andersson (2004). Structure, function, and mechanism of ribonucleotide reductases. BBA-Protein Proteomics 1699(1–2): 1-34.

Konig, H. and O. Kandler (1979). The amino acid sequence of the peptide moiety of the pseudomurein from *Methanobacterium thermoautotrophicum*. Arch Microbiol 121(3): 271-275.

Koonin, E. V., K. S. Makarova and L. Aravind (2001). Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55: 709-742.

Koval, S. F. and K. F. Jarrell (1987). Ultrastructure and biochemistry of the cell wall of *Methanococcus voltae*. J Bacteriol 169(3): 1298-1306.

Krogh, A., B. Larsson, G. von Heijne and E. L. Sonnhammer (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305(3): 567-580.

Kruskal, W. H. and W. A. Wallis (1952). Use of ranks in one-criterion variance analysis. J Amer Statist. Assoc 47(260): 583-621.

Krzycki, J. A. (2004). Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. Curr Opin Chem Biol 8(5): 484-491.

Kubo, I., K. Nihei and K. Tsujimoto (2003). Antibacterial action of anacardic acids against methicillin resistant *Staphylococcus aureus* (MRSA). J Agric Food Chem 51(26): 7624-7628.

Kühn, W., K. Fiebig, H. Hippe, R. A. Mah, B. A. Huser and G. Gottschalk (1983). Distribution of cytochromes in methanogenic bacteria. FEMS Microbiol Lett 20(3): 407-410.

Kulkarni, G., D. M. Kridelbaugh, A. M. Guss and W. W. Metcalf (2009). Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of *Methanosarcina barkeri*. Proc Natl Acad Sci U S A 106(37): 15915-15920.

Kullen, M. J. and T. R. Klaenhammer (1999). Identification of the pH-inducible, proton-translocating F_1F_0-ATPase (*atpBEFHAGDC*) operon of *Lactobacillus acidophilus* by differential display: gene structure, cloning and characterization. Mol Microbiol 33(6): 1152-1161.

Kunkel, A., J. A. Vorholt, R. K. Thauer and R. Hedderich (1998). An *Escherichia coli* hydrogenase-3-type hydrogenase in methanogenic archaea. Eur J Biochem 252(3): 467-476.

Kuzuyama, T. (2002). Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. Biosci Biotechnol Biochem 66(8): 1619-1627.

L'Haridon, S., M. Chalopin, D. Colombo and L. Toffin (2014). *Methanococcoides vulcani* sp. nov., a marine methylotrophic methanogen that uses betaine, choline and *N*,*N*-dimethylethanolamine for methanogenesis, isolated from a mud volcano, and emended description of the genus *Methanococcoides*. Int J Syst Evol Microbiol 64(6): 1978-1983.

Lambie, S. C., W. J. Kelly, S. C. Leahy, D. Li, K. Reilly, T. A. McAllister, E. R. Valle, G. T. Attwood and E. Altermann (2015). The complete genome sequence of the rumen methanogen *Methanosarcina barkeri* CM1. Stand Genomic Sci 10(1): 57. Retrieved from https://standardsingenomics.biomedcentral.com/articles/10.1186/s40793-015-0038-5.

Lamed, R., J. Naimark, E. Morgenstern and E. A. Bayer (1987). Specialized cell surface structures in cellulolytic bacteria. J Bacteriol 169(8): 3792-3800.

Lang, K., J. Schuldes, A. Klingl, A. Poehlein, R. Daniel and A. Brune (2015). New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome

analysis of "*Candidatus* Methanoplasma termitum". Appl Environ Microbiol 81(4): 1338-1352.

Langmead, B. and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4): 357-359.

Lanigan, G. W., A. L. Payne and J. E. Peterson (1978). Antimethanogenic drugs and *Heliotropium europaeum* poisoning in penned sheep. Crop Pasture Sci 29(6): 1281-1292.

Larimer, F. W. (1987). Cleavage by ApaI is inhibited by overlapping *dcm* methylation. Nucleic Acids Res 15(21): 9087.

Larsen, T. M., S. K. Boehlein, S. M. Schuster, N. G. J. Richards, J. B. Thoden, H. M. Holden and I. Rayment (1999). Three-dimensional structure of *Escherichia coli* asparagine synthetase B: a short journey from substrate to product. Biochemistry 38(49): 16146-16157.

Lazazzera, B. A. (2000). Quorum sensing and starvation: signals for entry into stationary phase. Curr Opin Microbiol 3(2): 177-182.

Lê, S., J. Josse and F. Husson (2008). FactoMineR: an R package for multivariate analysis. J Stat Softw 25(1): 1-18.

Leadbetter, J. R. and J. A. Breznak (1996). Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. Appl Environ Microbiol 62(10): 3620-3631.

Leadbetter, J. R., L. D. Crosby and J. A. Breznak (1998). *Methanobrevibacter filiformis* sp. nov., A filamentous methanogen from termite hindguts. Arch Microbiol 169(4): 287-292.

Leahy, S. C., W. J. Kelly, E. Altermann, R. S. Ronimus, C. J. Yeoman, D. M. Pacheco, D. Li, Z. Kong, S. McTavish, C. Sang, S. C. Lambie, P. H. Janssen, D. Dey and G. T. Attwood (2010). The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. PloS one 5(1): e8926. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008926.

Leahy, S. C., W. J. Kelly, R. S. Ronimus, N. Wedlock, E. Altermann and G. T. Attwood (2013). Genome sequencing of rumen bacteria and archaea and its application to methane mitigation strategies. Animal 7(s2): 235-243.

Leahy, S. C., W. J. Kelly, D. Li, Y. Li, E. Altermann, S. C. Lambie, F. Cox and G. T. Attwood (2013). The complete genome sequence of *Methanobrevibacter* sp. AbM4. Stand Genomic Sci 8(2): 215-227.

Lee, J. H., S. Kumar, G. H. Lee, D. H. Chang, M. S. Rhee, M. H. Yoon and B. C. Kim (2013). *Methanobrevibacter boviskoreani* sp. nov., isolated from the rumen of Korean native cattle. Int J Syst Evol Microbiol 63(11): 4196-4201.

Lee, J. H., M. S. Rhee, S. Kumar, G. H. Lee, D. H. Chang, D. S. Kim, S. H. Choi, D. W. Lee, M. H. Yoon and B. C. Kim (2013). Genome sequence of *Methanobrevibacter* sp. strain JH1, isolated from rumen of Korean native cattle. Genome Announc 1(1): e00002-13. Retrieved from http://genomea.asm.org/content/1/1/e00002-13.full.

Lee, J. K. and K. N. Houk (1997). A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase. Science 276(5314): 942-945.

Legendre, P. and E. D. Gallagher (2001). Ecologically meaningful transformations for ordination of species data. Oecol 129(2): 271-280.

Legendre, P., J. Oksanen and C. J. F. ter Braak (2011). Testing the significance of canonical axes in redundancy analysis. Methods Ecol Evol 2(3): 269-277.

Leigh, J. A. (2000). Nitrogen fixation in methanogens: the archaeal perspective. Curr Issues Mol Biol 2(4): 125-131.

Levis, R., P. Dunsmuir and G. M. Rubin (1980). Terminal repeats of the *Drosophila* transposable element copia: nucleotide sequence and genomic organization. Cell 21(2): 581-588.

Li, X., Y. Yang, L. Ma, X. Sun, S. Yang, X. Kong and X. Hu (2014). Comparative proteomics analyses of *Kobresia pygmaea* adaptation to environment along an elevational gradient on the central Tibetan Plateau. PloS one 9(6): e98410. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098410.

Lie, T. J., K. C. Costa, B. Lupa, S. Korpole, W. B. Whitman and J. A. Leigh (2012). Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. Proc Natl Acad Sci U S A 109(38): 15473-15478.

Lieb, M. (1991). Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair. Genetics 128(1): 23-27.

Lienard, T., B. Becher, M. Marschall, S. Bowien and G. Gottschalk (1996). Sodium ion translocation by N^5-methyltetrahydromethanopterin: coenzyme M methyltransferase from *Methanosarcina mazei* Gö1 reconstituted in ether lipid liposomes. Eur J Biochem 239(3): 857-864.

Lin, B., J. H. Wang, Y. Lu, Q. Liang and J. X. Liu (2013). *In vitro* rumen fermentation and methane production are influenced by active components of essential oils combined with fumarate. J Anim Physiol Anim Nutr (Berl) 97(1): 1-9.

Lin, C., L. Raskin and D. A. Stahl (1997). Microbial community structure in gastrointestinal tracts of domestic animals: Comparative analyses using rRNA-targeted oligonucleotide probes. FEMS Microbiol Ecol 22(4): 281-294.

Liu, Y. and W. B. Whitman (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. Ann N Y Acad Sci 1125(1): 171-189.

Liu, Y., R. H. White and W. B. Whitman (2010). *Methanococci* use the diaminopimelate aminotransferase (DapL) pathway for lysine biosynthesis. J Bacteriol 192(13): 3304-3310.

Lobo, S. A., A. Brindley, M. J. Warren and L. M. Saraiva (2009). Functional characterization of the early steps of tetrapyrrole biosynthesis and modification *in Desulfovibrio vulgaris* Hildenborough. Biochem J 420(2): 317-325.

Loenen, W. A. M. (2006). *S*-Adenosylmethionine: jack of all trades and master of everything? Biochem Soc Trans 34(2), 330-333.

Longstaff, D. G., S. K. Blight, L. Zhang, K. B. Green-Church and J. A. Krzycki (2007). *In vivo* contextual requirements for UAG translation as pyrrolysine. Mol Microbiol 63(1): 229-241.

Lowe, T. M. and S. R. Eddy (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5): 955-964.

Lu, G. and E. N. Moriyama (2004). Vector NTI, a balanced all-in-one sequence analysis suite. Brief Bioinform 5(4): 378-388.

Lund, E., J. E. Dahlberg, L. Lindahl, S. R. Jaskunas, P. P. Dennis and M. Nomura (1976). Transfer RNA genes between 16*S* and 23*S* rRNA genes in rRNA transcription units of *E. coli*. Cell 7(2): 165-177.

Luo, D., S. Ganesh, J. Koolaard and M. D. Luo (2014). Package 'predictmeans'. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.6475&rep=rep1&type=pdf.

Luton, P. E., J. M. Wayne, R. J. Sharp and P. W. Riley (2002). The *mcrA* gene as an alternative to 16*S* rRNA in the phylogenetic analysis of methanogen populations in landfill. Microbiology 148(11): 3521-3530.

Mackie, R. I. and S. Heath (1979). Enumeration and isolation of lactate-utilizing bacteria from the rumen of sheep. Appl Environ Microbiol 38(3): 416-421.

MacPherson, L. (2014). Agricultural production statistics: June 2014 (final). Retrieved 28.07.2015 from http://www.stats.govt.nz/browse_for_stats/industry_sectors/agriculture-horticulture-forestry/AgriculturalProduction_final_HOTPJun14final.aspx.

Maeder, D. L., I. Anderson, T. S. Brettin, D. C. Bruce, P. Gilna, C. S. Han, A. Lapidus, W. W. Metcalf, E. Saunders, R. Tapia and K. R. Sowers (2006). The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. J Bacteriol 188(22): 7922-7931.

Magni, G., G. Orsomando and N. Raffaelli (2006). Structural and functional properties of NAD kinase, a key enzyme in NADP biosynthesis. Mini Rev Med Chem 6(7): 739-746.

Magoc, T., D. Wood and S. L. Salzberg (2013). EDGE-pro: estimated degree of gene expression in prokaryotic genomes. Evol Bioinform Online 9: 127-136.

Mahapatra, A., A. Patel, J. A. Soares, R. C. Larue, J. K. Zhang, W. W. Metcalf and J. A. Krzycki (2006). Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine. Mol Microbiol 59(1): 56-66.

Mahillon, J. and M. Chandler (1998). Insertion sequences. Microbiol Mol Biol Rev 62(3): 725-774.

Maizel, J. V., A. A. Benson and N. E. Tolbert (1956). Identification of phosphoryl choline as an important constituent of plant sap. Plant Physiol 31(5): 407-408.

Major, T. A., Y. Liu and W. B. Whitman (2010). Characterization of energy-conserving hydrogenase B in *Methanococcus maripaludis*. J Bacteriol 192(15): 4022-4030.

Makarova, K. S., L. Aravind, Y. I. Wolf and E. V. Koonin (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol Direct 6(1): 38. Retrieved fom http://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-6-38.

Makarova, K. S., D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost and E. V. Koonin (2011). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9(6): 467-477.

Makkar, H. P., M. Blümmel and K. Becker (1995). *In vitro* effects of and interactions between tannins and saponins and fate of tannins in the rumen. J Sci Food Agric 69(4): 481-493.

Mander, G. J., A. J. Pierik, H. Huber and R. Hedderich (2004). Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon *Archaeoglobus profundus*. Eur J Biochem 271(6): 1106-1116.

Margolies, M. N. and R. F. Goldberger (1966). Isolation of the fourth enzyme (isomerase) of histidine biosynthesis from *Salmonella typhimurium*. J Biol Chem 241(14): 3262-3269.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley and J. M. Rothberg (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057): 376-380.

Markham, G. D., E. W. Hafner, C. W. Tabor and H. Tabor (1980). *S*-Adenosylmethionine synthetase from *Escherichia coli*. J Biol Chem 255(19): 9082-9092.

Marolewski, A., J. M. Smith and S. J. Benkovic (1994). Cloning and characterization of a new purine biosynthetic enzyme: a non-folate glycinamide ribonucleotide transformylase from *E. coli*. Biochemistry 33(9): 2531-2537.

Marraffini, L. A. and E. J. Sontheimer (2008). CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. Science 322(5909): 1843-1845.

Marraffini, L. A. and E. J. Sontheimer (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11(3): 181-190.

Martin, S. A. (1998). Manipulation of ruminal fermentation with organic acids: a review. J Anim Sci 76(12): 3123-3132.

Mashhadi, Z., H. Xu, L. L. Grochowski and R. H. White (2010). Archaeal RibL: a new FAD synthetase that is air sensitive. Biochemistry 49(40): 8748-8755.

Mashhadi, Z., H. Zhang, H. Xu and R. H. White (2008). Identification and characterization of an archaeon-specific riboflavin kinase. J Bacteriol 190(7): 2615-2618.

Matarazzo, F., A. C. Ribeiro, M. Faveri, C. Taddei, M. B. Martinez and M. P. Mayer (2012). The domain Archaea in human mucosal surfaces. Clin Microbiol Infect 18(9): 834-840.

May, M. S. and S. Hattman (1975). Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host-and R-factor-controlled enzymes. J Bacteriol 123(2): 768-770.

McAllister, T. A. and C. J. Newbold (2008). Redirecting rumen fermentation to reduce methanogenesis. Aust J Exp Agric 48(2): 7-13.

McArthur, J. M. and J. E. Miltimore (1961). Rumen gas analysis by gas-solid chromatography. Can J Anim Sci 41(2): 187-196.

McClure, R., D. Balasubramanian, Y. Sun, M. Bobrovskyy, P. Sumby, C. A. Genco, C. K. Vanderpool and B. Tjaden (2013). Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res 41(14): e140. Retrieved from http://nar.oxfordjournals.org/content/41/14/e140.long.

McDonald, M. J., C. H. Chou, K. B. Swamy, H. D. Huang and J. Y. Leu (2015). The evolutionary dynamics of tRNA-gene copy number and codon-use in *E. coli*. BMC Evol Biol 15(1): 163. Retrieved from https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-015-0441-y.

McGinn, S. M., K. A. Beauchemin, T. Coates and D. Colombatto (2004). Methane emissions from beef cattle: Effects of monensin, sunflower oil, enzymes, yeast, and fumaric acid. J Anim Sci 82(11): 3346-3356.

McIntosh, E. M. and R. H. Haynes (1986). Sequence and expression of the dCMP deaminase gene (DCD1) of *Saccharomyces cerevisiae*. Mol Cell Biol 6(5): 1711-1721.

McMillan, D. G. G., S. A. Ferguson, D. Dey, K. Schröder, H. L. Aung, V. Carbone, G. T. Attwood, R. S. Ronimus, T. Meier, P. H. Janssen and G. M. Cook (2011). A_1A_0-ATP synthase of *Methanobrevibacter ruminantium* couples sodium ions for ATP synthesis under physiological conditions. J Biol Chem 286(46): 39882-39892.

Mendez-Garcia, C., V. Mesa, R. R. Sprenger, M. Richter, M. S. Diez, J. Solano, R. Bargiela, O. V. Golyshina, A. Manteca, J. L. Ramos, J. R. Gallego, I. Llorente, V. A. Martins dos Santos, O. N. Jensen, A. I. Pelaez, J. Sanchez and M. Ferrer (2014). Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. ISME J 8(6): 1259-1274.

Mengin-Lecreulx, D. and J. van Heijenoort (1996). Characterization of the essential gene *glmM* encoding phosphoglucosamine mutase in *Escherichia coli*. J Biol Chem 271(1): 32-39.

Messenger, L. J. and H. Zalkin (1979). Glutamine phosphoribosylpyrophosphate amidotransferase from *Escherichia coli*. Purification and properties. J Biol Chem 254(9): 3382-3392.

Meyer, E., N. J. Leonard, B. Bhat, J. Stubbe and J. M. Smith (1992). Purification and characterization of the *purE*, *purK*, and *purC* gene products: identification of a previously unrecognized energy requirement in the purine biosynthetic pathway. Biochemistry 31(21): 5022-5032.

Middleton, B. (1974). The kinetic mechanism and properties of the cytoplasmic acetoacetyl-coenzyme A thiolase from rat liver. Biochem J 139(1): 109-121.

Millar, G., A. Lewendon, M. Hunter and J. Coggins (1986). The cloning and expression of the *aroL* gene from *Escherichia coli* K12. Purification and complete amino acid sequence of shikimate kinase II, the *aroL*-gene product. Biochem. J 237(2): 427-437.

Miller, M. B., K. Skorupski, D. H. Lenz, R. K. Taylor and B. L. Bassler (2002). Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. Cell 110(3): 303-314.

Miller, T. L. (2001). *Methanobrevibacter*. Bergey's Manual of Systematics of Archaea and Bacteria. W. B. Whitman. Online, Wiley. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781118960608.gbm00496/full.

Miller, T. L. and C. Lin (2002). Description of *Methanobrevibacter gottschalkii* sp. nov., *Methanobrevibacter thaueri* sp. nov., *Methanobrevibacter woesei* sp. nov. and *Methanobrevibacter wolinii* sp. nov. Int J Syst Evol Microbiol 52(3): 819-822.

Miller, T. L., M. J. Wolin, E. Conway de Macario and A. J. Macario (1982). Isolation of *Methanobrevibacter smithii* from human feces. Appl Environ Microbiol 43(1): 227-232.

Ministry for the Environment. (2015). New Zealand's Greenhouse Gas Inventory 1990-2013. Ministry for the Environment. Wellington, New Zealand.

Mitchell, A. D., A. Chappell and K. L. Knox (1979). Metabolism of betaine in the ruminant. J Anim Sci 49(3): 764-774.

Moe, P. W. and H. F. Tyrrell (1979). Methane production in dairy cows. J Dairy Sci 62(10): 1583-1586.

Moore, D. P. and J. F. Thompson (1967). Methionine biosynthesis from *S*-methylcysteine by thiomethyl transfer in *Neurospora*. Biochem Biophys Res Commun 28(3): 474-479.

Morgan, E. A., T. Ikemura and M. Nomura (1977). Identification of spacer tRNA genes in individual ribosomal RNA transcription units of *Escherichia coli*. Proc Natl Acad Sci U S A 74(7): 2710-2714.

Morgavi, D. P., W. J. Kelly, P. H. Janssen and G. Attwood (2012). Rumen microbial (meta)genomics and its application to ruminant production. Animal 7(s1): 184-201

Morgavi, D. P., C. Martin, J. P. Jouany and M. J. Ranilla (2012). Rumen protozoa and methanogenesis: Not a simple cause-effect relationship. Br J Nutr 107(3): 388-397.

Morii, H. and Y. Koga (2003). CDP-2,3-Di-*O*-geranylgeranyl-*sn*-glycerol:L-serine *O*-archaetidyltransferase (archaetidylserine synthase) in the methanogenic archaeon *Methanothermobacter thermautotrophicus*. J Bacteriol 185(4): 1181-1189.

Morii, H., M. Nishihara and Y. Koga (2000). CTP:2,3-di-*O*-geranylgeranyl-*sn*-glycero-1-phosphate cytidyltransferase in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*. J Biol Chem 275(47): 36568-36574.

Morrison, S. D., S. A. Roberts, A. M. Zegeer, W. R. Montfort and V. Bandarian (2008). A new use for a familiar fold: the X-ray crystal structure of GTP-bound GTP cyclohydrolase III from

Methanocaldococcus jannaschii reveals a two metal ion catalytic mechanism. Biochemistry 47(1): 230-242.

Morton, T. A., J. A. Runquist, S. W. Ragsdale, T. Shanmugasundaram, H. G. Wood and L. G. Ljungdahl (1991). The primary structure of the subunits of carbon monoxide dehydrogenase/acetyl-CoA synthase from *Clostridium thermoaceticum*. J Biol Chem 266(35): 23824-23828.

Morvan, B., F. Bonnemoy, G. Fonty and P. Gouet (1996). Quantitative determination of H_2-utilizing acetogenic and sulfate-reducing bacteria and methanogenic archaea from digestive tract of different mammals. Curr Microbiol 32(3): 129-133.

Moss, A. R., J. P. Jouany and J. Newbold (2000). Methane production by ruminants: its contribution to global warming. Ann Zootech 49(3): 231-253.

Mount, D. W. (2008). Using BLOSUM in sequence alignments. CSH Protoc 2008. Retrieved from http://cshprotocols.cshlp.org/content/2008/6/pdb.top39.full.

Muller, E., K. Fahlbusch, R. Walther and G. Gottschalk (1981). Formation of *N*,*N*-dimethylglycine, acetic acid, and butyric acid from betaine by *Eubacterium limosum*. Appl Environ Microbiol 42(3): 439-445.

Muller, V., M. Blaut and G. Gottschalk (1986). Utilization of methanol plus hydrogen by *Methanosarcina barkeri* for methanogenesis and growth. Appl Environ Microbiol 52(2): 269-274.

Muller, V., C. Ruppert and T. Lemker (1999). Structure and function of the A_1A_0-ATPases from methanogenic Archaea. J Bioenerg Biomembr 31(1): 15-27.

Muller, V. (2003). Energy conservation in acetogenic bacteria. Appl Environ Microbiol 69(11): 6345-6353.

Musfeldt, M., M. Selig and P. Schönheit (1999). Acetyl coenzyme A synthetase (ADP forming) from the hyperthermophilic archaeon *Pyrococcus furiosus*: identification, cloning, separate expression of the encoding genes, *acdAI* and *acdBI*, in *Escherichia coli*, and *in vitro* reconstitution of the active heterotetrameric enzyme from its recombinant subunits. J Bacteriol 181(18): 5885-5888.

Myllykallio, H., G. Lipowski, D. Leduc, J. Filee, P. Forterre and U. Liebl (2002). An alternative flavin-dependent mechanism for thymidylate synthesis. Science 297(5578): 105-107.

Nagata, Y., K. Tanaka, T. Iida, Y. Kera, R. Yamada, Y. Nakajima, T. Fujiwara, Y. Fukumori, T. Yamanaka, Y. Koga, S. Tsuji and K. Kawaguchi-Nagata (1999). Occurrence of D-amino acids in a few archaea and dehydrogenase activities in hyperthermophile *Pyrobaculum islandicum*. Biochim Biophys Acta 1435(1-2): 160-166.

Narindrasorasak, S. and W. A. Bridger (1977). Phosphoenolypyruvate synthetase of *Escherichia coli*: molecular weight, subunit composition, and identification of phosphohistidine in phosphoenzyme intermediate. J Biol Chem 252(10): 3121-3127.

Naumann, E., H. Hippe and G. Gottschalk (1983). Betaine: New Oxidant in the Stickland reaction and methanogenesis from betaine and L-Alanine by a *Clostridium sporogenes-Methanosarcina barkeri* Coculture. Appl Environ Microbiol 45(2): 474-483.

Nawrocki, E. P., S. W. Burge, A. Bateman, J. Daub, R. Y. Eberhardt, S. R. Eddy, E. W. Floden, P. P. Gardner, T. A. Jones, J. Tate and R. D. Finn (2015). Rfam 12.0: updates to the RNA families database. Nucleic Acids Res 43(Database issue): D130-D137. Retrieved from http://nar.oxfordjournals.org/content/early/2014/11/11/nar.gku1063.full.

Nawrocki, E. P. and S. R. Eddy (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22): 2933-2935.

Neill, A. R., D. W. Grime and R. M. C. Dawson (1978). Conversion of choline methyl groups through trimethylamine into methane in the rumen. Biochem J 170(3): 529-535.

Nejat, N., G. Vadamalai, R. E. Davis, N. A. Harrison, K. Sijam, M. Dickinson, S. N. Abdullah and Y. Zhao (2013). '*Candidatus* Phytoplasma malaysianum', a novel taxon associated with virescence and phyllody of Madagascar periwinkle (*Catharanthus roseus*). Int J Syst Evol Microbiol 63(2): 540-548.

Neuwald, A. F. and G. V. Stauffer (1985). DNA sequence and characterization of the *Escherichia coli serB* gene. Nucleic Acids Res 13(19): 7025-7039.

Newbold, C. J., S. Lopez, N. Nelson, J. O. Ouda, R. J. Wallace and A. R. Moss (2005). Propionate precursors and other metabolic intermediates as possible alternative electron acceptors to methanogenesis in ruminal fermentation in vitro. Br J Nutr 94(1): 27-35.

Ng, F., S. Kittelmann, M. L. Patchett, G. T. Attwood, P. H. Janssen, J. Rakonjac and D. Gagic (2016). An adhesin from hydrogen-utilizing rumen methanogen *Methanobrevibacter ruminantium* M1 binds a broad range of hydrogen-producing microorganisms. Environ Microbiol. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1462-2920.13155/pdf.

Nielsen, F. S., P. S. Andersen and K. F. Jensen (1996). The B form of dihydroorotate dehydrogenase from *Lactococcus lactis* consists of two different subunits, encoded by the *pyrDb* and *pyrK* genes, and contains FMN, FAD, and [FeS] redox centers. J Biol Chem 271(46): 29359-29365.

Nishihara, M. and Y. Koga (1997). Purification and properties of *sn*-glycerol-1-phosphate dehydrogenase from *Methanobacterium thermoautotrophicum*: characterization of the biosynthetic enzyme for the enantiomeric glycerophosphate backbone of ether polar lipids of Archaea. J Biochem 122(3): 572-576.

Nishimura, Y. and T. Eguchi (2006). Biosynthesis of archaeal membrane lipids: digeranylgeranylglycerophospholipid reductase of the thermoacidophilic archaeon *Thermoplasma acidophilum*. J Biochem 139(6): 1073-1081.

Niu, Y. H. (2014). Effect of different molecular weight of tannins from alpine plants on *in vitro* methane production in yaks [Thesis Abstract]. Master.

Noel, S. J., O. Hojberg, T. Urich and M. Poulsen (2016). Draft genome sequence of "*Candidatus* Methanomethylophilus" sp. 1R26, enriched from bovine rumen, a methanogenic archaeon belonging to the *Methanomassiliicoccales* Order. Genome Announc 4(1): e01734-15. Retrieved from http://genomea.asm.org/content/4/1/e01734-15.full.

Norager, S., K. F. Jensen, O. Bjornberg and S. Larsen (2002). *E. coli* dihydroorotate dehydrogenase reveals structural and functional distinctions between different classes of dihydroorotate dehydrogenases. Structure 10(9): 1211-1223.

Noll, K. M., K. L. Rinehart, R. S. Tanner and R. S. Wolfe (1986). Structure of component B (7-mercaptoheptanoylthreonine phosphate) of the methylcoenzyme M methylreductase system of *Methanobacterium thermoautotrophicum*. PNAS 83(12): 4238-4242.

Nusslein, B., K. J. Chin, W. Eckert and R. Conrad (2001). Evidence for anaerobic syntrophic acetate oxidation during methane production in the profundal sediment of subtropical Lake Kinneret (Israel). Environ Microbiol 3(7): 460-470.

O'Sullivan, O., M. Begley, R. P. Ross, P. D. Cotter and C. Hill (2011). Further identification of novel lantibiotic operons using LanM-based genome mining. Probiotics Antimicrob Proteins 3(1): 27-40.

Ochman, H., J. G. Lawrence and E. A. Groisman (2000). Lateral gene transfer and the nature of bacterial innovation. Nature 405(6784): 299-304.

Ogawa, H., T. Gomi, M. M. Mueckler, M. Fujioka, P. S. Backlund, R. R. Aksamit, C. G. Unson and G. L. Cantoni (1987). Amino acid sequence of *S*-adenosyl-L-homocysteine hydrolase from rat liver as derived from the cDNA sequence. PNAS 84(3): 719-723.

Ohkuma, S. and B. Poole (1978). Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. Proc Natl Acad Sci U S A 75(7): 3327-3331.

Ohmori, H., E. C. Friedberg, R. P. Fuchs, M. F. Goodman, F. Hanaoka, D. Hinkle, T. A. Kunkel, C. W. Lawrence, Z. Livneh, T. Nohmi, L. Prakash, S. Prakash, T. Todo, G. C. Walker, Z. Wang and R. Woodgate (2001). The Y-family of DNA polymerases. Mol Cell 8(1): 7-8.

Ohta, A., T. Obara, Y. Asami and I. Shibuya (1985). Molecular cloning of the *cls* gene responsible for cardiolipin synthesis in *Escherichia coli* and phenotypic consequences of its amplification. J Bacteriol 163(2): 506-514.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens and M. J. Oksanen (2007). The vegan package. Community ecology package: 631-637.

Omi, R., M. Goto, N. Nakagawa, I. Miyahara and K. Hirotsu (2004). Expression, purification and preliminary X-ray characterization of histidinol phosphate phosphatase. Acta Crystallogr D Biol Crystallogr 60(3): 574-576.

Oyeleke, S. B. and T. A. Okusanmi (2008). Isolation and characterization of cellulose hydrolysing microorganism from the rumen of ruminants. Afr J Biotechnol 7(10). Retrieved from http://www.ajol.info/index.php/ajb/article/download/58700/47026.

Ozment, C., J. Barchue, L. J. DeLucas and D. Chattopadhyay (1999). Structural study of *Escherichia coli* NAD synthetase: overexpression, purification, crystallization, and preliminary crystallographic analysis. J Struct Biol 127(3): 279-282.

Padmanabha, J., J. Liu, C. Kurekci, S. E. Denman and C. S. McSweeney (2013). A methylotrophic methanogen isolate from the *Thermoplasmatales* affiliated RCC clade may provide insight into the role of this group in the rumen. In: Proceedings of the 5th Greenhouse Gases and Animal Agriculture Conference, Dublin.

Padyana, A. K. and S. K. Burley (2003). Crystal structure of shikimate 5-dehydrogenase (SDH) bound to NADP: insights into function and evolution. Structure 11(8): 1005-1013.

Paradis, E., K. Strimmer, J. R. Claude, G. Jobb, R. Opgen-Rhein, J. Dutheil, Y. Noel, B. Bolker and J. Lemon (2008). The ape package. Anal Phylogenet Evol. Retrieved from http://www.math.mcmaster.ca/~bolker/eeid/2009/Evolution/2009_Day1_phylogenetics/R_phylo_resources/ape.pdf.

Paris, S., P. M. Wessel and R. Dumas (2002). Overproduction, purification, and characterization of recombinant aspartate semialdehyde dehydrogenase from *Arabidopsis thaliana*. Protein Express Purif 24(1): 99-104.

Patra, A. K. and Z. Yu (2012). Effects of essential oils on methane production and fermentation by, and abundance and diversity of, rumen microbial populations. Appl Environ Microbiol 78(12): 4271-4280.

Passarge, E., B. Horsthemke and R. A. Farber (1999). Incorrect use of the term synteny. Nat Genet 23(4): 387. Retrieved from http://www.boneandcancer.org/MOLab%20Publications%20in%20PDF%20files/He_co-author_Nature%20Genetics_1999.pdf.

Paul, K., J. O. Nonoh, L. Mikulski and A. Brune (2012). "Methanoplasmatales," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. Appl Environ Microbiol 78(23): 8245-8253.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10): 785-786.

Petti, C. A. (2007). Detection and identification of microorganisms by gene amplification and sequencing. Clin Infect Dis 44(8): 1108-1114.

Pfaltz, A., A. Kobelt, R. Huster and R. K. Thauer (1987). Biosynthesis of coenzyme F_{430} in methanogenic bacteria. Identification of 15,173-seco-F_{430}-173-acid as an intermediate. Eur J Biochem 170(1-2): 459-467.

Phillips, J. D., F. G. Whitby, J. P. Kushner and C. P. Hill (2003). Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase. EMBO J 22(23): 6225-6233.

Phillipson, A. T. (1939). The movements of the pouches of the stomach of sheep. Q J Exp Physiol Cogn Med Sci 29(4): 395-415.

Pihl, T. D., S. Sharma and J. N. Reeve (1994). Growth phase-dependent transcription of the genes that encode the two methyl coenzyme M reductase isoenzymes and N^5-methyltetrahydromethanopterin:coenzyme M methyltransferase in *Methanobacterium thermoautotrophicum* ΔH. J Bacteriol 176(20): 6384-6391.

Pitson, S. M., G. L. Mendz, S. Srinivasan and S. L. Hazell (1999). The tricarboxylic acid cycle of *Helicobacter pylori*. Eur J Biochem 260(1): 258-267.

Pohlert, T. (2014). The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). Retrieved from https://www.researchgate.net/profile/Thorsten_Pohlert/publication/274015222_PMCMR_Calculate_Pairwise_Multiple_Comparisons_of_Mean_Rank_Sums._R_package_version_1.1/links/551299b10cf20bfdad518fc9.pdf.

Pollak, N., C. Dölle and M. Ziegler (2007). The power to reduce: pyridine nucleotides - small molecules with a multitude of functions. Biochem J 402(2): 205-218.

Poppi, D. P., D. J. Minson and J. H. Ternouth (1981). Studies of cattle and sheep eating leaf and stem fractions of grasses. 1. The voluntary intake, digestibility and retention time in the reticulo-rumen. Crop Pasture Sci 32(1): 99-108.

Porat, I., B. W. Waters, Q. Teng and W. B. Whitman (2004). Two biosynthetic pathways for aromatic amino acids in the archaeon *Methanococcus maripaludis*. J Bacteriol 186(15): 4940-4950.

Poulsen, M., C. Schwab, B. B. Jensen, R. M. Engberg, A. Spang, N. Canibe, O. Hojberg, G. Milinovich, L. Fragner, C. Schleper, W. Weckwerth, P. Lund, A. Schramm and T. Urich

(2013). Methylotrophic methanogenic *Thermoplasmata* implicated in reduced methane emissions from bovine rumen. Nat Commun 4: 1428.

Prat, L., I. U. Heinemann, H. R. Aerni, J. Rinehart, P. O'Donoghue and D. Soll (2012). Carbon source-dependent expansion of the genetic code in bacteria. Proc Natl Acad Sci U S A 109(51): 21070-21075.

Prieto, C., C. Garcia-Estrada, D. Lorenzana and J. F. Martin (2012). NRPSsp: non-ribosomal peptide synthase substrate predictor. Bioinformatics 28(3): 426-427.

Primak, Y. A., M. Du, M. C. Miller, D. H. Wells, A. T. Nielsen, W. Weyler and Z. Q. Beck (2011). Characterization of a feedback-resistant mevalonate kinase from the archaeon *Methanosarcina mazei*. Appl Environ Microbiol 77(21): 7772-7778.

Pugh, D. M. (2002). The EU precautionary bans of animal feed additive antibiotics. Toxicol Lett 128(1–3): 35-44.

Puigbo, P., I. G. Bravo and S. Garcia-Vallve (2008). CAIcal: a combined set of tools to assess codon usage adaptation. Biol Direct 3(1): 38.

Quitterer, F., P. Beck, A. Bacher and M. Groll (2013). Structure and reaction mechanism of pyrrolysine synthase (PylD). Angew Chem Int Ed Engl 52(27): 7033-7037.

Quitterer, F., A. List, W. Eisenreich, A. Bacher and M. Groll (2012). Crystal structure of methylornithine synthase (PylB): insights into the pyrrolysine biosynthesis. Angew Chem Int Ed Engl 51(6): 1339-1342.

Raffaelli, N., F. M. Pisani, T. Lorenzi, M. Emanuelli, A. Amici, S. Ruggieri and G. Magni (1997). Characterization of nicotinamide mononucleotide adenylyltransferase from thermophilic Archaea. J Bacteriol 179(24): 7718-7723.

Rahmstorf, S. and D. Coumou (2011). Increase of extreme events in a warming world. Proc Natl Acad Sci U S A 108(44): 17905-17909.

Randau, L., K. Calvin, M. Hall, J. Yuan, M. Podar, H. Li and D. Soll (2005). The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. Proc Natl Acad Sci U S A 102(50): 17934-17939.

Rankin, C. A., G. C. Haslam and R. H. Himes (1993). Sequence and expression of the gene for N^{10}-formyltetrahydrofolate synthetase from *Clostridium cylindrosporum*. Protein Sci 2(2): 197-205.

Rea, S., J. P. Bowman, S. Popovski, C. Pimm and A. D. G. Wright (2007). *Methanobrevibacter millerae* sp. nov. and *Methanobrevibacter olleyae* sp. nov., methanogens from the ovine and bovine rumen that can utilize formate for growth. Int J Syst Evol Microbiol 57(3): 450-456.

Rees, W. M. R., D. Lloyd and A. G. Williams (1995). The effects of co-cultivation with the acetogen *Acetitomaculum ruminis* on the fermentative metabolism of the rumen fungi *Neocallimastix patriciarum* and *Neocallimastix* sp. strain L2. FEMS Microbiol Lett 133(1): 175-180.

Reeve, J. N., J. Nolling, R. M. Morgan and D. R. Smith (1997). Methanogenesis: genes, genomes, and who's on first? J Bacteriol 179(19): 5975-5986.

Reguera, G., K. D. McCarthy, T. Mehta, J. S. Nicoll, M. T. Tuominen and D. R. Lovley (2005). Extracellular electron transfer via microbial nanowires. Nature 435(7045): 1098-1101.

Rhodes, D. and A. D. Hanson (1993). Quaternary ammonium and tertiary sulfonium compounds in higher plants. Annu Rev Plant Biol 44(1): 357-384.

Rice, P., I. Longden and A. Bleasby (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16(6): 276-277.

Ricke, S. C., D. M. Schaefer and T. S. Chang (1994). Influence of methylamine on anaerobic rumen bacterial growth and plant fiber digestion. Bioresource Technol 50(3): 253-257.

Rinke, C., P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J. F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies, J. A. Dodsworth, B. P. Hedlund, G. Tsiamis, S. M. Sievert, W. T. Liu, J. A. Eisen, S. J. Hallam, N. C. Kyrpides, R. Stepanauskas, E. M. Rubin, P. Hugenholtz and T. Woyke (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature 499(7459): 431-437.

Ripley, B., B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth and M. B. Ripley (2015). Package 'MASS'. CRAN repository. Retrieved from ftp://ftp3.pl.freebsd.org/pub/mirrors/CRAN/web/packages/MASS/MASS.pdf.

Ritchie, R. J. and J. Gibson (1987). Permeability of ammonia and amines in *Rhodobacter sphaeroides* and *Bacillus firmus*. Arch Biochem Biophys 258(2): 332-341.

Romantsov, T., Z. Guan and J. M. Wood (2009). Cardiolipin and the osmotic stress responses of bacteria. Biochim Biophys Acta 1788(10): 2092-2100.

Roongsawang, N., K. Washio and M. Morikawa (2010). Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J Mol Sci 12(1): 141-172.

Rosenthal, A. Z., E. G. Matson, A. Eldar and J. R. Leadbetter (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. ISME J 5(7): 1133-1142.

Rospert, S., D. Linder, J. Ellermann and R. K. Thauer (1990). Two genetically distinct methyl-coenzyme M reductases in *Methanobacterium thermoautotrophicum* strain Marburg and ΔH. Eur J Biochem 194(3): 871-877.

Rosewarne, C. P., P. Greenfield and P. Hendry (2016). Draft genome sequence of *Methanobrevibacter acididurans* strain ATM^{T} from anaerobic fermenter. Unpublished raw data.

Rotaru, A. E., M. P. Shrestha, F. Liu, M. Shrestha, M. Embree, K. Zengler, C. Wardman, K. P. Nevin and D. R. Lovley (2014). A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to *Methanosaeta* for the reduction of carbon dioxide to methane. Energy Environ Sci 7(1): 408-415.

Rotaru, A. E., P. M. Shrestha, F. Liu, B. Markovaite, S. Chen, K. P. Nevin and D. R. Lovley (2014). Direct interspecies electron transfer between *Geobacter metallireducens* and *Methanosarcina barkeri*. Appl Environ Microbiol 80(15): 4599-4605.

Rowe, J. B., M. L. Loughnan, J. V. Nolan and R. A. Leng (1979). Secondary fermentation in the rumen of a sheep given a diet based on molasses. Br J Nutr 41(2): 393-397.

Rstudio Team (2015). Rstudio: integrated development for R. Rstudio, Inc., Boston, MA.

Russell, J. B. and G. M. Cook (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol Rev 59(1): 48-62.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream and B. Barrell (2000). Artemis: sequence visualization and annotation. Bioinformatics 16(10): 944-945.

Sachdeva, G., K. Kumar, P. Jain and S. Ramachandran (2005). SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. Bioinformatics 21(4): 483-491.

Sakai, S., H. Imachi, S. Hanada, A. Ohashi, H. Harada and Y. Kamagata (2008). *Methanocella paludicola* gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage 'Rice Cluster I', and proposal of the new archaeal order *Methanocellales* ord. nov. Int J Syst Evol Microbiol 58(4): 929-936.

Sakamoto, N., A. M. Kotre and M. A. Savageau (1975). Glutamate dehydrogenase from *Escherichia coli*: purification and properties. J Bacteriol 124(2): 775-783.

Saleem, F., S. Bouatra, A. C. Guo, N. Psychogios, R. Mandal, S. M. Dunn, B. N. Ametaj and D. S. Wishart (2013). The bovine ruminal fluid metabolome. Metabolomics 9(2): 360-378.

Samuel, B. S., E. E. Hansen, J. K. Manchester, P. M. Coutinho, B. Henrissat, R. Fulton, P. Latreille, K. Kim, R. K. Wilson and J. I. Gordon (2007). Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. Proc Natl Acad Sci U S A 104(25): 10643-10648.

Samuel, G. and P. Reeves (2003). Biosynthesis of *O*-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and *O*-antigen assembly. Carbohydr Res 338(23): 2503-2519.

Sandkvist, M. (2001). Biology of type II secretion. Mol Microbiol 40(2): 271-283.

Santiveri, C. M. and M. A. Jimenez (2010). Tryptophan residues: scarce in proteins but strong stabilizers of β-hairpin peptides. Biopolymers 94(6): 779-790.

Sasarman, A., A. Nepveu, Y. Echelard, J. Dymetryszyn, M. Drolet and C. Goyer (1987). Molecular cloning and sequencing of the *hemD* gene of *Escherichia coli* K-12 and preliminary data on the Uro operon. J Bacteriol 169(9): 4257-4262.

Sato, T., H. Atomi and T. Imanaka (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science 315(5814): 1003-1006.

Sauer, K. and R. K. Thauer (1998). Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*--identification of the active-site histidine in the corrinoid-harboring subunit MtaC by site-directed mutagenesis. Eur J Biochem 253(3): 698-705.

Saunders, M. A. and A. S. Lea (2008). Large contribution of sea surface warming to recent increase in Atlantic hurricane activity. Nature 451(7178): 557-560.

Savant, D. V., Y. S. Shouche, S. Prakash and D. R. Ranade (2002). *Methanobrevibacter acididurans* sp. nov., a novel methanogen from a sour anaerobic digester. Int J Syst Evol Microbiol 52(4): 1081-1087.

Sayers, E. W., T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, Z. Lu, T. L. Madden, T. Madej, D. R. Maglott, A. Marchler-Bauer, V. Miller, I. Mizrachi, J. Ostell, A. Panchenko, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, D. Slotta, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, Y. Wang, W. John Wilbur, E. Yaschenko and J. Ye (2011). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 38(Database issue): D5-16. Retrieved from http://nar.oxfordjournals.org/content/39/suppl_1/D38.long.

Scanlan, P. D., F. Shanahan and J. R. Marchesi (2008). Human methanogen diversity and incidence in healthy and diseased colonic groups using *mcrA* gene analysis. BMC Microbiol 8(1). Retrieved from http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-8-79.

Schäfer, G., M. Engelhard and V. Müller (1999). Bioenergetics of the Archaea. Microbiol Mol Biol Rev 63(3): 570-620.

Scheller, S., M. Goenrich, R. Boecher, R. K. Thauer and B. Jaun (2010). The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. Nature 465(7298): 606-608.

Schiermeier, Q. (2011). Increased flood risk linked to global warming. Nature 470(7334): 316.

Schlaberg, R., K. E. Simmon and M. A. Fisher (2012). A systematic approach for discovering novel, clinically relevant bacteria. Emerg Infect Dis 18(3): 422-430.

Schlame, M. (2008). Thematic Review Series: Glycerolipids. Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. J Lipid Res 49(8): 1607-1620.

Schlame, M. and M. Ren (2009). The role of cardiolipin in the structural organization of mitochondrial membranes. Biochim Biophys Acta 1788(10): 2080-2083.

Schleifer, K. H. and O. Kandler (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev 36(4): 407-477.

Schluter, D., E. A. Clifford, M. Nemethy and J. S. McKinnon (2004). Parallel evolution and inheritance of quantitative traits. Am Nat 163(6): 809-822.

Schultz, J. E. and A. Matin (1991). Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*. J Mol Biol 218(1): 129-140.

Scofield, M. A., W. S. Lewis and S. M. Schuster (1990). Nucleotide sequence of *Escherichia coli asnB* and deduced amino acid sequence of asparagine synthetase B. J Biol Chem 265(22): 12895-12902.

Seedorf, H., S. Kittelmann, G. Henderson and P. H. Janssen (2014). RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. PeerJ 2: e494. Retrieved from https://peerj.com/articles/494/.

Seedorf, H., S. Kittelmann and P. H. Janssen (2015). Few highly abundant operational taxonomic units dominate within rumen methanogenic archaeal species in New Zealand sheep and cattle. Appl Environ Microbiol 81(3): 986-995.

Sekowska, A., P. Bertin and A. Danchin (1998). Characterization of polyamine synthesis pathway in *Bacillus subtilis* 168. Mol Microbiol 29(3): 851-858.

Selbmann, L., D. Isola, M. Fenice, L. Zucconi, K. Sterflinger and S. Onofri (2012). Potential extinction of Antarctic endemic fungal species as a consequence of global warming. Sci Total Environ 438: 127-134.

Serina, L., C. Blondin, E. Krin, O. Sismeiro, A. Danchin, H. Sakamoto, A. M. Gilles and O. Barzu (1995). *Escherichia coli* UMP-kinase, a member of the aspartokinase family, is a hexamer regulated by guanine nucleotides and UTP. Biochemistry 34(15): 5066-5074.

Setzke, E., R. Hedderich, S. Heiden and R. K. Thauer (1994). H_2: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Composition and properties. Eur J Biochem 220(1): 139-148.

Shariat, N., R. E. Timme, J. B. Pettengill, R. Barrangou and E. G. Dudley (2015). Characterization and evolution of *Salmonella* CRISPR-Cas systems. Microbiology 161(2): 374-386.

Sharp, P. M. and W. H. Li (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15(3): 1281-1295.

She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C.-Y. Christina, I. G. Clausen, B. A. Curtis and A. De Moors (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. PNAS 98(14): 7835-7840.

Shen, B. W., D. H. Dyer, J. Y. Huang, L. D'Ari, J. Rabinowitz and B. L. Stoddard (1999). The crystal structure of a bacterial, bifunctional 5, 10 methylene-tetrahydrofolate dehydrogenase/cyclohydrolase. Prot Sci 8(6): 1342-1349.

Shi, W., C. D. Moon, S. C. Leahy, D. Kang, J. Froula, S. Kittelmann, C. Fan, S. Deutsch, D. Gagic, H. Seedorf, W. J. Kelly, R. Atua, C. Sang, P. Soni, D. Li, C. S. Pinares-Patino, J. C. McEwan, P. H. Janssen, F. Chen, A. Visel, Z. Wang, G. T. Attwood and E. M. Rubin (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Res 24(9): 1517-1525.

Shieh, J. S. and W. B. Whitman (1987). Pathway of acetate assimilation in autotrophic and heterotrophic methanococci. J Bacteriol 169(11): 5327-5329.

Shima, S., E. Warkentin, R. K. Thauer and U. Ermler (2002). Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen. J Biosci Bioeng 93(6): 519-530.

Shimizu, H., S. Yamagata, R. Masui, Y. Inoue, T. Shibata, S. Yokoyama, S. Kuramitsu and T. Iwama (2001). Cloning and overexpression of the *oah1* gene encoding *O*-acetyl-L-homoserine sulfhydrylase of *Thermus thermophilus* HB8 and characterization of the gene product. Biochim Biophys Acta 1549(1): 61-72.

Shimon, L. J., E. A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham and F. Frolow (1997). A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly. Structure 5(3): 381-390.

Shin, E. C., B. R. Choi, W. J. Lim, S. Y. Hong, C. L. An, K. M. Cho, Y. K. Kim, J. M. An, J. M. Kang, S. S. Lee, H. Kim and H. D. Yun (2004). Phylogenetic analysis of archaea in three fractions of cow rumen based on the 16*S* rDNA sequence. Anaerobe 10(6): 313-319.

Shinkai, T., O. Enishi, M. Mitsumori, K. Higuchi, Y. Kobayashi, A. Takenaka, K. Nagashima and M. Mochizuki (2012). Mitigation of methane production from cattle by feeding cashew nut shell liquid. J Dairy Sci 95(9): 5308-5316.

Shinzato, N., T. Matsumoto, I. Yamaoka, T. Oshima and A. Yamagishi (1999). Phylogenetic diversity of symbiotic methanogens living in the hindgut of the lower termite *Reticulitermes speratus* analyzed by PCR and *in situ* hybridization. Appl Environ Microbiol 65(2): 837-840.

Shrestha, P. M., A. E. Rotaru, Z. M. Summers, M. Shrestha, F. Liu and D. R. Lovley (2013). Transcriptomic and genetic analysis of direct interspecies electron transfer. Appl Environ Microbiol 79(7): 2397-2404.

Siguier, P., E. Gourbeyre, A. Varani, B. Ton-Hoang and M. Chandler (2015). Everyman's guide to bacterial insertion sequences. Microbiol Spectr 3(2). Retrieved from http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.MDNA3-0030-2014.

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon and M. Chandler (2006). ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34(Database issue): D32-36. Retrieved from http://nar.oxfordjournals.org/content/34/suppl_1/D32.long.

Silk, G. W., B. F. Matthews and B. G. Gengenbach (1994). Cloning and expression of the soybean *dapA* gene encoding dihydrodipicolinate synthase. Plant Mol Biol 26(3): 989-993.

Skillman, L. C., P. N. Evans, G. E. Naylor, B. Morvan, G. N. Jarvis and K. N. Joblin (2004). 16*S* ribosomal DNA-directed PCR primers for ruminal methanogens and identification of methanogens colonising young lambs. Anaerobe 10(5): 277-285.

Silva, K. L., E. R. Duarte, C. E. S. Freitas, F. O. Abrão and L. C. Geraseev (2014). Protozoários ruminais de novilhos de corte criados em pastagem tropical durante o período seco. Ciênc Anim Bras 15: 259-265.

Simcock, D. C., K. N. Joblin, I. Scott, D. M. Burgess, C. W. Rogers, W. E. Pomroy and H. V. Simpson (1999). Hypergastrinaemia, abomasal bacterial population densities and pH in sheep infected with *Ostertagia circumcincta*. Int J Parasitol 29(7): 1053-1063.

Siragusa, R. J., J. J. Cerda, M. M. Baig, C. W. Burgin and F. L. Robbins (1988). Methanol production from the degradation of pectin by human colonic bacteria. Am J Clin Nutr 47(5): 848-851.

Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush and J. N. Reeve (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. J Bacteriol 179(22): 7135-7155.

Smith, E. A. and G. T. Macfarlane (1996). Studies on amine production in the human colon: Enumeration of amine forming bacteria and physiological effects of carbohydrate and pH. Anaerobe 2(5): 285-297.

Smith, J. M. and H. A. Daum, 3rd (1986). Nucleotide sequence of the *purM* gene encoding 5'-phosphoribosyl-5-aminoimidazole synthetase of *Escherichia coli* K12. J Biol Chem 261(23): 10632-10636.

Smith, K. S. and J. G. Ferry (1999). A plant-type (beta-class) carbonic anhydrase in the thermophilic methanoarchaeon *Methanobacterium thermoautotrophicum*. J Bacteriol 181(20): 6247-6253.

Smith, P. H. and R. E. Hungate (1958). Isolation and characterization of *Methanobacterium ruminantiumn*. sp. J Bacteriol 75(6): 713-718.

Smith, T. F. and M. S. Waterman (1981). Identification of common molecular subsequences. J Mol Biol 147(1): 195-197.

Sneath, P. H. and R. R. Sokal (1962). Numerical taxonomy. Nature 193: 855-860.

Soliva, C. R., I. K. Hindrichsen, L. Meile, M. Kreuzer and A. Machmüller (2003). Effects of mixtures of lauric and myristic acid on rumen methanogens and methanogenesis *in vitro*. Lett Appl Microbiol 37(1): 35-39.

Sollinger, A., C. Schwab, T. Weinmaier, A. Loy, A. T. Tveit, C. Schleper and T. Urich (2016). Phylogenetic and genomic analysis of *Methanomassiliicoccales* in wetlands and animal intestinal tracts reveals clade-specific habitat preferences. FEMS Microbiol Ecol 92(1): fiv149. Retrieved from http://femsec.oxfordjournals.org/content/92/1/fiv149.

Soupene, E., L. He, D. Yan and S. Kustu (1998). Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein. Proc Natl Acad Sci U S A 95(12): 7030-7034.

Sparrow, C. P. and C. R. Raetz (1985). Purification and properties of the membrane-bound CDP-diglyceride synthetase from *Escherichia coli*. J Biol Chem 260(22): 12084-12091.

Sprenger, G. A. (1995). Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12. Arch Microbiol 164(5): 324-330.

Sprenger, G. A., U. Schorken, G. Sprenger and H. Sahm (1995). Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains. Eur J Biochem 230(2): 525-532.

Sprenger, G. A., U. Schorken, G. Sprenger and H. Sahm (1995). Transaldolase B of *Escherichia coli* K-12: cloning of its gene, *talB*, and characterization of the enzyme from recombinant strains. J Bacteriol 177(20): 5930-5936.

Sprott, G. D., I. Ekiel and G. B. Patel (1993). Metabolic pathways *in Methanococcus jannaschii* and other methanogenic bacteria. Appl Environ Microbiol 59(4): 1092-1098.

Srinivasan, G., C. M. James and J. A. Krzycki (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science 296(5572): 1459-1462.

Stackebrandt, E. and J. Ebers (2006). Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 33(4): 152.

Staden, R., K. F. Beal and J. K. Bonfield (2000). The Staden package, 1998. Methods Mol Biol 132: 115-130.

Staples, C. R., S. Lahiri, J. Raymond, L. Von Herbulis, B. Mukhophadhyay and R. E. Blankenship (2007). Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*. J Bacteriol 189(20): 7392-7398.

Statistics New Zealand. (2015). Global New Zealand – International trade, investment, and travel profile: Year ended December 2014. Retrieved 27.04.2015 from www.stats.govt.nz.

Staudenmaier, H., B. Van Hove, Z. Yaraghi and V. Braun (1989). Nucleotide sequences of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. J Bacteriol 171(5): 2626-2633.

Steenhoudt, O. and J. Vanderleyden (2000). *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. FEMS Microbiol Rev 24(4): 487-506.

Stotz, A. and P. Linder (1990). The ADE2 gene from *Saccharomyces cerevisiae*: sequence and new vectors. Gene 95(1): 91-98.

Strieker, M., A. Tanovic and M. A. Marahiel (2010). Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol 20(2): 234-240.

Stupperich, E., K. E. Hammel, G. Fuchs and R. K. Thauer (1983). Carbon monoxide fixation into the carboxyl group of acetyl coenzyme A during autotrophic growth of *Methanobacterium*. FEBS Lett 152(1): 21-23.

Subharat, S., D. Shu, T. Zheng, B. M. Buddle, P. H. Janssen, D. Luo and D. N. Wedlock (2015). Vaccination of cattle with a methanogen protein produces specific antibodies in the saliva which are stable in the rumen. Vet Immunol Immunopathol 164(3-4): 201-207.

Tajima, K., T. Nagamine, H. Matsui, M. Nakamura and R. I. Aminov (2001). Phylogenetic analysis of archaeal 16*S* rRNA libraries from the rumen suggests the existence of a novel group of archaea not associated with known methanogens. FEMS Microbiol Lett 200(1): 67-72.

Takai, K. and K. Horikoshi (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. Genetics 152(4): 1285-1297.

Tallant, T. C. and J. A. Krzycki (1997). Methylthiol:coenzyme M methyltransferase from *Methanosarcina barkeri*, an enzyme of methanogenesis from dimethylsulfide and methylmercaptopropionate. J Bacteriol 179(22): 6902-6911.

Tallant, T. C., L. Paul and J. A. Krzycki (2001). The MtsA subunit of the methylthiol:coenzyme M methyltransferase of *Methanosarcina barkeri* catalyses both half-reactions of corrinoid-dependent dimethylsulfide: coenzyme M methyl transfer. J Biol Chem 276(6): 4485-4493.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30(12): 2725-2729.

Tan, H. Y., C. C. Sieo, N. Abdullah, J. B. Liang, X. D. Huang and Y. W. Ho (2011). Effects of condensed tannins from *Leucaena* on methane production, rumen fermentation and populations of methanogens and protozoa *in vitro*. Anim Feed Sci Technol 169(3-4): 185-193.

Tan, H. Y., C. C. Sieo, C. M. Lee, N. Abdullah, J. B. Liang and Y. W. Ho (2011). Diversity of bovine rumen methanogens *in vitro* in the presence of condensed tannins, as determined by sequence analysis of 16*S* rRNA gene library. J Microbiol 49(3): 492-498.

Tang, X., S. Ezaki, S. Fujiwara, M. Takagi, H. Atomi and T. Imanaka (1999). The tryptophan biosynthesis gene cluster *trpCDEGFBA* from *Pyrococcus kodakaraensis* KOD1 is regulated at the transcriptional level and expressed as a single mRNA. MGG 262(4-5): 815-821.

Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova and E. V. Koonin (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29(1): 22-28.

Tavendale, M. H., L. P. Meagher, D. Pacheco, N. Walker, G. T. Attwood and S. Sivakumaran (2005). Methane production from *in vitro* rumen incubations with *Lotus pedunculatus* and *Medicago sativa*, and effects of extractable condensed tannin fractions on methanogenesis. Anim Feed Sci Technol 123(1): 403-419.

Taylor, C. D., B. C. McBride, R. S. Wolfe and M. P. Bryant (1974). Coenzyme M, essential for growth of a rumen strain of *Methanobacterium ruminantium*. J Bacteriol 120(2): 974-975.

Taylor, C. D. and R. S. Wolfe (1974). Structure and methylation of coenzyme M($HSCH_2CH_2SO_3$). J Biol Chem 249(15): 4879-4885.

Teeling, H., J. Waldmann, T. Lombardot, M. Bauer and F. O. Glockner (2004). TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. BMC Bioinformatics 5(1): 1.

Temin, H. M. (1981). Structure, variation and synthesis of retrovirus long terminal repeat. Cell 27(1): 1-3.

Tesmer, J. J., T. L. Stemmler, J. E. Penner-Hahn, V. J. Davisson and J. L. Smith (1994). Preliminary X-ray analysis of *Escherichia coli* GMP synthetase: determination of anomalous scattering factors for a cysteinyl mercury derivative. Proteins 18(4): 394-403.

Thauer, R. K. and L. G. Bonacker (1994). Biosynthesis of coenzyme F_{430}, a nickel porphinoid involved in methanogenesis. Ciba Found Symp 180: 210-222.

Thauer, R. K., K. Jungermann and K. Decker (1977). Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev 41(1): 100-180.

Thauer, R. K. (1998). Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture. Microbiology 144(9): 2377-2406.

Thauer, R. K., A. K. Kaster, H. Seedorf, W. Buckel and R. Hedderich (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol 6: 579-591.

Thauer, R. K., A. K. Kaster, M. Goenrich, M. Schick, T. Hiromoto and S. Shima (2010). Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and H_2 storage. Annu Rev Biochem 79: 507-536.

Thoden, J. B., G. N. Phillips, Jr., T. M. Neal, F. M. Raushel and H. M. Holden (2001). Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center. Biochemistry 40(24): 6989-6997.

Thoden, J. B., F. M. Raushel, M. M. Benning, I. Rayment and H. M. Holden (1999). The structure of carbamoyl phosphate synthetase determined to 2.1 A resolution. Acta Crystallogr D Biol Crystallogr 55(1): 8-24.

Thompson, G. A., Jr. (1969). The metabolism of 2-aminoethylphosphonate lipids in *Tetrahymena pyriformis*. Biochim Biophys Acta 176(2): 330-338.

Thompson, J. D., T. J. Gibson and D. G. Higgins (2002). Multiple sequence alignment using ClustalW and ClustalX. Current Protocols in Bioinformatics, 2-3.

Thompson, L. D., L. D. Brandon, D. T. Nieuwlandt and C. J. Daniels (1989). Transfer RNA intron processing in the halophilic archaebacteria. Can J Microbiol 35(1): 36-42.

Tietze, M., A. Beuchle, I. Lamla, N. Orth, M. Dehler, G. Greiner and U. Beifuss (2003). Redox potentials of methanophenazine and CoB-S-S-CoM, factors involved in electron transport in methanogenic archaea. Chembiochem 4(4): 333-335.

Tock, M. R. and D. T. Dryden (2005). The biology of restriction and anti-restriction. Curr Opin Microbiol 8(4): 466-472.

Ton-That, H., L. A. Marraffini and O. Schneewind (2004). Protein sorting to the cell wall envelope of Gram-positive bacteria. Biochim Biophys Acta 1694(1-3): 269-278.

Torrents, E., G. Buist, A. Liu, R. Eliasson, J. Kok, I. Gibert, A. Graslund and P. Reichard (2000). The anaerobic (class III) ribonucleotide reductase from *Lactococcus lactis*. Catalytic properties and allosteric regulation of the pure enzyme system. J Biol Chem 275(4): 2463-2471.

Towne, G., T. G. Nagaraja, R. T. Brandt, Jr. and K. E. Kemp (1990). Dynamics of ruminal ciliated protozoa in feedlot cattle. Appl Environ Microbiol 56(10): 3174-3178.

Tschech, A. and N. Pfennig (1984). Growth yield increase linked to caffeate reduction in *Acetobacterium wo*odii. Arch Microbiol 137(2): 163-167.

Tsukagoshi, N., G. Tamura and K. Arima (1970). A novel protoplast-bursting factor (surfactin) obtained from *Bacillus subtilis* IAM 1213. I. The effects of surfactin on *Bacillus megaterium* KM. Biochim Biophys Acta 196(2): 204-210.

Tu, J. and W. Zillig (1982). Organization of rRNA structural genes in the archaebacterium *Thermoplasma acidophilum*. Nucleic Acids Res 10(22): 7231-7245.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. Biometrics 5(2): 99-114.

Uchiyama, T., K. Ito, K. Mori, H. Tsurumaru and S. Harayama (2010). Iron-corroding methanogen isolated from a crude-oil storage tank. Appl Environ Microbiol 76(6): 1783-1788.

Van Soest, P. J. (1984). Some physical characteristics of dietary fibres and their influence on the microbial ecology of the human colon. Proc Nutr Soc 43(1): 25-33.

Vantcheva, Z. M., K. Prodhan and R. W. Hemken (1970). Rumen methanol *in vivo* and *in vitro*. J Dairy Sci 53(10): 1511-1514.

Vartak, N. B., L. Liu, B. M. Wang and C. M. Berg (1991). A functional *leuABCD* operon is required for leucine synthesis by the tyrosine-repressible transaminase in *Escherichia coli* K-12. J Bacteriol 173(12): 3864-3871.

Veiga, M. C., M. K. Jain, W. Wu, R. I. Hollingsworth and J. G. Zeikus (1997). Composition and role of extracellular polymers in methanogenic granules. Appl Environ Microbiol 63(2): 403-407.

Vernikos, G. S. and J. Parkhill (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics 22(18): 2196-2203.

Vesth, T., K. Lagesen, Ö. Acar and D. Ussery (2013). CMG-Biotools, a free workbench for basic comparative microbial genomics. PloS one 8(4): e60120. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0060120.

Vinokur, J. M., T. P. Korman, Z. Cao and J. U. Bowie (2014). Evidence of a novel mevalonate pathway in archaea. Biochemistry 53(25): 4161-4168.

Wadhwa, D. R. and A. D. Care (2002). The absorption of phosphate ions from the ovine reticulorumen. Vet J 163(2): 182-186.

Wallace, R. J., N. R. McEwan, F. M. McIntosh, B. Teferedegne and C. J. Newbold (2002). Natural products as manipulators of rumen fermentation. Asian Australas. J. Anim. Sci 15(10): 1458-1468.

Wang, H., D. P. Fewer, L. Holm, L. Rouhiainen and K. Sivonen (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc Natl Acad Sci U S A 111(25): 9259-9264.

Wang, J., K. A. Stieglitz, J. P. Cardia and E. R. Kantrowitz (2005). Structural basis for ordered substrate binding and cooperativity in aspartate transcarbamoylase. Proc Natl Acad Sci U S A 102(25): 8881-8886.

Wang, L., S. Chen, T. Xu, K. Taghizadeh, J. S. Wishnok, X. Zhou, D. You, Z. Deng and P. C. Dedon (2007). Phosphorothioation of DNA in bacteria by *dnd* genes. Nat Chem Biol 3(11): 709-710.

Ward, D. E., S. W. Kengen, J. van Der Oost and W. M. de Vos (2000). Purification and characterization of the alanine aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its role in alanine production. J Bacteriol 182(9): 2559-2566.

Wargo, M. J. (2013). Homeostasis and catabolism of choline and glycine betaine: lessons from *Pseudomonas aeruginosa*. Appl Environ Microbiol 79(7): 2112-2120.

Wassenaar, R. W., P. J. Daas, W. J. Geerts, J. T. Keltjens and C. van der Drift (1996). Involvement of methyltransferase-activating protein and methyltransferase 2 isoenzyme II in methylamine:coenzyme M methyltransferase reactions in *Methanosarcina barkeri* Fusaro. J Bacteriol 178(23): 6937-6944.

Watanabe, T., S. Asakawa, A. Nakamura, K. Nagaoka and M. Kimura (2004). DGGE method for analyzing 16S rDNA of methanogenic archaeal community in paddy field soil. FEMS Microbiol Lett 232(2): 153-163.

Watkins, A. J., E. G. Roussel, R. J. Parkes and H. Sass (2014). Glycine betaine as a direct substrate for methanogens (*Methanococcoides* spp.). Appl Environ Microbiol 80(1): 289-293.

Watson, G. M., J. P. Yu and F. R. Tabita (1999). Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea. J Bacteriol 181(5): 1569-1575.

Wegener, G., V. Krukenberg, D. Riedel, H. E. Tegetmeyer and A. Boetius (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. Nature 526(7574): 587-590.

Weidner, S., A. Becker, I. Bonilla, S. Jaenicke, J. Lloret, I. Margaret, A. Puhler, J. E. Ruiz-Sainz, S. Schneiker-Bekel, R. Szczepanowski, J. M. Vinardell, S. Zehner and M. Gottfert (2012). Genome sequence of the soybean symbiont *Sinorhizobium fredii* HH103. J Bacteriol 194(6): 1617-1618.

Weisburg, W. G., S. M. Barns, D. A. Pelletier and D. J. Lane (1991). 16*S* ribosomal DNA amplification for phylogenetic study. J Bacteriol 173(2): 697-703.

Welander, P. V. and W. W. Metcalf (2005). Loss of the mtr operon in *Methanosarcina* blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. Proc Natl Acad Sci U S A 102(30): 10664-10669.

Weljie, A. M., J. Newton, P. Mercier, E. Carlson and C. M. Slupsky (2006). Targeted profiling: quantitative analysis of 1H NMR metabolomics data. Anal Chem 78(13): 4430-4442.

Welte, C. and U. Deppenmeier (2011). Membrane-bound electron transport in *Methanosaeta thermophila*. J Bacteriol 193(11): 2868-2870

Welte, C. and U. Deppenmeier (2011). Re-evaluation of the function of the F_{420} dehydrogenase in electron transport of *Methanosarcina mazei*. FEBS J 278(8): 1277-1287.

Welte, C. and U. Deppenmeier (2014). Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. Biochim Biophys Acta 1837(7): 1130-1147.

White, P. J., G. Millar and J. R. Coggins (1988). The overexpression, purification and complete amino acid sequence of chorismate synthase from *Escherichia coli* K12 and its comparison with the enzyme from *Neurospora crassa*. Biochem. J 251: 313-322.

White, R. H. (1985). Biosynthesis of coenzyme M (2-mercaptoethanesulfonic acid). Biochemistry 24(23): 6487-6493.

White, R. H. (2003). The biosynthesis of cysteine and homocysteine in *Methanococcus jannaschii*. BBA-Gen Subjects 1624(1): 46-53.

White, R. H. (2004). L-Aspartate semialdehyde and a 6-deoxy-5-ketohexose 1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*. Biochemistry 43(23): 7618-7627.

White, R. H. and H. Xu (2006). Methylglyoxal is an intermediate in the biosynthesis of 6-deoxy-5-ketofructose-1-phosphate: a precursor for aromatic amino acid biosynthesis in *Methanocaldococcus jannaschii*. Biochemistry 45(40): 12366-12379.

White, R. W. (1969). Viable bacteria inside the rumen ciliate *Entodinium caudatum*. J Gen Microbiol 56(3): 403-408.

Widdel, F. and N. Pfennig (1981). Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. I. Isolation of new sulfate-reducing bacteria enriched with acetate from saline environments. Description of *Desulfobacter postgatei* gen. nov., sp. nov. Arch Microbiol 129(5): 395-400.

Widyastuti, Y., S. K. Lee, K. Suzuki and T. Mitsuoka (1992). Isolation and characterization of rice-straw degrading *Clostridia* from cattle rumen. J Vet Med Sci 54(1): 185-188.

Wiedemann, I., T. Bottiger, R. R. Bonelli, A. Wiese, S. O. Hagge, T. Gutsmann, U. Seydel, L. Deegan, C. Hill, P. Ross and H. G. Sahl (2006). The mode of action of the lantibiotic lacticin 3147 - a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II. Mol Microbiol 61(2): 285-296.

Willcott, M. R. (2009). MestRe Nova. J Am Chem Soc 131(36): 13180-13180.

Wilding, E. I., J. R. Brown, A. P. Bryant, A. F. Chalker, D. J. Holmes, K. A. Ingraham, S. Iordanescu, C. Y. So, M. Rosenberg and M. N. Gwynn (2000). Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in Gram-positive cocci. J Bacteriol 182(15): 4319-4327.

Wishart, D. S., T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert (2013). HMDB 3.0 - The Human Metabolome Database in 2013. Nucleic Acids Res 41(Database issue): D801-807. Retrieved from http://nar.oxfordjournals.org/content/early/2012/11/16/nar.gks1065.long.

Witting, K. and R. D. Sussmuth (2011). Discovery of antibacterials and other bioactive compounds from microorganisms-evaluating methodologies for discovery and generation of non-ribosomal peptide antibiotics. Curr Drug Targets 12(11): 1547-1559.

Wolfram, F., E. N. Kitova, H. Robinson, M. T. Walvoort, J. D. Codee, J. S. Klassen and P. L. Howell (2014). Catalytic mechanism and mode of action of the periplasmic alginate epimerase AlgG. J Biol Chem 289(9): 6006-6019.

Wolin, M. J. (1976). Interactions between H_2-producing and methane-producing species. Microbial production and utilization of gases (H_2, CH_4, CO). H. G. Schlegel, E. G. Gottingen and N. Pfennig. Göttingen, Deutsche Forschungsgemeinschaft: 141-150.

Wolin, M. J., T. L. Miller and C. S. Stewart (1997). Microbe-microbe interactions. The rumen microbial ecosystem. P. N. Hobson and C. S. Stewart. London, Blackie Academic & Professional: 467-491.

Wong, T. Y. and S. D. Schwartzbach (2015). Protein mis-termination initiates genetic diseases, cancers, and restricts bacterial genome expansion. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 33(3): 255-285.

Wood, G. E., A. K. Haydock and J. A. Leigh (2003). Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*. J Bacteriol 185(8): 2548-2554.

Wright, A. D., A. F. Toovey and C. L. Pimm (2006). Molecular identification of methanogenic archaea from sheep in Queensland, Australia reveal more uncultured novel archaea. Anaerobe 12(3): 134-139.

Wright, A. D. G., C. H. Auckland and D. H. Lynn (2007). Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. Appl Environ Microbiol 73(13): 4206-4210.

Wubbolts, M. G., P. Terpstra, J. B. van Beilen, J. Kingma, H. A. Meesters and B. Witholt (1990). Variation of cofactor levels in *Escherichia coli*. Sequence analysis and expression of the *pncB* gene encoding nicotinic acid phosphoribosyltransferase. J Biol Chem 265(29): 17665-17672.

Wugeditsch, T., N. E. Zachara, M. Puchberger, P. Kosma, A. A. Gooley and P. Messner (1999). Structural heterogeneity in the core oligosaccharide of the S-layer glycoprotein from *Aneurinibacillus thermoaerophilus* DSM 10155. Glycobiology 9(8): 787-795.

Xia, J., R. Mandal, I. V. Sinelnikov, D. Broadhurst and D. S. Wishart (2012). MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis. Nucleic Acids Res 40(Web Server issue): W127-W133. Retrieved from http://nar.oxfordjournals.org/content/40/W1/W127.long.

Xing, R. Y. and W. B. Whitman (1991). Characterization of enzymes of the branched-chain amino acid biosynthetic pathway in *Methanococcus* spp. J Bacteriol 173(6): 2086-2092.

Yamagata, S. (1987). Partial purification and some properties of homoserine *O*-acetyltransferase of a methionine auxotroph of *Saccharomyces cerevisiae*. J Bacteriol 169(8): 3458-3463.

Yamagata, S., M. Isaji, K. Nakamura, S. Fujisaki, K. Doi, S. Bawden and R. D'Andrea (1994). Overexpression of the *Saccharomyces cerevisiae* MET17/MET25 gene in *Escherichia coli* and

comparative characterization of the product with *O*-acetylserine. *O*-acetylhomoserine sulfhydrylase of the yeast. Appl Microbiol Biotechnol 42(1): 92-99.

Yamagishi, T., S. Ishida and S. Nishida (1964). Isolation of toxigenic strains of *Clostridium perfringens* from soil. J Bacteriol 88: 646-652.

Yamao, F., A. Muto, Y. Kawauchi, M. Iwami, S. Iwagami, Y. Azumi and S. Osawa (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. Proc Natl Acad Sci U S A 82(8): 2306-2309.

Yamashita, S., H. Hemmi, Y. Ikeda, T. Nakayama and T. Nishino (2004). Type 2 isopentenyl diphosphate isomerase from a thermoacidophilic archaeon *Sulfolobus shibatae*. Eur J Biochem 271(6): 1087-1093.

Yanagita, K., Y. Kamagata, M. Kawaharasaki, T. Suzuki, Y. Nakamura and H. Minato (2000). Phylogenetic analysis of methanogens in sheep rumen ecosystem and detection of *Methanomicrobium mobile* by fluorescence *in situ* hybridization. Biosci Biotechnol Biochem 64(8): 1737-1742.

Yang, Z., A. Savchenko, A. Yakunin, R. Zhang, A. Edwards, C. Arrowsmith and L. Tong (2003). Aspartate dehydrogenase, a novel enzyme identified from structural and functional studies of TM1643. J Biol Chem 278(10): 8804-8808.

Yasmin, A., J. G. Kenny, J. Shankar, A. C. Darby, N. Hall, C. Edwards and M. J. Horsburgh (2010). Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. J Bacteriol 192(4): 1122-1130.

Yasueda, H., Y. Kawahara and S. Sugimoto (1999). *Bacillus subtilis yckG* and *yckF* encode two key enzymes of the ribulose monophosphate pathway used by methylotrophs, and *yckH* is required for their expression. J Bacteriol 181(23): 7154-7160.

Yip, W. S., N. G. Vincent and S. J. Baserga (2013). Ribonucleoproteins in archaeal pre-rRNA processing and modification. Archaea 2013: 614735. Retrieved from http://downloads.hindawi.com/journals/arch/2013/614735.pdf.

Yoshihisa, T. (2014). Handling tRNA introns, archaeal way and eukaryotic way. Front Genet 5: 213.

Yoshinaga, M. Y., L. Wormer, M. Elvert and K. U. Hinrichs (2012). Novel cardiolipins from uncultured methane-metabolizing archaea. Archaea 2012: 832097. Retrieved from http://downloads.hindawi.com/journals/arch/2012/832097.pdf.

Yoshioka, Y., S. Kurei and Y. Machida (2001). Identification of a monofunctional aspartate kinase gene of *Arabidopsis thaliana* with spatially and temporally regulated expression. Genes Genet Syst 76(3): 189-198.

Yu, N. Y., J. R. Wagner, M. R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. C. Sahinalp, M. Ester, L. J. Foster and F. S. Brinkman (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26(13): 1608-1615.

Zadoks, R. N., H. M. Griffiths, M. A. Munoz, C. Ahlstrom, G. J. Bennett, E. Thomas and Y. H. Schukken (2011). Sources of *Klebsiella* and *Raoultella* species on dairy farms: be careful where you walk. J Dairy Sci 94(2): 1045-1051.

Zamenhof, S. and H. H. Eichhorn (1967). Study of microbial evolution through loss of biosynthetic functions: establishment of defective mutants. Nature 216(5114): 456-458.

Zeikus, J. G. and D. L. Henning (1975). *Methanobacterium arbophilicum* sp.nov. An obligate anaerobe isolated from wetwood of living trees. Antonie Van Leeuwenhoek 41(4): 543-552.

Zerbino, D. R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. Current Protocol of Bioinformatics 11-5. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2952100/.

Zhang, D. L., L. Daniels and C. D. Poulter (1990). Biosynthesis of archaebacterial membranes. Formation of isoprene ethers by a prenyl transfer reaction. J Am Chem Soc 112(3): 1264-1265.

Zhang, J. K., A. K. White, H. C. Kuettner, P. Boccazzi and W. W. Metcalf (2002). Directed mutagenesis and plasmid-based complementation in the methanogenic archaeon *Methanosarcina acetivorans* C2A demonstrated by genetic analysis of proline biosynthesis. J Bacteriol 184(5): 1449-1454.

Zhang, Y., P. V. Baranov, J. F. Atkins and V. N. Gladyshev (2005). Pyrrolysine and selenocysteine use dissimilar decoding strategies. J Biol Chem 280(21): 20740-20751.

Zhang, Y., J. Desharnais, S. E. Greasley, G. P. Beardsley, D. L. Boger and I. A. Wilson (2002). Crystal structures of human GAR Tfase at low and high pH and with substrate β-GAR. Biochemistry 41(48): 14206-14215.

Zhang, Y., M. Morar and S. E. Ealick (2008). Structural biology of the purine biosynthetic pathway. Cell Mol Life Sci 65(23): 3699-3724.

Zhao, C., Y. Kumada, H. Imanaka, K. Imamura and K. Nakanishi (2006). Cloning, overexpression, purification, and characterization of *O*-acetylserine sulfhydrylase-B from *Escherichia coli*. Prot Expr Purif 47(2): 607-613.

Zhao, X. and W. A. van der Donk (2016). Structural characterization and bioactivity analysis of the two-component lantibiotic Flv system from a ruminant bacterium. Cell Chem Biol 23(2): 246-256.

Zhou, M., E. Hernandez-Sanabria and L. L. Guan (2009). Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. Appl Environ Microbiol 75(20): 6524-6533.

Zwieb, C. and S. Bhuiyan (2010). Archaea signal recognition particle shows the way. Archaea 2010: 485051. Retrieved from http://downloads.hindawi.com/journals/arch/2010/485051.pdf.

Appendix

Table A.3.1. Manual functional annotation of the methanogenic archaeon isolate ISO4-H5 predicted ORFs. Table excludes hypothetical proteins.

****

AMINO ACID METABOLISM

Glutamate/glutamine

AR505_0145 glutamate dehydrogenase GdhA
AR505_0091 glutamine synthase alpha subunit GlnA

Arginine

AR505_0671 argininosuccinate lyase ArgH
AR505_0672 argininosuccinate synthase ArgG
AR505_0673 *N*-acetyl-gamma-glutamyl-phosphate reductase ArgC
AR505_0674 bifunctional ornithine acetyltransferase/*N*-acetylglutamate synthase protein ArgJ
AR505_0675 acetylglutamate kinase ArgB
AR505_0676 acetylornithine aminotransferase ArgD
AR505_1401 ornithine carbamoyltransferase ArgF

General

AR505_1429 aspartate aminotransferase Ast1
AR505_1666 aspartate aminotransferase Ast2

Alanine

AR505_1199 alanine aminotransferase Alt

Aspartate/asparagine

AR505_0736 asparagine synthase (glutamine-hydrolyzing) AsnB1
AR505_1303 asparagine synthase (glutamine-hydrolyzing) AsnB2

Lysine

AR505_0160 aspartate kinase LysC
AR505_0157 L,L-diaminopimelate aminotransferase DapL
AR505_0158 diaminopimelate epimserase DapF
AR505_0159 diaminopimelate decarboxylase LysA
AR505_0161 dihydrodipicolinate synthase DapA
AR505_0491 aspartate-semialdehyde dehydrogenase Asd
AR505_0728 dihydrodipicolinate reductase DapB
AR505_1120 aspartokinase with LysC domain

Pyrrolysine

AR505_1322 pyrrolysine synthase PylD
AR505_1323 (2R,3R)-3-methylornithyl-N^6-lysine synthase PylC
AR505_1324 methylornithine synthase PylB
AR505_1325 pyrrolysine--tRNA ligase PylS

Methionine

AR505_0293 homoserine dehydrogenase MetL
AR505_0694 homoserine *O*-succinyltransferase MetA
AR505_0738 *O*-acetylhomoserine/*O*-acetylserine sulfhydrylase MetZ/CysK3

SAM cycle

AR505_0547 *S*-adenosyl-L-homocysteine hydrolase AhcY1
AR505_1788 *S*-adenosyl-L-homocysteine hydrolase AhcY2
AR505_0467 *S*-adenosylmethionine synthetase MetK1
AR505_0579 *S*-adenosylmethionine synthetase MetK2
AR505_1118 *S*-adenosylhomocystein nucleosidase

Threonine

AR505_0610 homoserine kinase ThrB
AR505_0611 threonine synthase ThrC

Chorismate

AR505_1148 triose-phosphate isomerase TpiA
AR505_0508 fructose 1,6-bisphosphate aldolase Fba
AR505_0509 3-dehydroquinate synthase AroB
AR505_0510 shikimate 5-dehydrogenase AroE
AR505_0511 shikimate kinase AroL
AR505_0512 3-phosphoshikimate 1-carboxyvinyltransferase AroA
AR505_1449 chorismate synthase AroC

Tryptophan

AR505_1158 anthranilate synthase component I TrpE
AR505_1159 anthranilate synthase component II TrpG
AR505_1160 anthranilate phosphoribosyltransferase TrpD
AR505_1161 indole-3-glycerol phosphate synthase TrpC
AR505_1162 phosphoribosylanthranilate isomerase TrpF
AR505_1163 tryptophan synthase beta subunit TrpB
AR505_1164 tryptophan synthase alpha subunit TrpA
AR505_1151 tryptophan synthase beta subunit truncated TrpB

Phenylalanine/tyrosine

AR505_0490 chorismate mutase AroH
AR505_1450 chorismate mutase AroH
AR505_1745 prephenate dehydrogenase TyrA
AR505_0515 prephenate dehydratase PheA

Proline

AR505_0108 glutamate-5-semialdehyde dehydrogenase ProA
AR505_0109 glutamate 5-kinase ProB
AR505_1634 pyrroline-5-carboxylate reductase ProC

Histidine

AR505_0486 bifunctional phosphoribosyl-ATP pyrophosphatase/ phosphoribosyl-AMP cyclohydrolase HisI
AR505_0487 imidazoleglycerol phosphate synthase HisF
AR505_0488 imidazoleglycerol-phosphate dehydratase HisB
AR505_0494 ATP phosphoribosyl transferase
AR505_0495 histidinol-phosphate aminotransferase HisC
AR505_0496 imidazole glycerol phosphate synthase HisH
AR505_0497 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide ribonucleotide isomerase HisA
AR505_1073 histidinal dehydrogenase HisD
AR505_1770 histidinol-phosphate phosphatase HisK

Serine

AR505_1664 phosphoglycerate dehydrogenase SerA
AR505_1665 phosphoserine amidotransferase SerC
AR505_0073 phosphoserine phosphatase SerB1
AR505_0757 phosphoserine phosphatase SerB2

Cysteine

AR505_1192 serine acetyltransferase CysE
AR505_0799 cysteine synthase CysK1
AR505_0800 cysteine synthase CysM
AR505_0695 cysteine synthase CysK2

Isoleucine/leucine/valine

AR505_0152 acetolactate synthase large subunit IlvB1
AR505_0153 acetolactate synthase large subunit IlvB2
AR505_0542 acetolactate synthase large subunit IlvB3
AR505_1768 acetolactate synthase small subunit IlvN
AR505_1767 branched-chain-amino-acid aminotransferase IlvE
AR505_0150 ketol-acid reductoisomerase IlvC
AR505_1452 dihydroxy-acid dehydratase IlvD
AR505_0631 2-isopropylmalate/homocitrate/citramalate synthase LeuA/CimA
AR505_0632 3-isopropylmalate dehydratase large subunit LeuC
AR505_0633 3-isopropylmalate dehydratase small subunit LeuD
AR505_0634 3-isopropylmalate dehydrogenase LeuB

Polyamines

AR505_1607 arginase/agmatinase family protein SpeB
AR505_0376 *N*-carbamoyl-D-amino acid amidohydrolase AguB
AR505_0268 pyruvoyl-dependent arginine decarboxylase PdaD

Salvage- general

AR505_0999 indolepyruvate Fdx oxidoreductase alpha subunit IorA
AR505_0998 indolepyruvate Fdx oxidoreductase beta subunit IorB

CELL CYCLE

Cell division

AR505_0446 cell division control protein Cdc48
AR505_0854 cell division control protein Cdc48
AR505_1412 cell division ATPase MinD
AR505_0975 cell division protein FtsZ1
AR505_1792 cell division protein FtsZ2

Chromosome replication

AR505_1277 ATP-dependent DNA ligase DnlI
AR505_0001 Orc1/Cdc6 family replication initiation protein
AR505_1205 family replication initiation protein Cdc6

AR505_0130 DNA polymerase family B PolB
AR505_0374 DNA polymerase III PolC
AR505_1037 DNA polymerase IV
AR505_1438 DNA polymerase II large subunit DP2 PolD2
AR505_1816 DNA polymerase II small subunit DP1 PolD1
AR505_0009 DNA primase small subunit PriA
AR505_1142 DNA primase DnaG
AR505_1782 DNA primase large subunit PriB
AR505_1668 OB fold nucleic acid binding domain-containing protein
AR505_0828 replication factor C large subunit RfcL
AR505_1209 replication factor C small subunit
AR505_0018 replicative DNA helicase Mcm
AR505_0209 ribonuclease HII RnhB
AR505_0420 ribonuclease H
AR505_0461 flap endonuclease Fen
AR505_0603 DNA-binding protein
AR505_0125 replication-associated recombination protein A

Genome segregation

AR505_0033 DNA gyase B subunit GyrB
AR505_0034 DNA gyase A subunit GyrA
AR505_1609 chromosome segregation and condensation protein ScpB
AR505_1610 chromosome segregation and condensation protein ScpA
AR505_1611 chromosome segregation protein SMC
AR505_0388 DNA topoisomerase VI subunit A
AR505_0389 DNA topoisomerase VI subunit B
AR505_1299 DNA topoisomerase I TopA

CELL ENVELOPE

Cell surface proteins

AR505_0005 adhesin-like protein
AR505_0061 adhesin-like protein
AR505_0353 adhesin-like protein
AR505_0354 adhesin-like protein
AR505_0355 adhesin-like protein
AR505_0407 adhesin-like protein
AR505_0594 adhesin-like protein
AR505_0614 adhesin-like protein
AR505_0646 adhesin-like protein
AR505_0654 adhesin-like protein
AR505_0657 adhesin-like protein
AR505_0658 adhesin-like protein
AR505_0660 adhesin-like protein
AR505_0664 adhesin-like protein
AR505_0666 adhesin-like protein
AR505_0668 adhesin-like protein
AR505_0670 adhesin-like protein
AR505_0704 adhesin-like protein
AR505_0744 adhesin-like protein
AR505_0807 adhesin-like protein
AR505_0851 adhesin-like protein
AR505_0874 adhesin-like protein
AR505_0993 adhesin-like protein
AR505_0985 adhesin-like protein
AR505_0991 adhesin-like protein
AR505_0992 adhesin-like protein
AR505_1032 adhesin-like protein
AR505_1033 adhesin-like protein
AR505_1155 adhesin-like protein
AR505_1173 adhesin-like protein
AR505_1290 adhesin-like protein
AR505_1509 adhesin-like protein
AR505_1521 adhesin-like protein
AR505_1524 adhesin-like protein
AR505_1532 adhesin-like protein
AR505_1534 adhesin-like protein
AR505_1547 adhesin-like protein
AR505_1549 adhesin-like protein
AR505_1559 adhesin-like protein
AR505_1560 adhesin-like protein
AR505_1561 adhesin-like protein
AR505_1707 adhesin-like protein
AR505_1715 adhesin-like protein
AR505_1741 adhesin-like protein
AR505_1761 adhesin-like protein

Expolysaccharide synthesis

AR505_0539 polysaccharide biosynthesis protein
AR505_0540 cell wall biosynthesis protein glycosyltransferase family
AR505_0543 CDP-glucose 4,6-dehydratase RfbG
AR505_0549 glucose-1-phosphate cytidylyltransferase RfbF
AR505_0550 glycosyl transferase GT2 family
AR505_0551 UDP-galactopyranose mutase Glf
AR505_0552 glucose-1-phosphate thymidylyltransferase RfbA
AR505_0553 dTDP-glucose 4,6-dehydratase RfbB
AR505_0554 dTDP-4-dehydrorhamnose 3,5-epimerase RfbC
AR505_0555 polysaccharide/polyol phosphate ABC transporter permease protein
AR505_0556 polysaccharide/polyol phosphate ABC transporter ATP-binding protein
AR505_0558 LPS biosynthesis protein LICD family
AR505_0559 glycosyl transferase GT8 family
AR505_0561 glycosyl transferase GT2 family
AR505_1106 dolichol kinase (7 TMHs)
AR505_1754 D-sedoheptuloase 7-phosphate isomerase
AR505_1763 glycosyl transferase GT2 family
AR505_1770 D-glycerol-D-manno-heptose-7-phosphate kinase
AR505_1771 D-glycerol-D-manno-heptose 1,7-bisphosphate 7-phosphatase
AR505_1772 D-glycerol-D-manno-heptose 1-phosphate guanylyltransferase
AR505_1778 glycosyl transferase GT2 family
AR505_1786 glycosyl transferase family protein
AR505_1787 glycosyl transferase family protein

Cell envelope phospholipid biosynthesis

AR505_0081 UDP-glucose 4-epimerase GalE
AR505_0123 phosphoglucosamine mutase GlmM
AR505_0361 cell wall biosynthesis protein Mur ligase family
AR505_0541 NAD-dependent epimerase/dehydratase
AR505_0542 acetolactate synthase large subunit IlvB
AR505_1608 phosphoglucosamine mutase GlmM

CELLULAR PROCESSES

Oxidative stress response

AR505_0212 CoA-disulfide reductase cdr
AR505_0263 glycolate oxidase D subunit GlcD
AR505_0630 desulfoferrodoxin Dfx
AR505_0700 rubredoxin
AR505_0766 peroxiredoxin AhpC
AR505_1047 rubredoxin
AR505_1353 rubrerythrin Rbr
AR505_0047 thioredoxin TrxA
AR505_0048 thioredoxin-disulfide reductase TrxB1

Stress response

AR505_1343 universal stress protein UspA
AR505_0443 universal stress protein UspA
AR505_0294 universal stress protein UspA

CENTRAL CARBON METABOLISM

Acetate

AR505_0429 pyruvate Fdx oxidoreductase gamma subunit PorC
AR505_0430 pyruvate Fdx oxidoreductase delta subunit PorD
AR505_0431 pyruvate Fdx oxidoreductase alpha subunit PorA
AR505_0432 pyruvate Fdx oxidoreductase beta subunit PorB
AR505_1282 acetyl-CoA synthetase AcsA

Aromatic ccompounds

AR505_1280 carboxymuconolactone decarboxylase family protein

Bicarbonate

AR505_0752 nitrate/sulfonate/bicarbonate ABC transporter ATPase NtrD
AR505_0753 nitrate/sulfonate/bicarbonate ABC transporter permease protein NtrB
AR505_0754 nitrate/sulfonate/bicarbonate ABC transporter substrate-binding protein
AR505_1245 carbonic anhydrase Cab

Gluconeogenesis

AR505_1769 bifunctional phosphoglucose/phosphomannose isomerase
AR505_0154 glyceraldehyde-3-phosphate dehydrogenase Gap
AR505_0155 phosphoglycerate kinase Pgk
AR505_0427 phosphoenolpyruvate synthase PpsA
AR505_0470 phosphopyruvate hydratase Eno
AR505_0474 2,3-bisphosphoglycerate-independent phosphoglycerate mutase ApgM

AR505_0560 bifunctional phosphoglucose/phosphomannose isomerase
AR505_0942 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase GpmA
AR505_1104 phosphoenolpyruvate synthase PpsA
AR505_1148 triose-phosphate isomerase TpiA
AR505_1149 fructose 1,6-bisphosphatase Fbp
AR505_0082 aldose 1-epimerase
AR505_0932 fructose-bisphosphate aldolase Fba

Glycolate salvage pathway

AR505_0753 phosphoglycolate/pyridoxal phosphate phosphatase family
AR505_0604 phosphoglycolate phosphatase
AR505_0762 glycolate oxidase subunit GlcD

PRPP synthesis

AR505_1513 ribose 5-phosphate isomerase A RpiA
AR505_1685 ribose-phosphate diphosphokinase Prs

Ribulose monophosphate pathway

AR505_0014 bifunctional hexulose-6-phosphate synthase/ribonuclease regulator
AR505_0177 6-phospho 3-hexuloisomerase hxlB
AR505_1785 ribokinase RbsK
AR505_1648 transketolase subunit B
AR505_1649 transketolase subunit A
AR505_1647 fructose-6-phosphate aldolase Fsa
AR505_0461 ribulose-phosphate 3-epimerase Rpe
AR505_1604 ribulose-5-phosphate 4-epimerase and related epimerases and aldolases

Tricarboxylic cycle

AR505_0022 fumarate hydratase FumA
AR505_0023 fumarate hydratase FumB
AR505_1780 malate dehydrogenase Mdh
AR505_0592 aconitate hydratase 1 AcnA
AR505_0678 citrate synthase
AR505_0531 isopropylmalate/isohomocitrate dehydrogenases LeuB

ENERGY METABOLISM

Electron transfer

AR505_0289 NADPH-dependent FMN reductase
AR505_0446 flavodoxin
AR505_0451 Fdx
AR505_0452 Fdx
AR505_0627 NADPH-dependent FMN reductase
AR505_0681 4Fe-4S Fdx iron-sulfur binding domain-containing protein
AR505_0692 anaerobic sulfatase maturase
AR505_0725 4Fe-4S Fdx iron-sulfur binding domain-containing protein
AR505_0763 electron transfer flavoprotein beta subunit
AR505_0764 electron transfer flavoprotein alpha subunit
AR505_0775 NADPH-dependent FMN reductase
AR505_0825 flavodoxin
AR505_0902 Fe-S oxidoreductase
AR505_0915 NADPH-dependent FMN reductase
AR505_0953 4Fe-4S Fdx iron-sulfur binding domain-containing protein
AR505_0976 NADPH-dependent FMN reductase
AR505_0978 flavodoxin-like protein
AR505_1039 NADPH-dependent FMN reductase
AR505_1068 4Fe-4S Fdx iron-sulfur binding domain-containing protein
AR505_1189 flavoredoxin family protein
AR505_1320 4Fe-4S Fdx iron-sulfur binding domain-containing protein
AR505_1338 NADPH-dependent FMN reductase
AR505_1410 NADPH-dependent FMN reductase
AR505_1434 Fdx
AR505_1512 4Fe-4S-binding-domain containing ABC transporter ATP-binding protein
AR505_1579 4Fe-4S Fdx iron-sulfur binding domain-containing protein
AR505_1818 A_1A_0 archaeal ATP synthase subunit D
AR505_1819 A_1A_0 archaeal ATP synthase subunit B
AR505_1820 A_1A_0 archaeal ATP synthase subunit A
AR505_1821 A_1A_0 archaeal ATP synthase subunit F
AR505_1822 A_1A_0 archaeal ATP synthase subunit C
AR505_1823 A_1A_0 archaeal ATP synthase subunit E
AR505_1824 A_1A_0 archaeal ATP synthase subunit K
AR505_1825 A_1A_0 archaeal ATP synthase subunit I
AR505_1826 A_1A_0 archaeal ATP synthase subunit H

Alcohol metabolism

AR505_0483 NADP-dependent alcohol dehydrogenase Adh

H_2 metabolism

AR505_1476 methyl-viologen-reducing hydrogenase alpha subunit MvhA
AR505_1477 methyl-viologen-reducing hydrogenase gamma subunit MvhG
AR505_1478 methyl-viologen-reducing hydrogenase delta subunit MvD

Methanogenesis pathway

AR505_0040 CoB-CoM heterodisulfide reductase subunit D HdrD
AR505_0168 CoB-CoM heterodisulfide reductase subunit D HdrD
AR505_0273 CoB--CoM heterodisulfide reductase subunit C HdrC
AR505_0274 CoB--CoM heterodisulfide reductase subunit B HdrB
AR505_0679 CoB--CoM heterodisulfide reductase subunit B HdrB
AR505_1479 CoB--CoM heterodisulfide reductase subunit A HdrA
AR505_0949 methanol-cobalamin methyltransferase B subunit MtaB2
AR505_0950 methanol corrinoid protein MtaC2
AR505_0951 methanol-cobalamin methyltransferase B subunit MtaB1
AR505_0952 methanol corrinoid protein MtaC1
AR505_1064 methyltransferase cognate corrinoid protein
AR505_1066 methylcobalamin:CoM methyltransferase MtsA
AR505_1067 methylthiol:corrinoid protein mtsB
AR505_1327 monomethylamine methyltransferase MtmB1
AR505_1328 monomethylamine methyltransferase MtmB2
AR505_1329 methyltransferase cognate corrinoid proteins MtmC1
AR505_1330 methyltransferase cognate corrinoid proteins MtmC2
AR505_1332 dimethylamine:corrinoid methyltransferase MtbB
AR505_1333 dimethylamine corrinoid protein MtbC
AR505_0772 trimethylamine:corrinoid methyltransferase MttB
AR505_0773 trimethylamine: corrinoid proteins MttC
AR505_1331 methylamine permease
AR505_1035 methylcobalamin:CoM methyltransferase MtaA1
AR505_1063 methylcobalamin:CoM methyltransferase MtaA2
AR505_0110 methyltransferase cognate corrinoid protein
AR505_1404 methylcobalamin:CoM methyltransferase MtaA3
AR505_1575 methylcobalamin:CoM methyltransferase MtaA4
AR505_1391 methyl-CoM reductase C subunit McrC
AR505_1392 methyl-CoM reductase component A2 AtwA
AR505_1396 methyl-CoM reductase alpha subunit McrA
AR505_1397 methyl-CoM reductase gamma subunit McrG
AR505_1398 methyl-CoM reductase operon protein D McrD
AR505_1399 methyl-CoM reductase beta subunit McrB
AR505_1622 $F_{420}H_2$ dehydrogenase subunit N FpoN
AR505_1623 $F_{420}H_2$ dehydrogenase subunit M FpoM
AR505_1624 $F_{420}H_2$ dehydrogenase subunit L FpoL
AR505_1625 $F_{420}H_2$ dehydrogenase subunit K FpoK
AR505_1626 hypothetical transmembrane protein
AR505_1627 NADH:ubiquinone oxidoreductase J FpoJ
AR505_1628 NADH:quinone oxidoreductase I FpoI
AR505_1629 NADH dehydrogenase subunit H FpoH
AR505_1630 NADH dehydrogenase subunit D FpoD
AR505_1631 NADH dehydrogenase subunit C FpoC
AR505_1632 NADH dehydrogenase subunit B FpoB
AR505_1633 NADH dehydrogenase subunit A FpoA
AR505_1634 NADP oxidoreductase coenzyme F_{420} dependent
AR505_0362 methanogenesis marker protein 13
AR505_0724 methanogenesis marker protein 16
AR505_0786 methanogenesis marker protein 2
AR505_1203 methanogenesis marker protein 8
AR505_1385 methanogenesis marker protein 7
AR505_1386 methanogenesis marker protein 17
AR505_1387 methanogenesis marker protein 15
AR505_1388 methanogenesis marker protein 5
AR505_1389 methanogenesis marker protein 6
AR505_1390 methanogenesis marker protein 3
AR505_1405 methanogenesis marker protein 1
AR505_1417 methanogenesis marker protein 4
AR505_1637 methanogenesis marker protein 11

LIPID METABOLISM

Biosynthesis general

AR505_0892 geranylgeranyl reductase family protein
AR505_0928 acyl carrier protein phosphodiesterase AcpD
AR505_1433 geranylgeranyl reductase
AR505_1618 geranylgeranyl reductase
AR505_1646 geranylgeranyl reductase family protein

Lipid backbone

AR505_0626 NAD(P)-dependent glycerol-1-phosphate dehydrogenase

Phospholipid biosynthesis
AR505_0918 CDP-alcohol phosphatidyltransferase
AR505_0933 cardiolipin synthase
AR505_1147 phospholipase D/transphosphatidylase PlD
AR505_1433 digeranylgeranylglycerophospholipid reductase DGGGPR
AR505_1587 digeranylgeranylglyceryl phosphate synthase DGGGPS
AR505_1588 geranylgeranylglyceryl phosphate synthase GGGPS
AR505_1616 CDP-diglyceride synthetase CdsA
AR505_1783 CDP-diacylglycerol—glycerol-3-phosphate 3-phosphatidyltransferase
Mevalonate pathway
AR505_0004 acetyl-CoA acetyltransferase
AR505_0601 hydroxymethylglutaryl-CoA synthase
AR505_0602 acetyl-CoA acetyltransferase
AR505_0768 hydroxymethylglutaryl-CoA reductase (NADPH) HmgA
AR505_1431 mevalonate kinase Mvk
Elongation of isoprenoid side chains
AR505_0191 isopentenyl diphosphate delta-isomerase Fni
AR505_0192 isopentenyl phosphate kinase Ippk
AR505_1587 prenyltransferase UbiA
AR505_0190 bifunctional short chain isoprenyl diphosphate synthase IdsA/GGPS
AR505_1619 octaprenyl-diphosphate synthase IspB

MOBILE ELEMENTS

CRISPR-associated genes
AR505_1089 CRISPR-associated endonuclease Cas3-HD
AR505_1090 CRISPR type I-E/ECOLI-associated protein CasA/Cse1
AR505_1091 CRISPR type I-E/ECOLI-associated protein CasB/Cse2 complex
AR505_1092 CRISPR-associated protein Cas7/Cse4/CasC
AR505_1093 CRISPR-associated protein Cas5/CasD
AR505_1094 CRISPR-associated protein Cas6/Cse3/CasE
AR505_1095 CRISPR-associated endonuclease Cas1
Toxin/antitoxin
AR505_0857 prevent-host-death family protein
AR505_0858 addiction module toxin, RelE/StbE family
AR505_0980 toxin-antitoxin system, toxin component, HipA family
AR505_1466 addiction module antitoxin RelB/DinJ
AR505_1468 addiction module antitoxin RelB/DinJ
AR505_1566 death-on-curing family protein hypothetical protein
Transposase
AR505_0010 transposase
AR505_0106 transposase
AR505_0107 transposase
AR505_0147 transposase IS4 family
AR505_0151 transposase IS4 family
AR505_0213 transposase IS116/IS110/IS902 family
AR505_0215 transposase IS605 OrfB family
AR505_0276 transposase IS605 OrfB family
AR505_0290 transposase IS605 OrfB family
AR505_0371 transposase IS605 OrfB family
AR505_0372 transposase
AR505_0440 transposase IS4 family
AR505_0498 transposase IS605 OrfB family
AR505_0516 transposase IS605 OrfB family
AR505_0523 transposase IS605 OrfB family
AR505_0525 transposase
AR505_0528 transposase
AR505_0574 transposase
AR505_0580 transposase IS605 OrfB family
AR505_0639 transposase IS605 OrfB family
AR505_0677 transposase IS605 OrfB family
AR505_0708 transposase
AR505_0784 transposase IS4 family
AR505_0860 transposase IS605 OrfB family
AR505_0903 transposase IS605 OrfB family
AR505_0913 transposase IS605 OrfB family
AR505_1219 transposase IS605 OrfB family
AR505_1239 transposase
AR505_1356 transposase IS4 family
AR505_1423 transposase
AR505_1458 transposase IS4 family protein
AR505_1562 transposase
AR505_1563 transposase
AR505_1574 transposase
AR505_1692 transposase IS605 OrfB family
AR505_1703 transposase
AR505_1714 transposase IS4 family
AR505_1765 transposase IS605 OrfB family
AR505_1775 transposase IS605 OrfB family
AR505_1777 transposase IS605 OrfB family
Integrase
AR505_0313 phage integrase
AR505_0669 integrase catalytic subunit
AR505_0931 phage integrase
AR505_1543 phage integrase
AR505_1640 phage integrase
AR505_1697 phage integrase family protein
AR505_1772 group II intron maturase
AR505_1518 prophage Lp3 helicase
AR505_1570 phage integrase

NITROGEN METABOLISM

Potential Fixation
AR505_0186 nitroreductase family protein
AR505_0207 nitrilase/cyanide hydratase and apolipoprotein *N*-acyltransferase
AR505_0359 nitrogenase iron protein NifH
AR505_0591 nitroreductase family protein
AR505_0822 nitroreductase family protein
AR505_1268 dinitrogenase iron-molybdenum cofactor biosynthesis protein
AR505_1288 oxidoreductase/nitrogenase component 1
AR505_1289 nitrogenase
AR505_1502 oxidoreductase/nitrogenase component 1
Other
AR505_0956 hydroxylamine reductase Hcp
AR505_1059 FeS cluster assembly scaffold protein NifU
AR505_1060 cysteine desulfurase NifS
AR505_1447 nitrogenase cofactor biosynthesis protein NifB
AR505_1501 nitrogenase iron protein NifH
Regulation
AR505_0206 nitrogen regulatory protein P-II
AR505_0402 nitrogen regulatory protein P-II
Transport
AR505_0205 ammonium transporter Amt

NUCLEIC ACID METABOLISM

DNA-binding proteins
AR505_0051 histone acetyltransferase ELP3 family
AR505_0603 DNA-binding protein
AR505_0732 NAD-dependent deacetylase SIR2 family
AR505_0777 DNA-binding protein
AR505_1236 archaeal histone
AR505_1617 DNA/RNA-binding protein
AR505_1668 OB fold nucleic acid binding domain-containing protein
Helicase
AR505_0025 ATP-dependent DNA helicase
AR505_0039 ATP-dependent DNA helicase
AR505_0583 helicase SNF2 family
AR505_0613 helicase, superfamily II
AR505_0805 ATP-dependent DNA helicase
AR505_0814 DNA repair helicase Rad3
AR505_1115 DEAD/DEAH box helicase domain-containing protein
AR505_1117 ATP-dependent RNA helicase
AR505_1132 ATP-dependent DNA helicase UvrD/REP family
AR505_1251 DNA helicase
AR505_1267 helicase
AR505_1278 DEAD/DEAH box helicase domain-containing protein
AR505_1321 ATP-dependent RNA helicase
AR505_1337 DEAD/DEAH box helicase domain-containing protein
AR505_1369 ATP-dependent DNA helicase
AR505_1500 DNA helicase
AR505_1531 ATP-dependent DNA helicase
AR505_1662 DEAD/DEAH box helicase domain-containing protein
Recombination and repair
AR505_0027 8-oxoguanine DNA-glycosylase Ogg
AR505_0028 endonuclease IV
AR505_0117 ssDNA exonuclease RecJ
AR505_0125 replication-associated recombination protein A

AR505_0178 endonuclease IV
AR505_0187 RdgB/HAM1 family non-canonical purine NTP pyrophosphatase
AR505_0189 DNA repair and recombination protein RadB
AR505_0329 DNA mismatch endonuclease Vsr
AR505_0391 formamidopyrimidine-DNA glycosylase MutM
AR505_0420 ribonuclease H
AR505_0448 6-*O*-methylguanine DNA methyltransferase Ogt
AR505_0693 endoribonuclease L-PSP
AR505_0798 ssDNA exonuclease RecJ
AR505_1055 site-specific recombinase
AR505_1056 site-specific recombinase
AR505_1069 DNA repair photolyase
AR505_1084 exodeoxyribonuclease VII small subunit XseB
AR505_1085 exodeoxyribonuclease VII large subunit XseA
AR505_1139 exonuclease
AR505_1154 archaeal Holliday junction resolvase Hjc
AR505_1298 GTP:adenosylcobinamide-phosphate guanylyltransferase CobU
AR505_1305 endonuclease III Nth
AR505_1309 DNA alkylation repair enzyme
AR505_1402 ssDNA exonuclease RecJ
AR505_1578 excinuclease ABC A subunit UvrA
AR505_1585 DNA repair and recombination protein RadA
AR505_1614 uracil-DNA glycosylase Ung
AR505_1729 resolvase domain-containing protein

Regulation

AR505_0907 peptidase U62 modulator of DNA gyrase
AR505_0908 peptidase U62 modulator of DNA gyrase

Restriction and modification

AR505_0327 DNA-cytosine methyltransferase Dcm
AR505_0339 DNA-cytosine methyltransferase Dcm
AR505_0340 DNA-cytosine methyltransferase Dcm
AR505_0662 GTPase subunit of restriction endonuclease
AR505_0663 McrBC 5-methylcytosine restriction system component
AR505_0958 viral recombinase YgaJ family

PROTEIN FATE

Protein degradation

AR505_0013 peptidase M48 family (4 TMHs)
AR505_0115 universal archaeal protein Kae1
AR505_0122 metal-dependent protease
AR505_0144 peptidase M24 family
AR505_0164 hypothetical protein
AR505_0188 Kae1-associated kinase Bud32
AR505_0264 peptidase M50 family (6 TMHs)
AR505_0270 peptidase M16 family
AR505_0449 aminoacyl-histidine dipeptidase PepD
AR505_0615 ATP-dependent protease S16 family
AR505_0688 amidase
AR505_0702 cysteine proteinase
AR505_0785 peptidase M50 family (6 TMHs)
AR505_0919 proline-specific peptidases
AR505_0967 proteasome endopeptidase complex
AR505_1001 peptidase U32 family
AR505_1143 ATP-dependent protease
AR505_1318 peptidase M18 family
AR505_1443 methionine aminopeptidase Map
AR505_1674 proteasome-activating nucleotidase

Protein folding

AR505_0007 prefoldin alpha subunit PfdA
AR505_0092 thermosome subunit
AR505_0093 molecular chaperone GrpE
AR505_0094 chaperone protein DnaK
AR505_0095 chaperone protein DnaJ
AR505_0286 heat shock protein Hsp90
AR505_0836 peptidyl-prolyl cis-trans isomerase
AR505_1062 thermosome subunit
AR505_1228 chaperone protein DnaK
AR505_1232 chaperone protein DnaK
AR505_1248 ATP-dependent chaperone protein ClpB
AR505_1482 proteasome alpha subunit PsmA
AR505_1490 prefoldin beta subunit PfdB

Protein secretion

AR505_0006 signal recognition particle receptor FtsY
AR505_0241 preprotein translocase subunit SecY
AR505_0667 signal peptidase I
AR505_0758 signal recognition particle SRP54 protein
AR505_0962 Type II/IV secretion system protein E
AR505_0986 signal peptidase I
AR505_1473 signal recognition particle SRP19 protein
AR505_1612 signal peptidase I
AR505_1828 signal peptidase I
AR505_1799 protein translocase subunit Sss1

PROTEIN SYNTHESIS

Other

AR505_0131 translation-associated GTPase
AR505_1444 Sua5/YciO/YrdC/YwlC family translation factor
AR505_1683 tRNA methyltransferase subunit

Ribosomal proteins

AR505_0008 ribosomal protein LX
AR505_0057 ribosomal protein L30e Rpl30e
AR505_0077 ribosomal protein S17e Rps17e
AR505_0118 ribosomal protein S15P Rps15p
AR505_0129 ribosomal protein L15e Rpl15e
AR505_0135 ribosomal protein L21e Rpl21e
AR505_0179 ribosomal protein L18e Rpl18e
AR505_0180 ribosomal protein L13P Rpl13p
AR505_0181 ribosomal protein S9P Rps9p
AR505_0218 ribosomal protein L3P Rpl3p
AR505_0219 ribosomal protein L4P Rpl4p
AR505_0220 ribosomal protein L23P Rpl23p
AR505_0221 ribosomal protein L2P Rpl2p
AR505_0222 ribosomal protein S19P Rps19p
AR505_0223 ribosomal protein L22P Rpl22p
AR505_0224 ribosomal protein S3P Rps3p
AR505_0225 ribosomal protein L29P Rpl29p
AR505_0228 ribosomal protein S17P Rps17p
AR505_0229 ribosomal protein L14e Rpl14e
AR505_0230 ribosomal protein L24P Rpl24p
AR505_0231 ribosomal protein S4e Rps4e
AR505_0232 ribosomal protein L5P Rpl5p
AR505_0233 ribosomal protein S8P Rps8p
AR505_0234 ribosomal protein L6P Rpl6p
AR505_0235 ribosomal protein L32e Rpl32e
AR505_0236 ribosomal protein L19e Rpl19e
AR505_0237 ribosomal protein L18P Rpl18p
AR505_0238 ribosomal protein S5P Rps5p
AR505_0239 ribosomal protein L30P Rpl30p
AR505_0240 ribosomal protein L15P Rpl15p
AR505_0277 ribosomal protein S3Ae Rps3ae
AR505_0283 ribosomal protein L37e Rpl37e
AR505_0468 ribosomal protein L40e Rpl40e
AR505_0776 ribosomal protein S19e Rps19e
AR505_0778 ribosomal protein L39e Rpl39e
AR505_0779 ribosomal protein L31e Rpl31e
AR505_1370 ribosomal protein S27e Rps27e
AR505_1371 ribosomal protein S24e Rps24e
AR505_1421 ribosomal protein L10e Rpl10e
AR505_1440 ribosomal protein S11P Rps11p
AR505_1441 ribosomal protein S4P Rps4p
AR505_1442 ribosomal protein S13P Rps13p
AR505_1453 ribosomal protein S10P Rps10p
AR505_1456 ribosomal protein S7P Rps7p
AR505_1457 ribosomal protein S12P Rps12p
AR505_1472 ribosomal protein S8e Rps8e
AR505_1483 ribosome maturation protein SBDS
AR505_1487 ribosomal protein L37Ae Rpl37ae
AR505_1514 ribosomal protein L44e Rpl44e
AR505_1515 ribosomal protein S27e Rps27e
AR505_1671 ribosomal protein S2P Rps2p
AR505_1753 ribosomal protein S6e Rps6e
AR505_1757 ribosomal protein L24e Rpl24e
AR505_1758 ribosomal protein S28e Rps28e
AR505_1759 ribosomal protein L7Ae Rpl7ae
AR505_1794 ribosomal protein L12P Rpl12p
AR505_1795 acidic ribosomal protein P0 RplPO
AR505_1796 ribosomal protein L1P Rpl1p
AR505_1797 ribosomal protein L11P Rpl11p
AR505_1798 ribosomal protein L24 family

RNA processing

AR505_0827 wyosine biosynthesis protein TYW1
AR505_1601 RNA 3'-phosphate cyclase RtcA
AR505_0020 NMD3 family protein
AR505_0038 ribosomal RNA large subunit methyltransferase J RrmJ
AR505_0065 tRNA pseudouridine synthase A TruA
AR505_0120 fibrillarin
AR505_0132 dimethyladenosine transferase KsgA
AR505_0142 archease
AR505_0227 ribonuclease P subunit P29
AR505_0243 tRNA pseudouridine synthase B TruB
AR505_0438 archaeosine tRNA-ribosyltransferase TgtA
AR505_0464 ribonuclease Z Rnz
AR505_0637 archaeosine tRNA-ribosyltransferase TgtA
AR505_0809 tRNA-dihydrouridine synthase DusA
AR505_1021 2'-5' RNA ligase
AR505_1146 ribosomal-protein-alanine acetyltransferase RimI
AR505_1166 N^2,N^2-dimethylguanosine tRNA methyltransferase Trm1
AR505_1233 pseudouridylate synthase
AR505_1274 tRNA pseudouridine synthase D TruD
AR505_1336 YbaK/EbsC family protein
AR505_1358 exosome subunit
AR505_1406 RNA-splicing ligase RtcB
AR505_1414 7-cyano-7-deazaguanosine biosynthesis protein QueE
AR505_1415 queuosine biosynthesis protein QueC
AR505_1416 6-pyruvoyl tetrahydropterin synthase QueD
AR505_1424 ribonuclease P subunit RPR2
AR505_1460 MiaB-like tRNA modifying enzyme
AR505_1484 exosome complex RNA-binding protein Rrp4
AR505_1485 exosome complex exonuclease Rrp41
AR505_1486 exosome complex RNA-binding protein Rrp42
AR505_1583 tRNA nucleotidyltransferase Cca
AR505_1586 N^2,N^2-dimethylguanosine tRNA methyltransferase Trm
AR505_1638 RNA 2'-phosphotransferase Tpt1/KptA
AR505_1657 ribosomal protein L11 methyltransferase PrmA
AR505_1658 exosome complex RNA-binding protein Csl4
AR505_1791 RNA methyltransferase TrmH family
AR505_1808 tRNA intron endonuclease EndA

Translation factors

AR505_0139 translation initiation factor aIF-1A
AR505_0226 translation initiation factor aSUI1
AR505_0252 peptide chain release factor aRF1
AR505_0506 cell division protein pelota PelA
AR505_0642 peptidyl-tRNA hydrolase
AR505_0780 translation initiation factor aIF-6
AR505_1011 elongation factor Tu domain 2 protein
AR505_1284 diphthine synthase DphB
AR505_1420 diphthine synthase DphB
AR505_1454 translation elongation factor aEF-1 alpha
AR505_1455 translation elongation factor aEF-2
AR505_1459 translation initiation factor aIF-2 alpha subunit
AR505_1516 translation initiation factor aIF-2 alpha subunit
AR505_1603 translation initiation factor aIF-2 beta subunit
AR505_1606 translation initiation factor aIF-5A
AR505_1752 translation initiation factor aIF-2 gamma subunit
AR505_1755 translation initiation factor aIF-2
AR505_1774 translation elongation factor aEF-1 beta

tRNA aminoacylation

AR505_0068 glutamyl-$tRNA^{Gln}$ amidotransferase subunit D GatD
AR505_0069 glutamyl-$tRNA^{Gln}$ amidotransferase subunit E GatE
AR505_0198 alanyl-tRNA synthetase AlaS
AR505_0200 aspartate--tRNA synthetase AspS
AR505_0201 Asp-$tRNA^{Asn}$/Glu-$tRNA^{Gln}$ amidotransferase subunit B GatB
AR505_0202 Asp-$tRNA^{Asn}$/Glu-$tRNA^{Gln}$ amidotransferase subunit A GatA
AR505_0203 Asp-$tRNA^{Asn}$/Glu-$tRNA^{Gln}$ amidotransferase subunit C GatC
AR505_0247 tryptophanyl-tRNA synthetase TrpS
AR505_0248 phenylalanyl-tRNA synthetase alpha subunit PheS
AR505_0253 arginine-tRNA synthetase ArgS
AR505_0387 lysyl-tRNA synthetase LysS
AR505_0394 histidyl-tRNA synthetase HisS
AR505_0397 threonyl-tRNA synthetase ThrS
AR505_0455 isoleucyl-tRNA synthetase IleS
AR505_0709 methionyl-tRNA synthetase MetG
AR505_0710 methionyl-tRNA synthetase beta subunit MetG
AR505_0767 phenylalanyl-tRNA synthetase subunit beta PheT
AR505_0972 leucyl-tRNA synthetase LeuS
AR505_1121 glutamyl-tRNA synthetase GltX
AR505_1150 aspartyl-tRNA synthetase AspS
AR505_1191 cysteinyl-tRNA synthetase CysS
AR505_1325 pyrrolysine-tRNA ligase PylS
AR505_1339 valyl-tRNA synthetase ValS
AR505_1341 glycyl-tRNA synthetase GlyS
AR505_1426 seryl-tRNA synthetase SerS
AR505_1448 alanyl-tRNA synthetase AlaS
AR505_1494 tyrosyl-tRNA synthetase TyrS
AR505_1684 prolyl-tRNA synthetase ProS

PURINES AND PYRIMIDINES

Purine biosynthesis

AR505_0037 phosphoribosylformylglycinamidine cyclo-ligase PurM
AR505_0066 phosphoribosylamine--glycine ligase PurD
AR505_0084 adenine phosphoribosyltransferase Apt
AR505_0141 phosphoribosylglycinamide formyltransferase PurN
AR505_0257 5-formaminoimidazole-4-carboxamide-1-β-D- ribofuranosyl 5'-monophosphate synthetase-like protein
AR505_0284 amidophosphoribosyltransferase PurF
AR505_0301 adenine phosphoribosyltransferase Apt
AR505_0622 phosphoribosylaminoimidazole- succinocarboxamide synthase PurC
AR505_0705 AMP phosphorylase
AR505_0742 adenylosuccinate lyase PurB
AR505_1071 exopolyphosphatase Ppx
AR505_1496 adenine phosphoribosyltransferase Apt
AR505_1595 phosphoribosylaminoimidazole carboxylase PurE
AR505_1656 bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase PurH
AR505_1667 5-formaminoimidazole-4-carboxamide-1-(beta)-D-ribofuranosyl 5'-monophosphate-formate ligase PurP
AR505_1680 phosphoribosylformylglycinamidine synthase I PurQ
AR505_1681 phosphoribosylformylglycinamidine synthase II PurL
AR505_1682 phosphoribosylformylglycinamidine synthase PurS
AR505_1804 phosphoribosylaminoimidazole carboxylase PurE

Purine interconversion

AR505_0262 adenylate kinase Adk
AR505_1168 adenylosuccinate synthase PurA
AR505_1784 adenylate kinase Adk

Pyrimidine biosynthesis

AR505_0060 orotidine 5'-phosphate decarboxylase PyrF
AR505_0281 dihydroorotase PyrC
AR505_0436 aspartate carbamoyltransferase regulatory subunit PyrI
AR505_0437 aspartate carbamoyltransferase PyrB
AR505_0500 carbamoyl-phosphate synthase large subunit CarB
AR505_0501 carbamoyl-phosphate synthase small subunit CarA
AR505_0503 carbamoyl-phosphate synthase large subunit CarB
AR505_0504 carbamoyl-phosphate synthase small subunit CarA
AR505_1029 dihydroorotate dehydrogenase PyrD
AR505_1030 dihydroorotate dehydrogenase electron transfer subunit PyrK
AR505_1581 cytosine deaminase
AR505_1615 orotate phosphoribosyltransferase PyrE

Pyrimidine interconversion

AR505_1827 CMP/dCMP deaminase CMP/dCMP deaminase
AR505_0076 thymidylate synthase ThyX
AR505_0251 uridylate kinase PyrH
AR505_1102 thymidylate kinase Tmk
AR505_1287 deoxycytidine triphosphate deaminase Dcd
AR505_1346 thioredoxin-disulfide reductase TrxB
AR505_1690 thymidylate kinase Tmk

Interconversion

AR505_0706 GMP synthase subunit A GuaA
AR505_1165 NTPase
AR505_1247 anaerobic ribonucleoside-triphosphate reductase NrdD
AR505_1743 5'-nucleotidase SurE
AR505_1756 nucleoside diphosphate kinase Ndk
AR505_1802 GMP synthase subunit A GuaA
AR505_1803 GMP synthase subunit B GuaAb
AR505_1827 CMP/dCMP deaminase

Salvage

AR505_1636 undecaprenyl pyrophosphate synthetase UppS
AR505_1344 cytosine deaminase
AR505_1382 *N*-methylhydantoinase A/acetone carboxylase beta subunit

AR505_1641 thymidine phosphorylase

Salvage and carbon fixation

AR505_1642 ribulose bisphosphate carboxylase RbcL
AR505_1643 ribose-1,5-bisphosphate isomerase E2b2

Transport

AR505_0995 xanthine/uracil permease family protein

REGULATION

Protein interaction

AR505_0138 serine/threonine protein kinase RIO1 family
AR505_0204 TPR repeat-containing protein
AR505_0259 phosphate uptake regulator PhoU
AR505_0260 phosphate uptake regulator PhoU
AR505_0275 TPR repeat-containing protein
AR505_0366 TPR repeat-containing protein
AR505_0367 TPR repeat-containing protein
AR505_0368 TPR repeat-containing protein
AR505_0370 TPR repeat-containing protein
AR505_0373 TPR repeat-containing protein
AR505_0419 Sel1 repeat-containing protein
AR505_0536 TPR repeat-containing protein
AR505_0557 TPR repeat-containing protein
AR505_0641 TPR repeat-containing protein
AR505_0793 TPR repeat-containing protein
AR505_0795 TPR repeat-containing protein
AR505_0824 HPr kinase
AR505_0898 TPR repeat-containing protein
AR505_0899 TPR repeat-containing protein
AR505_0917 TPR repeat-containing protein
AR505_1020 TPR repeat-containing protein
AR505_1048 TPR repeat-containing protein
AR505_1145 TPR repeat-containing protein
AR505_1212 TPR repeat-containing protein
AR505_1213 TPR repeat-containing protein
AR505_1214 TPR repeat-containing protein
AR505_1216 TPR repeat-containing protein
AR505_1225 TPR repeat-containing protein
AR505_1273 TPR repeat-containing protein
AR505_1334 TPR repeat-containing protein
AR505_1351 TPR repeat-containing protein
AR505_1463 TPR repeat-containing protein

Signal transduction

AR505_0425 signal transduction protein with CBS domains
AR505_0609 low molecular weight phosphotyrosine protein phosphatase
AR505_0624 protein-tyrosine phosphatase
AR505_0788 signal transduction histidine kinase
AR505_1169 signal transduction histidine kinase
AR505_1170 signal transduction histidine kinase

Transcriptional regulator

AR505_0043 transcriptional regulator ArsR family
AR505_0049 XRE family transcriptional regulator
AR505_0162 transcriptional regulator AsnC family
AR505_0199 transcriptional regulator AsnC family
AR505_0272 transcriptional regulator LysR family
AR505_0383 nickel responsive transcriptional regulator NikR
AR505_0457 transcriptional regulator LysR family
AR505_0462 transcriptional regulator ArsR family
AR505_0465 HTH/CBS domain-containing protein
AR505_0532 HTH domain-containing protein
AR505_0599 transcriptional regulator
AR505_0600 HTH domain-containing protein
AR505_0718 TfoX N-terminal domain protein
AR505_0734 transcriptional regulator HxlR family
AR505_0774 transcriptional regulator MarR family
AR505_0790 transcriptional regulator LytR family
AR505_0791 transcriptional regulator LytS family
AR505_0797 transcriptional regulator HxlR family
AR505_0801 HTH domain-containing protein
AR505_0831 transcriptional regulator MarR family
AR505_0843 transcriptional regulator
AR505_0859 transcriptional regulator TetR family
AR505_0955 transcriptional regulator HxlR family
AR505_0979 HTH domain-containing protein
AR505_0982 transcriptional regulator ArsR family
AR505_1012 transcriptional regulator XRE family
AR505_1099 transcriptional regulator MarR family
AR505_1176 transcriptional regulator TetR family
AR505_1179 metal dependent transcriptional regulator
AR505_1226 transcriptional regulator ArsR family
AR505_1237 transcriptional regulator HxlR family
AR505_1302 nickel-responsive transcriptional regulator NikR
AR505_1314 HTH domain-containing protein
AR505_1673 HTH domain-containing protein
AR505_1679 transcriptional regulator MarR family
AR505_1809 transcriptional regulator ArsR family
AR505_1811 transcriptional regulator

Other

AR505_0645 carbon starvation protein CstA
AR505_0984 sugar fermentation stimulation protein SfsA

TRANSCRIPTION

RNA polymerase

AR505_0053 DNA-directed RNA polymerase subunit H RpoH
AR505_0054 DNA-directed RNA polymerase subunit B RpoB
AR505_0055 DNA-directed RNA polymerase subunit A' RpoA1
AR505_0056 DNA-directed RNA polymerase subunit A" RpoA2
AR505_0182 DNA-directed RNA polymerase subunit N RpoN
AR505_0922 DNA-directed RNA polymerase subunit K RpoK
AR505_1373 DNA-directed RNA polymerase subunit E RpoE
AR505_1374 DNA-directed RNA polymerase RpoE
AR505_1439 DNA-directed RNA polymerase subunit D RpoD
AR505_1488 DNA-directed RNA polymerase subunit P RpoP

Translation factors

AR505_1807 transcription initiation factor TFIIB Tfb2
AR505_0810 transcription initiation factor TFIIB Tfb
AR505_1651 transcription factor S Tfs

Other

AR505_1605 C/D box methylation guide ribonucleoprotein complex aNOP56 subunit
AR505_1601 RNA 3'-phosphate cyclase RtcA
AR505_0598 polyphosphate:AMP phosphotransferase
AR505_0578 ribonuclease III Rnc
AR505_0137 RNA-binding protein
AR505_1408 RNA-binding protein
AR505_1411 RNA-binding protein

TRANSPORTERS

Amino acids

AR505_0830 amino acid/peptide transporter
AR505_0944 amino acid carrier protein AgcS
AR505_1242 amino acid/peptide transporter

Cations

AR505_0507 transporter Na^+/H^+ antiporter family
AR505_0258 divalent cation transporter mgtE family
AR505_0369 sodium bile acid symporter family
AR505_0385 transporter Na^+/H^+ antiporter family
AR505_0696 transporter CDF family
AR505_0812 Na/Pi-cotransporter
AR505_0835 transporter Na^+/H^+ antiporter family
AR505_1074 potassium uptake protein TrkA family
AR505_1075 potassium uptake protein TrkH family
AR505_1133 cation-transporting P-type ATPase heavy metal-translocating P-type ATPase
AR505_1510 copper ion binding protein
AR505_1511 copper translocating P-type ATPase
AR505_1222 ferrous iron transport protein B FeoB
AR505_1223 ferrous iron transport protein A FeoA
AR505_1506 iron chelate uptake ABC transporter permease
AR505_1507 iron chelate uptake ABC transporter permease inner membrane subunit
AR505_0620 sodium:solute symporter
AR505_1100 K^+-dependent Na^+/Ca^+ exchanger
AR505_0287 sodium/proline symporter PutP
AR505_1227 heavy metal translocating P-type ATPase
AR505_1271 heavy metal translocating P-type ATPase

Other

AR505_0030 ABC transporter permease protein
AR505_0031 ABC transporter ATP-binding protein
AR505_0111 ABC transporter ATP-binding protein

AR505_0112 ABC transporter substrate-binding protein
AR505_0113 ABC transporter permease protein
AR505_0127 permease
AR505_0163 mechanosensitive ion channel protein
AR505_0193 drug resistance MFS transporter
AR505_0285 Na^+-dependent transporter SNF family
AR505_0296 MatE efflux family protein
AR505_0303 ABC transporter solute binding protein
AR505_0304 ABC transporter ATP binding protein
AR505_0305 ABC transporter permease protein
AR505_0308 ABC transporter ATP-binding protein
AR505_0309 ABC transporter permease protein
AR505_0310 ABC transporter solute-binding protein
AR505_0311 transporter DMT family
AR505_0364 MFS transporter
AR505_0408 ABC transporter permease protein
AR505_0409 ABC transporter permease protein
AR505_0410 ABC transporter ATP-binding protein
AR505_0411 ABC transporter permease protein
AR505_0412 ABC transporter ATP-binding protein
AR505_0417 ABC transporter ATP-binding protein
AR505_0418 ABC transporter permease protein
AR505_0422 ABC transporter ATP-binding protein
AR505_0423 ABC transporter permease protein
AR505_0456 transporter DMT family
AR505_0476 ABC transporter solute-binding protein
AR505_0477 ABC transporter permease protein
AR505_0478 ABC transporter ATP-binding protein
AR505_0517 ABC transporter ATP-binding protein
AR505_0520 ABC transporter ATP-binding protein
AR505_0521 ABC transporter permease protein
AR505_0526 transporter RND family
AR505_0605 ABC transporter permease protein
AR505_0606 ABC transporter ATP-binding protein
AR505_0701 ABC transporter ATP-binding protein
AR505_0703 MATE efflux family protein
AR505_0712 ABC transporter ATP-binding protein
AR505_0713 ABC transporter permease protein
AR505_0714 ABC transporter permease protein
AR505_0715 ABC transporter solute-binding protein
AR505_0737 Na^+dependent transporter SBF family
AR505_0739 MFS transporter
AR505_0746 mechanosensitive ion channel protein
AR505_0806 MFS transporter
AR505_0815 Na^+dependent transporter SBF family
AR505_0905 EamA-like transporter family
AR505_0911 MFS transporter
AR505_0941 MFS transporter
AR505_0957 HPP family protein
AR505_1014 ABC transporter ATP-binding protein
AR505_1015 ABC transporter permease protein
AR505_1034 MatE efflux family protein
AR505_1052 MFS transporter
AR505_1065 MFS transporter
AR505_1138 transporter SDF family
AR505_1195 ABC transporter substrate-binding protein
AR505_1197 ABC transporter permease protein
AR505_1198 ABC transporter ATP-binding protein
AR505_1246 Na^+-driven multidrug efflux pump
AR505_1272 transporter MIP family
AR505_1291 ABC transporter ATP-binding protein
AR505_1292 ABC transporter permease protein
AR505_1293 ABC transporter substrate-binding protein
AR505_1310 small multidrug resistance protein
AR505_1311 small multidrug resistance protein
AR505_1378 small multidrug resistance protein
AR505_1379 small multidrug resistance protein
AR505_1462 MFS transporter
AR505_1491 ABC transporter ATP-binding protein
AR505_1492 ABC transporter permease protein
AR505_1498 MFS transporter
AR505_1504 ABC transporter permease protein
AR505_1505 ABC transporter ATP-binding protein
AR505_1508 ABC transporter ATP-binding protein
AR505_1523 ABC transporter substrate-binding protein
AR505_1525 ABC transporter permease protein
AR505_1526 ABC transporter ATP-binding protein
AR505_1548 ABC transporter substrate-binding protein
AR505_1550 ABC transporter permease protein
AR505_1551 ABC transporter ATP-binding protein
AR505_1576 MFS transporter
AR505_1686 sulfate permease SulP
AR505_1705 ABC transporter ATP-binding protein
AR505_1706 ABC transporter permease protein
AR505_1708 ABC transporter substrate-binding protein
AR505_1076 drug resistance transporter Bcr/CflA subfamily
AR505_1081 small multidrug resistance protein
AR505_1082 small multidrug resistance protein
AR505_1101 ABC transporter ATP-binding protein

UNKNOWN FUNCTION

Enzyme

AR505_0011 hydrolase, TatD family
AR505_0099 methyltransferase
AR505_0254 acetyltransferase
AR505_0426 ATPase (AAA+ superfamily)
AR505_0447 oxidoreductase DSBA family
AR505_0459 DNAse TatD family
AR505_0505 ATPase (AAA+ superfamily)
AR505_0522 ATPase (AAA+ superfamily)
AR505_0530 ATPase (AAA+ superfamily)
AR505_0537 SAM-dependent methyltransferase
AR505_0587 ATPase (AAA+ superfamily)
AR505_0717 ATPase (AAA+ superfamily)
AR505_0781 Zn-dependent hydrolase
AR505_0863 ATPase
AR505_0893 ATPase
AR505_0895 ATPase
AR505_0900 ATPase
AR505_0904 GNAT family acetyltransferases
AR505_0943 GNAT family acetyltransferase
AR505_0963 hydroase HD superfamily
AR505_0970 oxidoreductase
AR505_0973 SAM-dependent methyltransferase
AR505_0981 SAM-dependent methyltransferases
AR505_0983 ATPase
AR505_0996 acetyltransferase
AR505_1000 short-chain dehydrogenase/reductase SDR
AR505_1008 SAM-dependent methyltransferases
AR505_1079 ATPase
AR505_1194 SAM-dependent methyltransferase
AR505_1196 SAM-dependent methyltransferase
AR505_1244 GNAT family acetyltransferases
AR505_1285 SAM-dependent methyltransferase
AR505_1367 archaeal ATPase
AR505_1381 haloacid dehalogenase-like hydrolase
AR505_1403 phosphoesterase
AR505_1422 haloacid dehalogenase-like hydrolase
AR505_1427 haloacid dehalogenase-like hydrolase
AR505_1471 phosphodiesterase
AR505_1474 carbohydrate kinase YjeF family
AR505_1475 phosphoesterase
AR505_1493 phosphoesterase DHHA1
AR505_1557 Atpase
AR505_1558 nucleotidyltransferase
AR505_1593 ATPase
AR505_1613 haloacid dehalogenase-like hydrolase
AR505_1695 Metal-dependent hydrolases beta-lactamase family
AR505_1738 ATPase
AR505_1744 GNAT family acetyltransferases
AR505_1781 Fe-S oxidoreductase
AR505_1814 sugar phosphate isomerase/epimerase
AR505_0685 peptidyl-prolyl cis-trans isomerase FKBP-type
AR505_1028 peptidyl-prolyl cis-trans isomerase
AR505_1235 peptidyl-prolyl cis-trans isomerase
AR505_0070 methyltransferase
AR505_0307 SAM-dependent methyltransferase
AR505_0844 GNAT family acetyltransferase
AR505_1218 haloacid dehalogenase-like hydrolase
AR505_1437 GTPase

AR505_1653 RNA-metabolising metallo-beta-lactamase
AR505_1676 GTPase

Other

AR505_1352 alpha/beta hydrolase fold protein
AR505_0968 universal archaeal KH-domain/beta-lactamase-domain protein
AR505_1691 TOPRIM domain-containing protein
AR505_0012 beta-lactamase domain-containing protein
AR505_0071 metallo-beta-lactamase domain-containing protein
AR505_0098 pyridoxamine 5'-phosphate oxidase-related protein
AR505_0119 NUDIX domain-containing protein
AR505_0261 crcB protein CrcB
AR505_0298 Cob/MinD domain containing protein
AR505_0546 amidohydrolase family protein
AR505_0636 pap2 family protein
AR505_0699 PHP domain-containing protein
AR505_0711 PHP domain-containing protein
AR505_0721 CAAX amino terminal protease family
AR505_0722 von Willebrand factor type A domain protein
AR505_0733 macro domain protein
AR505_0890 SMC domain-containing protein
AR505_0891 SMC domain-containing protein
AR505_0924 pyridoxamine 5'-phosphate oxidase family protein
AR505_0926 NfeD family protein
AR505_0935 PHP domain-containing protein
AR505_1204 PP-loop domain-containing protein
AR505_1234 Pirin family protein
AR505_1257 NYN domain-containing protein
AR505_1340 CBS domain-containing protein
AR505_1342 CBS domain-containing protein
AR505_1407 Metal-dependent phosphohydrolase HD superfamily
AR505_1499 thioesterase family protein
AR505_1544 virulence-associated E family protein
AR505_1582 hydrolase beta-lactamase family
AR505_1621 radical SAM domain protein
AR505_1675 radical SAM domain-containing proteinB
AR505_1696 CBS domain-containing protein
AR505_1813 CAAX amino terminal protease family protein
AR505_0064 translin family DNA-binding protein
AR505_1177 MMPL domain-containing protein
AR505_1445 metallo-beta-lactamase domain protein
AR505_0059 radical SAM domain protein
AR505_0074 MORN repeat-containing protein
AR505_0075 TraB family protein
AR505_0165 radical SAM domain-containing protein
AR505_0195 alpha-NAC homolog
AR505_0208 sterol binding protein
AR505_0216 TIGR01210 family protein
AR505_0332 MORN repeat-containing protein
AR505_0393 small GTP-binding protein
AR505_1019 Raf kinase inhibitor-like protein
AR505_1023 ATP/GTP-binding protein
AR505_1178 radical SAM domain containing protein
AR505_1645 Met-10+ like-protein

VITAMINS AND COFACTORS

Cobalamin

AR505_0747 pyridoxamine 5'-phosphate oxidase PPOX
AR505_0680 3-oxoadipate enol-lactonase PcaD
AR505_0299 cobalamin biosynthesis protein CbiX
AR505_0360 cobalamin biosynthesis protein CbiX
AR505_0363 cobyrinic acid a,c-diamide synthase CbiA
AR505_0378 cobalt-precorrin-5 (C1)-methyltransferase CbiD
AR505_0379 cobalt-precorrin 5A hydrolase CbiG
AR505_0380 precorrin-6X reductase CbiJ
AR505_0381 cobyrinic acid a,c-diamide synthase CbiA
AR505_0398 cobaltochelatase CobN
AR505_0404 magnesium chelatase ChlD
AR505_0405 magnesium chelatase ChlI
AR505_0406 cobaltochelatase CobN
AR505_0413 precorrin-2 C20-methyltransferase CbiL
AR505_0414 precorrin-4 C11-methyltransferase CbiF
AR505_0415 precorrin-3B C17-methyltransferase CbiH
AR505_0416 precorrin-8X methylmutase CbiC
AR505_0444 cobyrinic acid a,c-diamide adenosyltransferase CobA
AR505_0473 nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase CobT
AR505_0597 anaerobic cobalt chelatase CbiK
AR505_0716 cobalamin (vitamin B12) biosynthesis protein CbiX
AR505_0769 cobaltochelatase subunit
AR505_0770 cobaltochelatase subunit
AR505_0771 cobaltochelatase subunit
AR505_0837 cobalt ABC transporter ATP-binding protein CbiO
AR505_0838 cobalt ABC transporter ATP-binding protein CbiO
AR505_0839 cobalt ABC transporter, permease protein CbiQ
AR505_0840 cobalt transport protein CbiN
AR505_0841 cobalt transport protein CbiM
AR505_0947 cobalamin biosynthesis protein CobW
AR505_1040 uroporphyrinogen III synthase HemD
AR505_1041 uroporphyrin-III C-methyltransferase CobA
AR505_1042 porphobilinogen deaminase HemC
AR505_1043 glutamate-1-semialdehyde aminomutase HemL
AR505_1044 porphobilinogen synthase HemB
AR505_1045 glutamyl-tRNA reductase HemA
AR505_1046 siroheme synthase CysG
AR505_1136 cobyric acid synthase CbiP
AR505_1137 cobyric acid synthase CbiP
AR505_1295 adenosylcobinamide-phopsphate synthase CbiB
AR505_1296 adenosylcobinamide amidohydrolase CbiZ
AR505_1297 adenosylcobalamin synthase CobS
AR505_1377 uroporphyrinogen decarboxylase
AR505_1503 cobalamin biosynthesis protein CbiX
AR505_1771 L-threonine kinase PduX

F_{430}

AR505_1211 F_{390} synthetase FtsA

Tetrahydrofolate

AR505_1639 bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase FolD

Metal-binding pterin

AR505_0832 molybdate ABC transporter ATP-binding protein ModC
AR505_1435 molybdenum cofactor biosynthesis protein A MoaA1
AR505_1263 molybdenum cofactor biosynthesis protein A MoaA1

Nicotinate

AR505_0121 ATP-NAD kinase
AR505_0143 L-aspartate dehydrogenase
AR505_0377 NAD synthetase NadE
AR505_0421 aspartate 1-decarboxylase PanD
AR505_0471 quinolinate phosphoribosyltransferase (decarboxylating) NadC
AR505_0616 nicotinamide-nucleotide adenylyltransferase
AR505_1155 quinolinate synthetase A protein NadA
AR505_1654 quinolinate phosphoribosyltransferase (decarboxylating) NadC

Others

AR505_1027 cysteine desulfurase SufS subfamily SufS
AR505_0169 dihydropteroate synthase FolP
AR505_0729 bifunctional protein FolC
AR505_1304 bifunctional protein FolC
AR505_0015 arsenate reductase
AR505_1206 NIF3 family protein
AR505_1316 1,4-dihydroxy-2-naphthoate octaprenyltransferase MenA
AR505_0938 FeS assembly ATPase SufC
AR505_0939 FeS assembly protein SufBD
AR505_1279 iron-sulfur cluster assembly protein
AR505_1447 nitrogenase cofactor biosynthesis protein NifB
AR505_1268 dinitrogenase iron-molybdenum cofactor biosynthesis protein
AR505_1413 dinitrogenase iron-molybdenum cofactor biosynthesis protein

Pantothenate

AR505_1689 bifunctional phosphopantothenate-cysteine ligase/phosphopantothenoylcysteine decarboxylase CoaBC
AR505_0680 3-oxoadipate enol-lactonase PcaD

Riboflavin

AR505_1294 threonine-phosphate decarboxylase CobD
AR505_1180 3,4-dihydroxy-2-butanone-4-phosphate synthase RibB
AR505_1182 riboflavin synthase RibC
AR505_1183 6,7-dimethyl-8-ribityllumazine synthase RibH
AR505_1243 nucleotide pyrophosphatase MazG
AR505_1602 bifunctional diaminohydroxyphosphoribosylaminopyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino) uracil reductase RibD

Thiamine

AR505_0032 thiamine biosynthesis protein ThiC1

AR505_0100	thiamine monophosphate synthase ThiE
AR505_0101	tyrosine lyase ThiH
AR505_0102	thiazole synthase ThiG
AR505_0103	thiamine biosynthesis protein ThiF
AR505_0104	ThiS sulfur-carrier protein
AR505_0128	thiazole biosynthesis/tRNA modification protein ThiI
AR505_0166	thiamine-monophosphate kinase ThiL
AR505_0173	thiazole biosynthesis adenylyltransferase ThiF
AR505_0458	thiazole biosynthesis adenylyltransferase ThiF
AR505_0618	thiazole biosynthesis adenylyltransferase ThiF
AR505_0643	thiamine monophosphate synthase ThiE
AR505_0644	hydroxyethylthiazole kinase ThiM
AR505_0650	4-amino-2-methyl-5-hydroxymethylpyrimidine phosphate kinase ThiD
AR505_0726	ApbE family protein
AR505_0765	phosphomethylpyrimidine kinase ThiD
AR505_1060	cysteine desulfurase NifS
AR505_1660	thiamine biosynthesis ATP pyrophosphatase ThiI
Ubiquinone	
AR505_0114	methylase involved in ubiquinone/menaquinone biosynthesis
AR505_0312	ubiquinone/menaquinone biosynthesis methyltransferase UbiE
AR505_0969	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
AR505_1172	3-octaprenyl-4-hydroxybenzoate decarboxylyase UbiD
AR505_1317	ubiquinone/menaquinone biosynthesis methyltransferase

Table A.4.19. Percentage codon usage in Methanomassiliicoccales genomes

Amino acid[#]	Codon	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10	Mpt1
Ser (S)	UCA	0.5 [7.1]	0.5 [8.6]	0.3 [5.2]	0.3 [4.0]	0.3 [4.9]	0.7 [9.8]	0.4 [5.2]	0.3 [4.3]	2.1 [28.2]	0.3 [5.2]	1 [15.1]
	UCC	2.8 [42.8]	2.5 [39.9]	3.4 [51.0]	3.4 [54.3]	2 [32.9]	2.7 [39.2]	2.8 [39.5]	3.1 [48.4]	0.9 [11.9]	2.3 [36.8]	1.5 [22.7]
	UCG	1.0 [15.5]	1.4 [21.6]	1.2 [17.5]	1.0 [15]	1.8 [30.2]	1.1 [15.6]	1.7 [23.7]	1.2 [18.3]	0.5 [6.5]	1.4 [22.7]	1.4 [21.6]
	UCU	0.5 [8.5]	0.4 [5.8]	0.3 [4.9]	0.3 [4.7]	0.2 [4]	0.6 [9.2]	0.4 [5.7]	0.4 [6.2]	2.0 [26.5]	0.3 [5.2]	0.8 [12.5]
	AGC	1.3 [19.4]	1.2 [19.8]	1.1 [16.9]	1.2 [18.8]	1.6 [25.8]	1.5 [22.4]	1.7 [23.5]	1.1 [17.3]	1.2 [15.8]	1.7 [27.3]	1.4 [20.6]
	AGU	0.4 [6.7]	0.3 [4.4]	0.3 [4.4]	0.2 [3.1]	0.1 [2.3]	0.3 [3.8]	0.2 [2.3]	0.4 [5.5]	0.8 [11.0]	0.2 [2.7]	0.5 [7.5]
Phe (F)	UUC	3.6 [92.3]	3.6 [95.3]	3.7 [95.1]	3.7 [95.9]	3.2 [93.8]	3.5 [93.5]	3.5 [96.4]	3.5 [93.8]	1.7 [45.1]	3.3 [91.7]	3.1 [80.0]
	UUU	0.3 [7.7]	0.2 [4.7]	0.2 [4.9]	0.2 [4.1]	0.2 [6.2]	0.2 [6.5]	0.1 [3.6]	0.2 [6.2]	2.0 [54.9]	0.3 [8.3]	0.8 [20.0]
Thr (T)	ACA	0.6 [10.4]	0.9 [16.5]	0.5 [9.3]	0.4 [7.0]	0.2 [4.4]	0.9 [17.8]	0.4 [8.4]	0.6 [10.7]	2.2 [41.2]	0.3 [5.8]	1.8 [31.4]
	ACC	3.8 [66.2]	2.8 [51.0]	3.6 [64.9]	3.9 [73.3]	3 [60.7]	2.2 [42.4]	3.2 [60.4]	3.0 [58.7]	0.9 [16.5]	3.2 [64.5]	1.8 [30.6]
	ACG	0.7 [12.4]	1.4 [25.8]	1.1 [19.8]	0.8 [14]	1.5 [29.8]	1.5 [30]	1.2 [23.7]	1.3 [24.5]	0.6 [10.3]	1.1 [22.0]	1.6 [27.1]
	ACU	0.6 [10.9]	0.4 [6.7]	0.3 [6.0]	0.3 [5.7]	0.2 [5]	0.5 [9.8]	0.4 [7.6]	0.3 [6.0]	1.7 [32.0]	0.4 [7.6]	0.6 [19.8]
Asn (N)	AAC	3.0 [76.9]	3.1 [84.0]	2.9 [82.9]	3.0 [88.1]	2.3 [84.5]	2.2 [66.8]	2.5 [82.8]	2.7 [79.3]	1.8 [38.8]	2.4 [81.4]	2.7 [65.2]
	AAU	0.9 [23.1]	0.6 [16.0]	0.6 [17.7]	0.4 [11.9]	0.4 [15.5]	1.1 [33.2]	0.5 [17.2]	0.7 [20.7]	2.8 [61.2]	0.5 [18.6]	1.5 [34.8]
Lys (K)	AAA	1.7 [29.3]	0.9 [15.8]	1.2 [23.3]	0.9 [15.3]	0.3 [7.5]	1.4 [25.4]	1 [19.4]	1.5 [25.3]	4.3 [67.4]	0.6 [13.4]	3.3 [48.7]
	AAG	4.2 [70.7]	5.0 [84.2]	4.0 [76.7]	4.8 [84.7]	3.7 [92.5]	4.2 [74.6]	4 [80.6]	4.4 [74.7]	2.1 [32.6]	4.3 [86.6]	3.5 [51.3]
Glu (E)	GAA	2.3 [34.1]	1.5 [22.9]	1.7 [25.9]	1.2 [17.5]	1 [14.7]	2 [28]	1.3 [21]	1.9 [28.7]	4.7 [69.1]	1.3 [19.0]	3.1 [47.6]
	GAG	4.4 [65.9]	5.2 [77.1]	4.9 [74.1]	5.5 [82.5]	6 [85.3]	5 [72]	5.1 [79]	4.8 [71.3]	2.1 [30.9]	5.5 [81.0]	3.4 [52.4]
Tyr (Y)	UAC	2.6 [70.4]	2.8 [76.9]	2.8 [77.5]	2.8 [82.2]	2 [73.7]	2 [64.6]	2.7 [77.3]	2.3 [65.7]	1.4 [38.4]	2.2 [73.7]	1.7 [52.7]
	UAU	1.1 [29.6]	0.8 [23.1]	0.8 [22.5]	0.6 [17.8]	0.7 [26.3]	1.1 [35.4]	0.8 [22.7]	1.2 [34.3]	2.3 [61.6]	0.8 [26.3]	1.6 [47.3]
Val (V)	GUA	1.0 [13.1]	0.7 [9.2]	0.6 [8.3]	0.5 [6.6]	0.4 [5.4]	1 [12.9]	0.4 [5.9]	1.0 [13.2]	2.3 [32.7]	0.6 [7.5]	1.4 [19.4]
	GUC	3.5 [47.5]	4.2 [55.1]	4.3 [55.2]	4.7 [59.8]	3.4 [41.1]	2.8 [36.7]	4.2 [56.9]	4.0 [51.0]	1.3 [19.0]	3.1 [38.4]	2.5 [33.6]
	GUG	2.1 [28.4]	2.2 [28]	2.3 [29.3]	2.1 [25.8]	3.9 [48.3]	2.8 [36.6]	2.3 [30.9]	2.4 [30.3]	1.2 [17.6]	3.9 [47.8]	2.1 [28.6]
	GUU	0.8 [11]	0.6 [7.7]	0.6 [7.2]	0.6 [7.8]	0.4 [5.2]	1 [13.8]	0.5 [6.4]	0.4 [5.6]	2.2 [30.7]	0.5 [6.3]	1.4 [18.4]
Gln (Q)	CAA	0.3 [10.9]	0.2 [7.9]	0.2 [9.9]	0.1 [5.5]	0.2 [9.9]	0.3 [13.3]	0.1 [6.1]	0.2 [11.1]	0.7 [30.4]	0.3 [11.8]	0.6 [28.8]
	CAG	2.1 [89.1]	2.2 [92.1]	1.9 [90.1]	2.2 [94.5]	2.3 [90.1]	1.6 [86.7]	2 [93.9]	1.9 [88.9]	1.6 [69.6]	2.2 [88.2]	1.5 [71.2]
Met (M)	AUG	3.2 [100.0]	3.6 [100.0]	3.3 [100.0]	3.3 [100.0]	3.2 [100]	3.6 [100]	3.2 [100]	3.5 [100.0]	2.8 [100.0]	3.1 [100.0]	3.3 [100.0]
Cys (C)	UGC	1.2 [74.5]	1.3 [81.6]	1.3 [79.8]	1.4 [87.2]	1.1 [89]	1.2 [83.9]	1.6 [94]	1.3 [75.2]	0.9 [55.6]	1.1 [90.5]	1.0 [69.6]
	UGU	0.4 [25.5]	0.3 [18.4]	0.3 [20.2]	0.2 [12.8]	0.1 [11]	0.2 [16.1]	0.1 [6]	0.4 [24.8]	0.7 [44.4]	0.1 [9.5]	0.4 [30.4]
Leu (L)	CUA	0.2 [2.7]	0.2 [2.9]	0.2 [2.4]	0.2 [2.3]	0.4 [3.7]	0.2 [2.5]	0.1 [1.4]	0.2 [2.2]	0.6 [7.2]	0.3 [3.1]	0.3 [3.3]
	CUC	3.5 [42.5]	3.1 [38.5]	4.3 [54.0]	4.0 [50.5]	2.5 [25.7]	2.2 [26.8]	3.5 [43.9]	3.2 [40.1]	1.1 [13.2]	2.6 [27.3]	1.7 [21.9]
	CUG	2.8 [34.4]	3.5 [42.2]	2.4 [29.4]	2.7 [34.8]	5.5 [56.9]	3.7 [44.2]	3.4 [43.7]	2.9 [37.0]	2.3 [26.8]	5.1 [54.1]	2.4 [31]
	CUU	1.0 [11.6]	0.7 [8.7]	0.6 [6.9]	0.7 [8.8]	0.6 [6.4]	1.3 [15.5]	0.6 [7.1]	0.9 [11.3]	2.0 [22.8]	0.7 [7.3]	1.8 [22.6]
	UUA	0.2 [2.7]	0.1 [1.5]	0.1 [1.7]	0.1 [0.8]	0.1 [1]	0.2 [2.1]	0.1 [0.7]	0.1 [1.6]	1.5 [17.1]	0.1 [1.4]	0.5 [5.9]
	UUG	0.5 [6.1]	0.5 [6.2]	0.4 [5.5]	0.2 [2.8]	0.6 [6.4]	0.7 [8.9]	0.2 [3.2]	0.6 [7.9]	1.1 [12.9]	0.6 [6.8]	1.2 [15.3]
Ala (A)	GCA	2.0 [25.7]	1.7 [22.9]	1.1 [13.6]	1.5 [17.5]	0.6 [6.1]	1.7 [20.5]	1 [11]	1.5 [19.1]	3.1 [42.0]	0.6 [6.7]	2.4 [31.3]
	GCC	3.5 [44]	3.1 [40.4]	4.4 [53.5]	4.1 [48.7]	5.3 [57.6]	3.3 [41.1]	4.9 [51.4]	4.1 [50.6]	1.3 [17.5]	4.8 [52.4]	2.0 [25.9]
	GCG	1.5 [18.6]	2.2 [28.4]	2.0 [24.1]	2.3 [27.1]	2.8 [30]	2.2 [26.6]	2.9 [30]	2.0 [24.8]	0.7 [9.1]	3.0 [33.1]	2.4 [30.2]
	GCU	0.9 [11.8]	0.6 [8.3]	0.7 [8.8]	0.6 [6.8]	0.6 [6.2]	1 [11.9]	0.7 [7.6]	0.4 [5.5]	2.3 [31.3]	0.7 [7.8]	1.0 [12.6]
O (Pyl)	UAG	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]	0.0 [100.0]
Trp (W)	UGG	1.0 [100.0]	0.9 [100.0]	1.0 [100.0]	0.9 [100.0]	1.1 [100]	0.9 [100]	1 [100]	0.9 [100.0]	0.9 [100.0]	1.1 [100.0]	0.9 [100.0]
Pro (P)	CCA	0.2 [5.3]	0.3 [6.5]	0.2 [3.9]	0.2 [4.0]	0.3 [7.5]	0.2 [5.4]	0.2 [3.7]	0.2 [4.5]	1.2 [32.8]	0.3 [6.2]	0.4 [10.5]
	CCC	2.2 [56.6]	1.8 [47.7]	2.5 [62.7]	2.5 [61.6]	1.9 [41.3]	1.6 [40.4]	1.7 [41.7]	1.8 [46.9]	0.5 [13.2]	2.0 [45.4]	1.0 [27.4]
	CCG	0.8 [20.1]	1.3 [32.9]	1.0 [23.7]	1.0 [25.9]	1.9 [41.1]	1.4 [35.6]	1.7 [41.5]	1.3 [33.2]	0.8 [21.4]	1.7 [38.5]	1.7 [45.5]
	CCU	0.7 [18.0]	0.5 [12.8]	0.4 [9.6]	0.3 [8.5]	0.5 [10.1]	0.7 [18.6]	0.5 [13]	0.6 [15.4]	1.2 [32.5]	0.4 [9.9]	0.6 [16.6]

His (H)	CAC	1.1 [67.4]	1.2 [70.6]	1.2 [72.3]	1.3 [80.1]	1.2 [62.5]	1 [60.2]	1 [63.8]	1.0 [60.8]	0.6 [38.2]	1.3 [71.1]	0.9 [58.8]
	CAU	0.6 [32.6]	0.5 [29.4]	0.5 [27.7]	0.3 [19.9]	0.7 [37.5]	0.6 [39.8]	0.6 [36.2]	0.6 [39.2]	1.0 [61.8]	0.5 [28.9]	0.6 [41.2]
Asp (D)	GAC	4.0 [61.6]	4.6 [67.4]	4.7 [68.1]	5.1 [76.0]	4.6 [77.5]	3.6 [55.9]	4.7 [70.8]	4.9 [70.6]	2.0 [33.0]	4.5 [75.9]	3.1 [50.8]
	GAU	2.5 [38.4]	2.2 [32.6]	2.2 [31.4]	1.6 [24.0]	1.3 [22.5]	2.8 [44.1]	1.9 [29.2]	2.0 [29.4]	4.1 [67.0]	1.4 [24.1]	3.0 [49.2]
Arg (R)	CGA	0.0 [0.3]	0.0 [0.3]	0.0 [0.2]	0.0 [0.2]	0.5 [6.6]	0.1 [2.1]	0.1 [1.4]	0.0 [0.5]	0.2 [4.3]	0.2 [3.7]	0.1 [2.6]
	CGC	1.2 [23.3]	0.9 [18.6]	1.7 [32.4]	1.5 [28.0]	2.3 [32.3]	0.9 [15.2]	1.9 [31.5]	1.0 [19.2]	0.5 [12.6]	1.8 [29.4]	0.6 [12.5]
	CGG	0.2 [3.3]	0.1 [1.5]	0.2 [2.9]	0.1 [2.0]	1.7 [24.3]	0.4 [7.1]	0.5 [7.7]	0.2 [4.0]	0.2 [4.9]	1.4 [22.8]	0.2 [4.8]
	CGU	0.7 [14.6]	0.5 [9.6]	0.7 [13.6]	0.4 [8.3]	0.4 [5.8]	0.4 [6.2]	0.3 [4.7]	0.7 [12.1]	0.5 [13.5]	0.2 [3.8]	0.3 [7.4]
	AGA	0.7 [14.8]	0.7 [13.6]	0.5 [9.7]	0.5 [9.0]	0.4 [5.1]	1.4 [23.1]	0.5 [23.1]	0.9 [16.4]	1.8 [45.2]	0.4 [6.5]	1.6 [34.6]
	AGG	2.2 [43.6]	2.7 [56.4]	2.2 [41.4]	2.8 [52.6]	1.8 [25.8]	2.7 [46.4]	0.5 [7.9]	2.6 [47.9]	0.8 [19.5]	2.1 [33.8]	1.7 [38]
Ile (I)	AUA	1.0 [14.3]	1.5 [19.7]	1.3 [18.4]	0.8 [11.3]	1 [17.2]	3.7 [53.1]	2.8 [46.8]	2.3 [34.5]	2.5 [31.4]	1.3 [22.2]	3.2 [41.8]
	AUC	5.0 [73.3]	5.4 [73.8]	5.0 [74.7]	5.6 [82.1]	4.4 [77]	2.8 [40.3]	4.7 [75.5]	4.0 [60.2]	2.6 [32.0]	4.3 [70.4]	3.5 [45.3]
	AUU	0.8 [12.4]	0.5 [6.5]	0.5 [6.9]	0.4 [6.6]	0.3 [5.8]	0.5 [6.6]	0.3 [5.2]	0.3 [5.3]	2.9 [36.7]	0.4 [7.4]	1.0 [13.0]
Gly (G)	GGA	3.2 [41.2]	2.8 [37.3]	1.9 [23.9]	2.7 [34.6]	1.3 [15.9]	2 [24.4]	1.5 [18.4]	2.5 [30.9]	3.2 [44.4]	1.1 [13.0]	2.6 [32.4]
	GGC	1.9 [25.2]	2.0 [26.2]	2.5 [31]	2.3 [28.8]	4.2 [51.4]	2.9 [35.6]	3.5 [42.1]	1.9 [23.3]	1.6 [22.1]	4.0 [47.8]	2.5 [30.9]
	GGG	0.9 [12.0]	1.4 [18.1]	1.9 [23.2]	1.9 [23.5]	2 [24.7]	2.1 [26]	2.8 [33.4]	2.3 [28.2]	0.7 [9.5]	2.7 [32.5]	1.6 [19.9]
	GGU	1.7 [21.6]	1.4 [18.3]	1.7 [21.8]	1.0 [13.1]	0.7 [8]	1.1 [14]	0.5 [6.2]	1.4 [17.6]	1.7 [24.0]	0.5 [6.7]	1.4 [16.8]
Stop codons*												
ochre	UAA	13.4	14.1	11.3	8.8	13.3	18	10.9	14.1	45.2	17.5	33.1
amber	UAG	1.4 (2.7)	1.6 (3.2)	1.0 (2.4)	0.5 (2.5)	11.1 (16.2)	6.6 (10.1)	2.1 (5.1)	1.6 (2.8)	5.0 (5.3)	11.0 (11.3)	1.8 (2.3)
opal	UGA	84.1	84.4	86.3	89.1	70.4	72	83.9	84.3	49.7	71.5	65
Translation initiator												
M	AUG	93.6	93.5	89.7	84.5	76.4	85	81.2	83.7	84.4	75.8	85.5
L	GUG	1.7	4.0	7.1	11.2	6.8	5	5	5.2	8.0	9.2	9.7
V	UUG	4.7	2.5	3.2	4.2	11.9	6.6	12	11.1	7.5	15.0	4.7
I	ATT	0	0	0	0	0.7	0.1	0.8	0	0	0	0

[#] number indicates percentage of codon in genome, number in square bracket indicates percentage of codon used for each amino acid.

*the number in brackets indicates total in-frame UAG, without considering its use as a STOP codon or as a Pyl codon.

Table A.4.20. CRISPR associated genes in Methanomassiliicoccales genomes analysed

gene	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1	Reference*
cas1	-	-	-	-	-	-	-	+	+	+	+	SERP2463
cas2	-	-	-	-	-	-	-	-	-	-	-	SERP2462
cas3'	+	+	+	+	+	+	+	-	+	-	+	APE1232
cas3''	-	-	-	-	-	-	-	-	-	-	-	APE1231
cas4	-	-	-	-	-	-	+	-	-	-	-	APE1239
cas5	-	-	-	-	-	-	-	-	-	-	-	APE1234
cas6	+	-	-	-	-	-	-	-	-	-	-	PF1131
cas6e	-	-	-	-	-	-	-	-	-	-	-	*ygcH*
cas6f	-	-	-	-	-	-	-	-	-	-	-	y1727
cas7	+	-	-	-	-	-	-	-	-	-	-	*ygcJ*
cas8a1	-	-	-	-	-	-	-	-	-	-	-	LA3191
cas8a2	-	-	-	-	-	-	-	-	-	-	-	MJ0385
cas8b	-	-	-	-	-	-	-	-	-	-	-	MTH1090
cas8c	-	-	-	-	-	-	-	-	-	+		BH0338
cas9	-	-	-	-	-	-	-	-	-	-	-	Spy1046
cas10/csm1	-	-	-	-	-	-	-	-	-	-	-	MTH326
cas10d	-	-	-	-	-	-	-	-	-	-	-	Slr7011
csy1	-	-	-	-	-	-	-	-	-	-	-	Y1724
csy2	-	-	-	-	-	-	-	-	-	-	-	Y1725
csy3	-	-	-	-	-	-	-	-	-	-	-	Y1726
cse1	+	-	-	-	-	-	-	-	-	-	-	*ygcL*
cse2	-	-	-	-	-	-	-	-	-	-	-	*ygcK*
csc1	-	-	-	-	-	-	-	-	-	-	-	alr1563
csc2	-	-	-	-	-	-	-	-	-	-	-	slr7012
csa5	-	-	-	-	-	-	-	-	-	-	-	MJ0380
csn2	-	-	-	-	-	-	-	-	-	-	-	SPy1049
csm2	-	-	-	-	-	-	-	-	-	-	-	MTH1081
csm3	-	-	-	-	-	-	-	-	-	-	-	MTH1080
csm4	-	-	-	-	-	-	-	-	-	-	-	MTH1079
csm5	-	-	-	-	-	-	-	-	-	-	-	MTH1078
csm6	-	-	-	-	-	-	-	-	-	-	-	APE2256
cmr1	-	-	-	-	-	-	-	-	-	-	-	PF1130
cmr3	-	-	-	-	-	-	-	-	-	-	-	PF1128
cmr4	-	-	-	-	-	-	-	-	-	-	-	PF1126
cmr5	-	-	-	-	-	-	-	-	-	-	-	MTH324
cmr6	-	-	-	-	-	-	-	-	-	-	-	PF1124
csb1	-	-	-	-	-	-	-	-	-	-	-	Balac_1306
csb2	-	-	-	-	-	-	-	-	-	-	-	Balac_1305
csb3	-	-	-	-	-	-	-	-	-	-	-	Balac_1303
csx17	-	-	-	-	-	-	-	-	-	-	-	Btus_2683
csx14	-	-	-	-	-	-	-	-	-	-	-	GSU0052
csx10	-	-	-	-	-	-	-	-	-	-	-	Caur_2274
csx16	-	-	-	-	-	-	-	-	-	-	-	VVA1548
csaX	-	-	-	-	-	-	-	-	-	-	-	SSO1438
csx3	-	-	-	-	-	-	-	-	-	-	-	AF1864
csx1	-	-	-	-	-	-	-	-	-	-	-	MJ1666
csx15	-	-	-	-	-	-	-	-	-	-	-	TTE2665
csf1	-	-	-	-	-	-	-	-	-	-	-	AFE_1038
csf2	-	-	-	-	-	-	-	-	-	-	-	AFE_1039
csf3	-	-	-	-	-	-	-	-	-	-	-	AFE_1040
csf4	-	-	-	-	-	-	-	-	-	-	-	AFE_1037

Prediction cutoff: E-04. +Represents the gene is predicted in the particular genome. *Reference protein is displayed in either locus tag or gene name within the NCBI protein database

Table A.4.21. Predicted pyrrolysine usage in the ISO4-G11 genome

Locus tag	Predicted gene product	Distance (codons)*	Class#	Merge with
ISO4G11_0002	**trimethylamine:corrinoid methyltransferase**	164	1	ISO4G11_0003
ISO4G11_0009	**monomethylamine:corrinoid methyltransferase MttB**	259	1	ISO4G11_0010
ISO4G11_0011	**monomethylamine:corrinoid methyltransferase MtmB**	257	1	ISO4G11_0012
ISO4G11_0065	adenosylcobinamide amidohydrolase	77	1	ISO4G11_0066
ISO4G11_0237	serine hydroxymethyltransferase	350	1	ISO4G11_0238
ISO4G11_0250	hypothetical protein	3	3	-
ISO4G11_0274	FMN reductase	44	3	-
ISO4G11_0371	adhesin-like protein	332	1	ISO4G11_0372
ISO4G11_0437	**DNA helicase**	374	1	ISO4G11_0438
ISO4G11_0457	hypothetical protein	24, 18	3	-
ISO4G11_0513	uncharacterized protein	13	1	-
ISO4G11_0515	ATPase family	174	1	-
ISO4G11_0551	**dimethylamine:corrinoid methyltransferase MtbB**	115	1	ISO4G11_0552
ISO4G11_0579	uncharacterized protein	147	3	-
ISO4G11_0596	transcriptional regulator	149	1	ISO4G11_0597
ISO4G11_0645	phage-type endonuclease	81	1	ISO4G11_0646
ISO4G11_0683	hypothetical protein	1	3	-
ISO4G11_0686	hypothetical protein	8	2	-
ISO4G11_0719	hypothetical protein	28	1	-
ISO4G11_0741	hypothetical protein	224	1	ISO4G11_0740
ISO4G11_0750	type III restriction enzyme	695	1	ISO4G11_0751
ISO4G11_0768	protein tyrosine phosphatase	7	2	-
ISO4G11_0770	modE molybdate transport repressor domain	134	1	ISO4G11_0771
ISO4G11_0858	adenosylhomocysteinase	42	1	-
ISO4G11_0862	dephospho-CoA kinase	10	2	-
ISO4G11_0872	mobile mystery protein	3	2	-
ISO4G11_1175	transposase	230	1	ISO4G11_1176
ISO4G11_1233	hypothetical protein	218	1	ISO4G11_1232
ISO4G11_1353	TPR-repeat lipoprotein	26	2	-
ISO4G11_1446	hypothetical protein	276	3	-
ISO4G11_1514	**dimethylamine:corrinoid methyltransferase MtbB**	112	1	ISO4G11_1513
ISO4G11_1647	hypothetical protein	93	3	-
ISO4G11_1701	hypothetical protein	253	3	-
ISO4G11_1730	hypothetical protein	4	3	-

*Distance between amber codon to the next opal or ochre stop codon. # Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.22. Predicted pyrrolysine usage in the ISO4-G1 genome

Locus_tag	Predicted gene product	Distance (codons)*	Class#	Merge with
ISO4G1_0043	**nucleotide kinase**	18	2	-
ISO4G1_0180	**phosphomannomutase/phosphoglucomutase**	54	1	ISO4G1_0181
ISO4G1_0211	type IV leader peptidase family protein	140	1	ISO4G1_0212
ISO4G1_0245	amidohydrolase	298	1	ISO4G1_0246
ISO4G1_0292	radical SAM enzyme TIGR01210 family	86	1	ISO4G1_0293
ISO4G1_0371	hypothetical protein	20	3	-
ISO4G1_0394	hypothetical protein	122	3	-
ISO4G1_0470	**trimethylamine:corrinoid methyltransferase MttB**	161	1	ISO4G1_0471
ISO4G1_0477	**dimethylamine:corrinoid methyltransferase MtbB**	103	1	ISO4G1_0478
ISO4G1_0490	hypothetical protein	318, 269	2	-
ISO4G1_0554	type 11 methyltransferase	276	1	ISO4G1_0555
ISO4G1_0572	aconitate hydratase 1 AcnA	285	1	ISO4G1_0573
ISO4G1_0606	hypothetical protein	30, 7	2	-
ISO4G1_0654	**monomethylamine:corrinoid methyltransferase MtmB**	259	1	ISO4G1_0655
ISO4G1_0687	hypothetical protein	99	2	-
ISO4G1_0698	hypothetical protein	404	3	-
ISO4G1_0710	ATP-dependent DNA helicase	297, 222	1	ISO4G1_0711, ISO4G1_0712
ISO4G1_0713	hypothetical protein	6	3	-
ISO4G1_0714	hypothetical protein	135	3	-
ISO4G1_0767	hypothetical protein	199	3	-
ISO4G1_0792	ATP-dependent DNA helicase	552	1	ISO4G1_0793
ISO4G1_0800	sugar fermentation stimulation protein A	62	1	-
ISO4G1_0803	VTC domain-containing protein	65	1	ISO4G1_0804
ISO4G1_0806	hypothetical protein	83	3	-
ISO4G1_0851	hypothetical protein	178	3	-
ISO4G1_0871	type II secretion system protein	299	1	ISO4G1_0872
ISO4G1_0938	**dimethylamine:corrinoid methyltransferase MtbB**	112	1	ISO4G1_0939
ISO4G1_0946	**trimethylamine:corrinoid methyltransferase MttB**	160	1	ISO4G1_0947
ISO4G1_0968	hypothetical protein	369, 312, 305, 100, 98	3	-
ISO4G1_00984	non-ribosomal peptide synthetase	4141, 1774	1	ISO4G1_0985, ISO4G1_0986
ISO4G1_1003	X-prolyl dipeptidylaminopeptidase	69	1	
ISO4G1_1018	ATP-dependent DNA ligase	542	1	ISO4G1_1019
ISO4G1_1063	nitrogenase component 1 type oxidoreductase	11	1	-
ISO4G1_1090	Na/Pi-cotransporter II-related protein	160	1	ISO4G1_1089
ISO4G1_1122	archaeosine tRNA-ribosyltransferase	25	2	-
ISO4G1_1230	**trimethylamine:corrinoid methyltransferase MttB**	164	1	ISO4G1_1231
ISO4G1_1234	**dimethylamine:corrinoid methyltransferase MtbB**	113	1	ISO4G1_1235
ISO4G1_1238	**monomethylamine:corrinoid methyltransferase MtmB**	279	1	ISO4G1_1239
ISO4G1_1296	tripartite tricarboxylate transporter TctA family	5	2	-
ISO4G1_1349	iron ABC transporter ATP-binding protein	8	1	-
ISO4G1_1444	**trimethylamine:corrinoid methyltransferase MttB**	162	1	ISO4G1_1445
ISO4G1_1478	**nitrogenase iron protein NifH**	171	1	ISO4G1_1477
ISO4G1_1480	**glycosyl transferase GT8 family**	207, 101	1	ISO4G1_1481, ISO4G1_1482

*Distance between amber codon to the next opal or ochre stop codon. # Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.23. Predicted pyrrolysine usage in the BRNA1 genome

Locus_tag	Predicted gene product	Distance (codons)*	Class#	Merge with
TALC_00050	**putative nucleotide kinase (CMP and AMP kinases -like protein)**	50	2	-
TALC_00054	eukaryotic-type DNA primase, large subunit	91	1	-
TALC_00068	**glycosyltransferases involved in cell wall biogenesis**	190	1	TALC_00067
TALC_00088	hypothetical protein	4	3	-
TALC_00154	hypothetical protein	169	2	TALC_00155
TALC_00163	hypothetical protein	12	2	-
TALC_00237	BC-type spermidine/putrescine transport systems, ATPase component	69	1	TALC_00238
TALC_00301	**dimethylamine:corrinoid methyltransferase MtbB**	111	1	-
TALC_00305	**trimethylamine:corrinoid methyltransferase MttB**	171	1	-
TALC_00366	hypothetical protein	20	2	-
TALC_00418	dinitrogenase iron-molybdenum cofactor	352	1	TALC_00417
TALC_00490	Acetyltransferase (GNAT) family	47	1	TALC_00491
TALC_00532	NaMN:DMB phosphoribosyltransferase	142	1	TALC_00531
TALC_00554	carbamoylphosphate synthase small subunit	218	1	TALC_00555
TALC_00591	hypothetical protein	365	1	TALC_00592
TALC_00657	hypothetical protein	164	1	TALC_00655
TALC_00720	phenylacetate-CoA ligase	392	1	TALC_00719
TALC_00726	putative phosphatase	181	1	TALC_00727
TALC_00896	ABC-type cobalamin/$Fe3^{+}$-siderophores transport systems, ATPase	39	1	-
TALC_00943	phosphoglycerate mutase	140	1	TALC_00944
TALC_01010	hypothetical protein	469	2	TALC_01011
TALC_01032	uncharacterized protein	23	1	-
TALC_01093	**monomethylamine:corrinoid methyltransferase MtmB**	257	1	-
TALC_01094	**monomethylamine:corrinoid methyltransferase MtmB**	259	1	-
TALC_01102	**trimethylamine:corrinoid methyltransferase MttB**	168	1	-
TALC_01103	**dimethylamine:corrinoid methyltransferase MtbB**	112	1	-
TALC_01170	hypothetical protein	66	2	-
TALC_01197	cobalamin biosynthesis protein CobN-related Mg-chelatase	281	1	TALC_01196
TALC_01236	putative amidohydrolase	55	1	TALC_01237
TALC_01357	Na^{+}/proline symporter	141	1	TALC_01358
TALC_01508	L-asparaginase/archaeal Glu-tRNAGln amidotransferase subunit D	128	1	TALC_01507
TALC_01548	8-oxoguanine DNA glycosylase, N-terminal domain protein	99	1	TALC_01549

*Distance between amber codon to the next opal or ochre stop codon. # Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.24. Predicted pyrrolysine usage in the MpT1 genome

Locus_tag	Predicted gene product	Distance (codons)*	Class#	Merge with
Mpt1_c01150	integrase core domain protein	3	2	-
Mpt1_c02990	hypothetical protein	59	1	-
Mpt1_c03550	hypothetical protein	14	3	-
Mpt1_c03980	putative nickel-responsive regulator	21	2	-
Mpt1_c04300	hypothetical protein	7	3	-
Mpt1_c04430	integrase core domain protein	3	2	-
Mpt1_c05280	hypothetical protein	62	2	-
Mpt1_c05350	hypothetical protein	34, 1	3	-
Mpt1_c05360	hypothetical protein	36, 31	3	-
Mpt1_c05730	**monomethylamine:corrinoid methyltransferase MtmB**	258	1	-
Mpt1_c05760	**monomethylamine:corrinoid methyltransferase MtmB**	259	1	-
Mpt1_c07680	**integrase core domain protein**	3	2	-
Mpt1_c08050	sirohydrochlorin cobaltochelatase	154	1	Mpt1_c08040
Mpt1_c08460	hypothetical protein	16	2	-
Mpt1_c08720	hypothetical protein	127	2	-
Mpt1_c09890	hypothetical protein	31	3	-
Mpt1_c10480	hypothetical protein	14	2	-
Mpt1_c11820	archaea bacterial proteins of unknown function	11	2	-
Mpt1_c11860	**dephospho-CoA kinase**	10	2	-
Mpt1_c11940	cell division protein FtsZ 1	7	2	-
Mpt1_c12980	hypothetical protein	32	3	-
Mpt1_c13190	hypothetical protein	134	3	-
Mpt1_c13990	3’ ribonuclease Y	45	2	-
Mpt1_c14260	hypothetical protein	12	3	-
Mpt1_c14440	3’ bifunctional enzyme Fae/Hps	2	2	-

*Distance between amber codon to the next opal or ochre stop codon. # Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.25. Predicted pyrrolysine usage in the RumEn M1 genome

Locus_tag	Predicted gene product	Distance (codons)*	Class#	Merge with
AOA80_00040	geranylgeranylglyceryl phosphate synthase	45	2	-
AOA80_00095	4Fe-4S Fdx	16, 32, 69, 80, 159	2	-
AOA80_00125	hypothetical protein	73, 82	2	-
AOA80_00170	ATPase	32	2	-
AOA80_00200	hypothetical protein	4	2	-
AOA80_00225	hypothetical protein	8	2	-
AOA80_00235	hypothetical protein	99	2	-
AOA80_00240	UDP-glucose 4-epimerase	17	2	-
AOA80_00265	hypothetical protein	365	2	-
AOA80_00280	hypothetical protein	6	2	-
AOA80_00335	aspartate carbamoyltransferase	151	2	-
AOA80_00450	hypothetical protein	31	2	-
AOA80_00530	arginyl-tRNA synthetase	161	2	-
AOA80_00615	reductase	29	2	-
AOA80_00700	hypothetical protein	182	2	-
AOA80_00750	hypothetical protein	27	2	-
AOA80_00785	transferase	5	2	-
AOA80_00805	hypothetical protein	11	2	-
AOA80_00820	hypothetical protein	9, 31	2	-
AOA80_00860	hypothetical protein	13, 25	2	-
AOA80_00865	hypothetical protein	26	2	-
AOA80_00935	UDP-*N*-acetylglucosamine 2-epimerase	3	2	-
AOA80_00955	hypothetical protein	7, 26	2	-
AOA80_00965	hypothetical protein	7	3	-
AOA80_01005	hypothetical protein	30	2	-
AOA80_01030	hypothetical protein	4	2	-
AOA80_01040	hypothetical protein	22	2	-
AOA80_01070	hypothetical protein	118, 396	2	-
AOA80_01075	hypothetical protein	6	2	-
AOA80_01255	hypothetical protein	81	2	-
AOA80_01320	hypothetical protein	11	3	-
AOA80_01330	hydrogenase assembly protein HypC	100	2	-
AOA80_01335	hypothetical protein	60	3	-
AOA80_01345	hypothetical protein	71	2	-
AOA80_01390	hypothetical protein	1	2	-
AOA80_01445	hypothetical protein	296	2	-
AOA80_01470	hypothetical protein	2, 16	2	-
AOA80_01515	hypothetical protein	271	2	-
AOA80_01560	hypothetical protein	24	2	-
AOA80_01590	CopG family transcriptional regulator	16	2	-
AOA80_01610	cell division protein FtsZ	32	2	-
AOA80_01615	cell division protein FtsZ	17, 58	2	-
AOA80_01670	hypothetical protein	38, 77, 88, 99, 107, 146, 165	2	-
AOA80_01690	hypothetical protein	7	2	-
AOA80_01705	hypothetical protein	13	2	-
AOA80_01710	hypothetical protein	6	2	-
AOA80_01720	hypothetical protein	3	2	-
AOA80_01755	riboflavin deaminase	70	2	-
AOA80_01765	hypothetical protein	28	2	-
AOA80_01790	hypothetical protein	46	2	-
AOA80_01830	hypothetical protein	8	2	-
AOA80_01915	hypothetical protein	12	2	-
AOA80_01925	hypothetical protein	30	3	-
AOA80_01970	hypothetical protein	38	2	-
AOA80_02040	hypothetical protein	34	2	-
AOA80_02130	hypothetical protein	167	3	-
AOA80_02165	hypothetical protein	2	2	-
AOA80_02190	hypothetical protein	79, 238	2	-
AOA80_02285	hypothetical protein	47, 83	2	-
AOA80_02290	hypothetical protein	77	2	-
AOA80_02325	hypothetical protein	13, 42	2	-
AOA80_02370	methylthiol--CoM methyltransferase	8	2	-
AOA80_02395	hypothetical protein	57	2	-
AOA80_02480	RNA-associated protein	5	2	-
AOA80_02540	radical SAM protein	109, 15	2	-
AOA80_02545	hypothetical protein	15	2	-
AOA80_02580	DEAD/DEAH box helicase	37	2	-
AOA80_02595	hypothetical protein	6, 28, 57	2	-
AOA80_02615	hypothetical protein	108, 115	3	-

AOA80_02620	hypothetical protein	66, 88, 359,	3	-
AOA80_02705	hypothetical protein	31	2	-
AOA80_02725	nucleic acid-binding protein	48	2	-
AOA80_02805	glutamine amidotransferase	31	2	-
AOA80_02825	hypothetical protein	273	2	-
AOA80_02860	hypothetical protein	28	2	-
AOA80_02925	hypothetical protein	129	2	-
AOA80_02950	hypothetical protein	60	3	-
AOA80_03095	hypothetical protein	186	2	-
AOA80_03125	ABC transporter ATP-binding protein	479	2	-
AOA80_03220	hypothetical protein	7	2	-
AOA80_03250	hypothetical protein	10, 33	2	-
AOA80_03330	hypothetical protein	33, 68	2	-
AOA80_03360	hypothetical protein	51	3	-
AOA80_03380	hypothetical protein	33, 155, 159, 612	2	-
AOA80_03480	hypothetical protein	36	2	-
AOA80_03545	hypothetical protein	47, 56, 128	2	-
AOA80_03575	hypothetical protein	8	2	-
AOA80_03600	hypothetical protein	2	2	-
AOA80_03775	ATPase	25	2	-
AOA80_03820	hypothetical protein	80, 89, 240	2	-
AOA80_03845	glycosyl transferase family 2	8	2	-
AOA80_03915	hypothetical protein	130	2	-
AOA80_03960	hypothetical protein	4, 281	2	-
AOA80_03970	hypothetical protein	2, 59, 71, 76, 80	2	-
AOA80_03975	aspartate aminotransferase	7, 10, 26, 77	2	-
AOA80_04090	hypothetical protein	27, 141	3	-
AOA80_04145	hypothetical protein	66, 162	2	-
AOA80_04210	RNA-binding protein	34	2	-
AOA80_04235	hypothetical protein	16, 23, 251, 335, 350	2	-
AOA80_04240	hypothetical protein	84, 99	2	-
AOA80_04245	hypothetical protein	3	2	-
AOA80_04330	hypothetical protein	61, 84, 125, 142, 143, 287	2	-
AOA80_04335	hypothetical protein	146, 230, 657	2	-
AOA80_04370	hypothetical protein	17	2	-
AOA80_04385	hypothetical protein	19	3	-
AOA80_04460	hypothetical protein	3	3	-
AOA80_04520	hypothetical protein	20	2	-
AOA80_04570	hypothetical protein	17, 51	2	-
AOA80_04645	hypothetical protein	10, 15, 20, 36, 108	3	-
AOA80_04705	hypothetical protein	32	2	-
AOA80_04720	hypothetical protein	6	2	-
AOA80_04765	hypothetical protein	15	3	-
AOA80_04790	hypothetical protein	26	3	-
AOA80_04825	hypothetical protein	28, 119, 127, 170	2	-
AOA80_04835	hypothetical protein	50	2	-
AOA80_04880	hypothetical protein	2	2	-
AOA80_04890	hypothetical protein	6	2	-
AOA80_04905	hypothetical protein	0 (end of contig)	3	-
AOA80_04920	hypothetical protein	57	3	-
AOA80_05005	hypothetical protein	19, 29, 41, 44	2	-
AOA80_05050	hypothetical protein	43	2	-
AOA80_05055	hypothetical protein	265	2	-
AOA80_05060	hypothetical protein	81, 260, 331, 337	2	-
AOA80_05080	phosphoribosylaminoimidazole-succinocarboxamide synthase	6	2	-
AOA80_05090	hypothetical protein	41	3	-
AOA80_05110	hypothetical protein	16	2	-
AOA80_05200	hypothetical protein	25	2	-
AOA80_05220	hypothetical protein	13	2	-
AOA80_05260	hypothetical protein	66, 73	2	-
AOA80_05290	hypothetical protein	4, 81	2	-
AOA80_05300	hypothetical protein	22	2	-

AOA80_05315	hypothetical protein	9	2	-
AOA80_05330	hypothetical protein	15	2	-
AOA80_05360	hypothetical protein	24	2	-
AOA80_05380	hypothetical protein	14, 38, 69	2	-
AOA80_05410	hypothetical protein	50	3	-
AOA80_05490	hypothetical protein	152, 157	2	-
AOA80_05510	hypothetical protein	2	2	-
AOA80_05535	hypothetical protein	15, 39, 56, 65, 67, 71	3	-
AOA80_05545	hypothetical protein	13	2	-
AOA80_05570	hypothetical protein	17	2	-
AOA80_05575	GTPase	8	2	-
AOA80_05580	hypothetical protein	19	2	-
AOA80_05595	hypothetical protein	32, 130	2	-
AOA80_05630	hypothetical protein	21	2	-
AOA80_05670	hypothetical protein	17	2	-
AOA80_05695	hypothetical protein	32, 55, 545	2	-
AOA80_05700	hypothetical protein	3	2	-
AOA80_05705	translation initiation factor IF-2 subunit gamma	17	2	-
AOA80_05740	hypothetical protein	68	2	-
AOA80_05755	hypothetical protein	105, 176, 316, 494, 676, 773	2	-
AOA80_05900	hypothetical protein	13	2	-
AOA80_05920	hypothetical protein	6	2	-
AOA80_05925	hypothetical protein	66	2	-
AOA80_05945	hypothetical protein	54, 159, 334	2	-
AOA80_05965	ATPase	11	2	-
AOA80_05975	hypothetical protein	3, 14	2	-
AOA80_06020	nitrate ABC transporter ATP-binding protein	124	2	-
AOA80_06025	XRE family transcriptional regulator	121	2	-
AOA80_06095	mRNA 3'-end processing factor	2	2	-
AOA80_06170	CO dehydrogenase	1	2	-
AOA80_06205	nitroreductase	27	2	-
AOA80_06260	aspartate aminotransferase	164, 168	2	-
AOA80_06395	hypothetical protein	25	2	-
AOA80_06495	50*S* ribosomal protein L31	3	2	-
AOA80_06520	aromatic acid decarboxylase	43	2	-
AOA80_06585	glyoxalase	17	2	-
AOA80_06600	hypothetical protein	78	2	-
AOA80_06665	hypothetical protein	109	2	-
AOA80_06795	hypothetical protein	34	2	-
AOA80_06800	glutamate synthase	14, 59	2	-
AOA80_06810	hypothetical protein	37	2	-
AOA80_06830	hypothetical protein	4	3	-
AOA80_06885	hypothetical protein	41, 99	3	-
AOA80_06890	hypothetical protein	86, 126, 185	3	-
AOA80_06920	DNA-directed RNA polymerase subunit E	64	2	-
AOA80_06995	hypothetical protein	9, 22	2	-
AOA80_07025	hypothetical protein	74	2	-
AOA80_07045	hypothetical protein	102	2	-
AOA80_07085 (contig end, no start codon)	hypothetical protein	138	3	-
AOA80_07150	hypothetical protein	23	2	-
AOA80_07185	hypothetical protein	113	2	-
AOA80_07210	hypothetical protein	10, 533	2	-
AOA80_07225	ketoisovalerate oxidoreductase	194	2	-
AOA80_07240	hypothetical protein	54	2	-
AOA80_07265	hypothetical protein	6	2	-
AOA80_07310	hypothetical protein	112	2	-
AOA80_07330	hypothetical protein	30, 52	3	-
AOA80_07355	hypothetical protein	94, 118	2	-
AOA80_07360	hypothetical protein	45	2	-
AOA80_07375	hypothetical protein	28	2	-
AOA80_07405	hypothetical protein	226	2	-
AOA80_07515	hypothetical protein	8	3	-
AOA80_07530	hypothetical protein	75	2	-
AOA80_07540	hypothetical protein	21	2	-
AOA80_07545	hypothetical protein	31	3	-
AOA80_07575	tryptophan synthase subunit beta	7	2	-
AOA80_07680	**hypothetical protein (wrongly annotated monomethylamine:corrinoid methyltransferase)**	257	1	AOA80_07675

AOA80_07710	Seryl-tRNA synthetase	46	2	-
AOA80_07720	**hypothetical protein (wrongly annotated monomethylamine:corrinoid methyltransferase)**	257	1	AOA80_07725
AOA80_07740	hypothetical protein	135	2	-
AOA80_07745	hypothetical protein	31	2	-
AOA80_07795	hypothetical protein	296	2	-
AOA80_07855	hypothetical protein	19	2	-
AOA80_07870	hypothetical protein	3	2	-
AOA80_07965	arylsulfatase	14	2	-
AOA80_07970	hypothetical protein	39	2	-
AOA80_07990	hypothetical protein	1	2	-
AOA80_08050	diphthine synthase	22, 144, 196, 211	2	-
AOA80_08055	hypothetical protein	115, 121, 228, 606	2	-
AOA80_08085	hypothetical protein	20	3	-
AOA80_08105	hypothetical protein	29	2	-
AOA80_08155	hypothetical protein	267, 355	2	-
AOA80_08255	hypothetical protein	58	2	-
AOA80_08275	hypothetical protein	44, 47, 60	2	-
AOA80_08315	NAD-dependent dehydratase	103, 109, 435	2	-
AOA80_08375	hypothetical protein	282	2	-
AOA80_08415	hypothetical protein	17, 35	2	-
AOA80_08475	hypothetical protein	11	2	-
AOA80_08520	hypothetical protein	2	2	-
AOA80_08540	hypothetical protein	37, 84, 100	2	-
AOA80_08585	phosphatidylglycerophosphate synthase	6	2	-
AOA80_08590	hypothetical protein	17	2	-
AOA80_08760	cytidylate kinase	134	2	-
AOA80_08765	H/ACA RNA-protein complex component Cbf5p	3	2	-
AOA80_08790	hypothetical protein	243, 260	2	-
AOA80_08810	hypothetical protein	12	2	-
AOA80_08825	hypothetical protein	6, 8	2	-
AOA80_08845	hypothetical protein	55	2	-
AOA80_08870	tRNA pseudouridine synthase	23	2	-
AOA80_08910	hypothetical protein	12	2	-
AOA80_08935	hypothetical protein	40	2	-
AOA80_08945	hypothetical protein	2, 5	3	-
AOA80_09005	hypothetical protein	7	2	-
AOA80_09020	hypothetical protein	14, 93, 105, 124, 160	2	-
AOA80_09050	hypothetical protein	49, 73, 129, 174, 179	2	-
AOA80_09055	hypothetical protein	58	2	-
AOA80_09065	hypothetical protein	1, 45	2	-
AOA80_09075	hypothetical protein	292	2	-
AOA80_09080	hypothetical protein	84	2	-
AOA80_09110	hypothetical protein	17	2	-
AOA80_09155	hypothetical protein	31, 34, 68	3	-
AOA80_09190	hypothetical protein	10	2	-
AOA80_09220	hypothetical protein	99, 127, 241, 434, 453, 559, 598	2	-
AOA80_09225	hypothetical protein	452, 551, 579, 693, 886, 905, 1011, 1050	2	-
AOA80_09250	hypothetical protein	34	2	-
AOA80_09375	site-specific recombinase	194	2	-
AOA80_09405	hypothetical protein	42, 55	2	-
AOA80_09425	XRE family transcriptional regulator	9, 59	2	-
AOA80_09430	hypothetical protein	9, 77, 235	2	-
AOA80_09470	hypothetical protein	261	2	-
AOA80_09520	cob(I)yrinic acid a,c-diamide adenosyltransferase	49	2	-
AOA80_09540	thioredoxin	45, 121	2	-
AOA80_09545	hypothetical protein	58, 119	2	-
AOA80_09635	ABC transporter ATP-binding protein	361	2	-
AOA80_09645	hypothetical protein	16	2	-
AOA80_09665	hypothetical protein	7, 141	2	-
AOA80_09685	hypothetical protein	14	3	-
AOA80_09735	hypothetical protein	8	3	-
AOA80_09760	hypothetical protein	62	2	-

AOA80_09815	hypothetical protein	89, 124, 287, 332, 344, 413	2	-
AOA80_09820	glucose-1-phosphate thymidylyltransferase	12, 31, 84, 143	2	-
AOA80_09850	polyphosphate kinase	82	2	-
AOA80_09960	hypothetical protein	4	2	-
AOA80_09990	hypothetical protein	163	2	-
AOA80_10060	hypothetical protein	21	3	-
AOA80_10120	hypothetical protein	2, 78	3	-
AOA80_10260	hypothetical protein	57	2	-
AOA80_10265	hypothetical protein	24, 105	2	-
AOA80_10320	hypothetical protein	11	2	-
AOA80_10345	hypothetical protein	374	3	-
AOA80_10445	ATP-dependent helicase	47, 195, 219	2	-
AOA80_10450	acylpyruvase	8, 25, 214, 235, 275, 337, 449, 496, 513	2	-
AOA80_10540	hypothetical protein	19, 398	2	-
AOA80_10560	phosphomethylpyrimidine synthase	20, 29	2	-
AOA80_10625	hypothetical protein	278	3	-
AOA80_10710	hypothetical protein	138	2	-
AOA80_10745	hypothetical protein	22	3	-
AOA80_10775	hypothetical protein	23, 24, 34	3	-
AOA80_10780	hypothetical protein	48, 150	2	-
AOA80_10825	hypothetical protein	81, 108	2	-
AOA80_10850	hypothetical protein	86, 117, 122	2	-
AOA80_10870	hypothetical protein	1	3	-
AOA80_10935	hypothetical protein	32	2	-
AOA80_10980	transcriptional regulator	48	2	-
AOA80_11025	hypothetical protein	127, 215	3	-
AOA80_11030	hypothetical protein	88	2	-
AOA80_11185	hypothetical protein	12, 16, 27	2	-
AOA80_11270	ATP-dependent protease	10, 320	2	-
AOA80_11275	hypothetical protein	36	2	-
AOA80_11410	hypothetical protein	40	2	-
AOA80_11420	hypothetical protein	3	2	-
AOA80_11475	hypothetical protein	61	2	-
AOA80_11485	hypothetical protein	9	2	-
AOA80_11535	hypothetical protein	7	3	-

*Distance between amber codon to the next opal or ochre stop codon. # Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.26. Predicted pyrrolysine usage in the RumEn M2 genome

Locus_tag	Predicted gene product	Distance (codons)*	Class#	-
AOA81_00085	YgiQ family radical SAM protein	39	2	-
AOA81_00090	hypothetical protein	1, 11, 14	2	-
AOA81_00120	hypothetical protein	31, 168	3	-
AOA81_00125	hypothetical protein	38	2	-
AOA81_00250	hypothetical protein	19	2	-
AOA81_00255	hypothetical protein	48, 120, 192	2	-
AOA81_00290	hypothetical protein	7	2	-
AOA81_00295	hypothetical protein	79	2	-
AOA81_00375	hypothetical protein	6	3	-
AOA81_00490	hypothetical protein	7	3	-
AOA81_00525	hypothetical protein	8	2	-
AOA81_00535	hypothetical protein	17	2	-
AOA81_00590	hypothetical protein	9	2	-
AOA81_00645	glutamate 5-kinase	22	2	-
AOA81_00660	hypothetical protein	34	2	-
AOA81_00720	hypothetical protein	1	3	-
AOA81_00810	dihydropteroate synthase DHPS	19	2	-
AOA81_00900	hypothetical protein	10, 13, 53	2	-
AOA81_00940	excinuclease ABC subunit C	29	2	-
AOA81_01005	hypothetical protein	88	3	-
AOA81_01210	seryl-tRNA synthetase	44, 136	2	-
AOA81_01275	GTP-binding protein	5, 14	2	-
AOA81_01310	hypothetical protein	7, 24, 51	3	-
AOA81_01500	hypothetical protein	12	2	-
AOA81_01515	hypothetical protein	19, 64, 101, 104	2	-
AOA81_01530	acetolactate synthase	313	3	-
AOA81_01550	hypothetical protein	29	2	-
AOA81_01555	hypothetical protein	12	2	-
AOA81_01560	hypothetical protein	1	3	-
AOA81_01565	hypothetical protein	80	2	-
AOA81_01620	twitching motility protein PilT	109, 356	2	-
AOA81_01625	hypothetical protein	33	2	-
AOA81_01660	transcriptional regulator	14	2	-
AOA81_01680	hypothetical protein	9, 12, 28	2	-
AOA81_01720	hypothetical protein	55	2	-
AOA81_01860	hypothetical protein	41	2	-
AOA81_02180	3-phosphoglycerate kinase	17	2	-
AOA81_02325	ATP-dependent helicase	45, 150	2	-
AOA81_02345	hypothetical protein	2	2	-
AOA81_02470	metal-dependent hydrolase	6	2	-
AOA81_02495	signal recognition particle-docking protein FtsY	4, 22	2	-
AOA81_02560	ATP synthase subunit F	93	2	-
AOA81_02755	hypothetical protein	10	2	-
AOA81_02915	hypothetical protein	2	2	-
AOA81_02995	ATP-dependent DNA ligase	2	2	-
AOA81_03120	hypothetical protein	126	2	-
AOA81_03250	ABC transporter ATP-binding protein	20	2	-
AOA81_03305	endonuclease IV	2	2	-
AOA81_03495	orotidine 5'-phosphate decarboxylase	12, 1178	2	-
AOA81_03525	hypothetical protein	23	2	-
AOA81_03555	modification methylase HemK	121	2	-
AOA81_03565	hypothetical protein	85, 153	2	-
AOA81_03690	hypothetical protein	17, 96, 142, 172	3	-
AOA81_03715	molecular chaperone DnaJ	58	2	-
AOA81_03740	hypothetical protein	92, 154, 176, 220	2	-
AOA81_03750	hypothetical protein	6	2	-
AOA81_03910	geranylgeranyl pyrophosphate synthase	43	2	-
AOA81_03940	hypothetical protein	7	2	-
AOA81_03970	hypothetical protein	17	2	-
AOA81_03975	GTP cyclohydrolase	66, 83, 86	2	-
AOA81_04165	hypothetical protein	24	2	-
AOA81_04185	hypothetical protein	13, 78, 92, 115	2	-
AOA81_04205	arginyl-tRNA synthetase	167	2	-
AOA81_04305	hypothetical protein	28, 102	3	-
AOA81_04350	hypothetical protein	1	3	-
AOA81_04390	hypothetical protein	21	2	-
AOA81_04395	hypothetical protein	52	2	-

AOA81_04425	homoserine dehydrogenase	524	2	-
AOA81_04435	hypothetical protein	6	2	-
AOA81_04525	hypothetical protein	6	2	-
AOA81_04575	hypothetical protein	8	3	-
AOA81_04595	tRNA-modifying protein	15	2	-
AOA81_04625	hypothetical protein	4	3	-
AOA81_04695	hypothetical protein	16, 76, 238	2	-
AOA81_04755	hypothetical protein	12	3	-
AOA81_04910	hypothetical protein	23	2	-
AOA81_04920	hypothetical protein	31, 97	3	-
AOA81_04950	hypothetical protein	14	3	-
AOA81_04985	fructose 1,6-bisphosphatase	222	2	-
AOA81_05000	quinolinate synthetase	4	2	-
AOA81_05070	hypothetical protein	15	3	-
AOA81_05145	ATPase	50	2	-
AOA81_05165	hypothetical protein	211	2	-
AOA81_05220	hypothetical protein	8	2	-
AOA81_05460	hydroxyethylthiazole kinase	48, 64, 94	2	-
AOA81_05560	ubiquinone biosynthesis methyltransferase UbiE	14	2	-
AOA81_05570	hypothetical protein	102, 131	2	-
AOA81_05600	hypothetical protein	20, 23	2	-
AOA81_05605	hypothetical protein	93	2	-
AOA81_05625	topoisomerase I	7	2	-
AOA81_05665	hypothetical protein	27	2	-
AOA81_05710	hypothetical protein	21	3	-
AOA81_05810	hypothetical protein	614	1	-
AOA81_05840	tRNA (pseudouridine-N_1)-methyltransferase	27	2	-
AOA81_05860	hypothetical protein	7	2	-
AOA81_05880	3-hydroxy-3-methylglutaryl-CoA reductase	64, 100, 164, 216, 245	2	-
AOA81_05965	hypothetical protein	18	2	-
AOA81_06015	hypothetical protein	126	2	-
AOA81_06100	hypothetical protein	43	2	-
AOA81_06310	hypothetical protein	6, 20	3	-
AOA81_06320	hypothetical protein	45	2	-
AOA81_06425	hypothetical protein	7	2	-
AOA81_06445	hypothetical protein	16	2	-
AOA81_06635	hypothetical protein	69	3	-
AOA81_06660	hypothetical protein	23, 279	3	-
AOA81_06705	hypothetical protein	9, 32	2	-
AOA81_06715	hypothetical protein	7	2	-
AOA81_06830	hypothetical protein	3	2	-
AOA81_06845	threonine synthase	163	2	-
AOA81_06925	hypothetical protein	13	2	-
AOA81_06935	hypothetical protein	7	2	-
AOA81_06950	hypothetical protein	6	3	-
AOA81_06990	hypothetical protein	2	2	-

*Distance between amber codon to the next opal or ochre stop codon. # Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.27. Predicted pyrrolysine usage in 1R26 genome

Locus_tag	Predicted gene product	Distance (codons)*	Class	Merge with
AUQ37_00065	hypothetical protein	34	3	-
AUQ37_00335	hypothetical protein	77	2	-
AUQ37_00520#	hypothetical protein	58, 111, 177	3	-
AUQ37_04330	**dimethylamine methyltransferase**	112	1	AUQ37_04335
AUQ37_04480#	hypothetical protein	58, 106, 217, 287	3	-
AUQ37_04495	hypothetical protein	50	2	-
AUQ37_04535	hypothetical protein	98	3	-
AUQ37_04740	hypothetical protein	31	3	-
AUQ37_04745	hypothetical protein	54	2	-
AUQ37_04935	hypothetical protein	94	3	-
AUQ37_05005	hypothetical protein	7	3	-
AUQ37_05015	hypothetical protein	43	2	-
AUQ37_05040		19	2	-
AUQ37_05140	hypothetical protein	8	2	-
AUQ37_05225	hypothetical protein	132	2	-
AUQ37_05365	hypothetical protein	276	2	-
AUQ37_05400	hypothetical protein	11, 83	3	-
AUQ37_05455	ArsR family transcriptional regulator	23, 40	2	-
AUQ37_05460	hypothetical protein	87	2	-
AUQ37_05815	hypothetical protein	170	1	AUQ37_05810
AUQ37_05940	hypothetical protein	63	2	-
AUQ37_05965	**dimethylamine methyltransferase**	112	1	AUQ37_05970
AUQ37_05980	**trimethylamine methyltransferase**	0*	1	AUQ37_05975 (pseudo)
AUQ37_06000	**hypothetical protein – wrong annotation, monomethylamine methyltransferase**	259	1	AUQ37_06005
AUQ37_06010	**hypothetical protein – wrong annotation, monomethylamine methyltransferase**	257	1	AUQ37_06015
AUQ37_06190	hypothetical protein	21	3	-
AUQ37_06225	hypothetical protein	62, 80	3	-
AUQ37_06455	hypothetical protein	1	2	-
AUQ37_06480	hypothetical protein	385	1	AUQ37_06485
AUQ37_06510	hypothetical protein	55	2	-
AUQ37_06525	hypothetical protein	0	2	-
AUQ37_06545	hypothetical protein	199	1	AUQ37_06540
AUQ37_06620	hypothetical protein	81, 190	3	-
AUQ37_06625	hypothetical protein	108	3	-
AUQ37_06870	hypothetical protein	2	2	-
AUQ37_07100	hypothetical protein	28	2	-
AUQ37_01045	hypothetical protein	156, 187	3	-
AUQ37_01125	hypothetical protein	55	1	-
AUQ37_07200	hypothetical protein	1	2	-
AUQ37_07335	hypothetical protein	38	3	-
AUQ37_07630	hypothetical protein	30, 57	3	-
AUQ37_07805	hypothetical protein	40	3	-
AUQ37_07950	hypothetical protein	227	2	-
AUQ37_08100	hypothetical protein	221	1	-
AUQ37_08175	AMP kinase	30	2	-
AUQ37_08385	hypothetical protein	23	2	-
AUQ37_01425	hypothetical protein	54, 109	2	-
AUQ37_01585	hypothetical protein	1, 37	2	-
AUQ37_08580	hypothetical protein	24	3	-
AUQ37_08765	hypothetical protein	33	2	-
AUQ37_01855	hypothetical protein	4, 149	2	-
AUQ37_02030	hypothetical protein	3, 28, 133	3	-
AUQ37_02055	ribosome biogenesis protein	42	2	-
AUQ37_02245	hypothetical protein	174	2	-
AUQ37_02300	hypothetical protein	19	3	-
AUQ37_02400	hypothetical protein	44, 71	3	-
AUQ37_02460	hypothetical protein	871	2	-
AUQ37_02465	hypothetical protein	9, 33, 63	2	-
AUQ37_02495	hypothetical protein	164	1	AUQ37_02500
AUQ37_02845	hypothetical protein	72	3	-
AUQ37_03025	hypothetical protein	43, 87	3	-
AUQ37_03075	hypothetical protein	31	2	-
AUQ37_03390	hypothetical protein	8, 14	2	-
AUQ37_03520	hypothetical protein	40	2	-
AUQ37_03530	hypothetical protein	23	2	-
AUQ37_03870	hypothetical protein	1	3	-
AUQ37_03880	hypothetical protein	13	2	-

AUQ37_03985	hypothetical protein	2	3	-
AUQ37_03990	hypothetical protein	18, 54	3	-
AUQ37_04025	hypothetical protein	210	2	-
AUQ37_04140	hypothetical protein	729	2	-
AUQ37_04145	hypothetical protein	12	2	-
AUQ37_04180	hypothetical protein	45	2	-
AUQ37_04220	hypothetical protein	96	3	-
AUQ37_04265	hypothetical protein	291	3	-

*Distance between amber codon to the next opal or ochre stop codon. [#] Class 1: genes which have amber codon read-through and subsequent incorporation of pyrrolysine; Class 2: genes that utilised the amber codon as a stop codon; Class 3: genes with uncertain amber codon usage due to lack of homologous genes. Genes in bold indicate homologous genes with in-frame amber codon exist in other Methanomassiliicoccales

Table A.4.28. Domains of predicted secretome in the Methanomassiliicoccales

Pfam/TIGRfam	ISO4-H5	ISO4-G1	ISO4-G11	BRNA1	RumEn M1	RumEn M2	1R26	Mx1201	Mx1	B10	MpT1
Repeat domains (containing PF13306, Leucine rich repeats (6 copies), PF13754, Bacterial Ig-like domain (group 3) , PF09479, Listeria-Bacteroides repeat domain, PF05345, Putative Ig domain, PF02368, Bacterial Ig-like domain (group 2), TIGR02167: bacterial surface protein 26-residue repeat)	16	5	20	9	4	4	10	4	23	10	7
Lipo-box containing protein											
TIGR03075: PQQ-dependent dehydrogenase, methanol/ethanol family	1	-	-	-	-	-	-	-	-	-	-
PF03306: Alpha-acetolactate decarboxylase	1	-	-	-	-	-	-	-	-	-	-
PF01048: Phosphorylase superfamily	1	-	1	-	-	-	-	-	-	-	-
PF04519: Polymer-forming cytoskeletal	-	1	-	-	-	-	-	-	-	-	-
PF02129: X-Pro dipeptidyl-peptidase (S15 family)	-	1	-	-	-	-	-	-	-	-	-
PF12682: Flavodoxin	-	-	1	-	-	-	-	-	-	-	-
PF13360: PQQ-like domain	-	-	1	-	-	-	-	-	-	-	-
PF13229: Right handed beta helix region	-	-	1	-	-	-	-	-	-	-	-
PF06508, Queuosine biosynthesis protein QueC	-	-	-	-	-	-	1	1	-	-	-
PF01981, Peptidyl-tRNA hydrolase PTH2	-	-	-	-	-	-	-	-	1	-	-
PF01040: UbiA prenyltransferase family	-	-	-	-	1	-	1	-	-	-	-
PF16927: N-terminal 7TM region of histidine kinase	-	-	-	-	1	-	-	-	-	-	-
PF00091: Tubulin/FtsZ family, GTPase domain	-	-	-	-	1	-	-	-	-	-	-
PF01058: NADH ubiquinone oxidoreductase, 20 Kd subunit	-	-	-	-	2	-	-	-	-	-	-
PF07992: Pyridine nucleotide-disulphide oxidoreductase	-	-	-	-	1	-	-	-	-	-	-
PF12697: Alpha/beta hydrolase family	-	-	-	-	1	-	-	-	-	-	-
PF01855: Pyruvate flavodoxin/ferredoxin oxidoreductase, thiamine diP-bdg	-	-	-	-	1	-	-	-	-	-	-
PF06197: Protein of unknown function (DUF998)	-	-	-	-	-	1	-	-	-	-	-
PF07690: Major Facilitator Superfamily	-	-	-	-	-	-	1	-	-	-	-
PF01226: Formate/nitrite transporter	-	-	-	-	-	-	1	-	-	-	-
PF01497: Periplasmic binding protein	-	-	-	-	-	-	1	-	-	-	-
PF00801: PKD domain	-	-	-	-	-	-	1	-	-	-	-
PF00561: alpha/beta hydrolase fold	-	-	-	-	-	-	1	-	-	-	-
Integral membrane protein (3 or more TMH)											
PF00924: Mechanosensitive ion channel	1	1	1	1	1	1	-	-	-	1	1
PF01032: FecCD transport family	2	1	1	1	1	-	1	1	-	-	-
PF07690: Major Facilitator Superfamily	1	2	-	-	-	-	1	2	1	2	-
PF00909: Ammonium Transporter Family	-	1	1	-	-	-	-	-	1	1	-
PF06961: Protein of unknown function	1	-	-	1	-	-	-	-	-	-	-
PF01940: Integral membrane protein DUF92	1	-	-	-	-	-	-	-	-	-	-
PF00892: EamA-like transporter family	1	-	1	-	-	-	-	-	-	-	-
PF00950: ABC 3 transport family	1	1	1	-	-	-	1	-	-	-	-
PF02554: Carbon starvation protein CstA	1	-	1	-	-	-	-	-	-	-	-
PF00999: Sodium/hydrogen exchanger family	1	-	1	-	-	-	-	-	-	1	-
PF01554: MatE: The MatE domain	1	-	-	-	-	-	-	-	-	-	-
PF00893: Small Multidrug Resistance protein	-	1	1	-	1	-	-	1	-	-	-
PF06197: Protein of unknown function (DUF998)	-	1	-	1	1	-	-	-	-	1	-
TIGR01525: heavy metal translocating P-type ATPase	-	1	-	-	-	-	-	-	-	-	-
PF01970: Tripartite tricarboxylate transporter TctA family	-	1	-	-	-	-	-	-	-	-	-
TIGR01511 (Role:96): (Role:145): copper-translocating P-type ATPase	-	1	-	-	-	-	-	-	-	-	-
PF01497: Periplasmic binding protein	-	-	1	-	1	-	-	-	-	-	-
PF01040: UbiA prenyltransferase family	-	-	1	-	-	-	-	-	-	-	-
PF13593: SBF-like CPA transporter family	-	-	1	-	-	-	-	-	-	-	-
PF01595, Domain of unknown function DUF21	-	-	-	1	-	-	-	1	1	-	1
PF00474, Sodium:solute symporter family	-	-	-	1	-	-	-	-	-	-	-
PF02659, Putative manganese efflux pump	-	-	1	1	-	-	1	1	-	1	-
PF01758, Sodium Bile acid symporter family	-	-	-	1	-	1	1	-	-	1	1
PF02386: Cation transport protein	-	-	-	-	1	-	-	-	-	-	-
PF03176, MMPL family	-	-	-	1	-	-	-	-	-	-	-

PF00083, Sugar (and other) transporter	-	-	-	1	-	-	-	-	-	-	-
PF09594: Protein of unknown function (DUF2029)	-	-	-	-	-	1	-	-	-	-	-
PF03186: CobD/Cbib protein	-	-	-	-	1	1	-	-	-	-	-
PF00361: Proton-conducting membrane transporter	-	-	-	-	1	-	-	-	-	-	-
PF13620, Carboxypeptidase regulatory-like domain	-	-	-	-	-	-	-	-	-	-	1
PF01864, Putative integral membrane protein DUF46	-	-	-	-	1	-	-	-	-	-	1
PF13360: PQQ-like domain	-	-	-	-	-	-	-	1	1	-	1
PF01956, Integral membrane protein DUF106	-	-	-	-	-	-	-	1	-	-	-
PF13515, Fusaric acid resistance protein-like	-	-	-	-	-	-	-	1	-	-	-
PF01957, NfeD-like C-terminal, partner-binding	-	-	-	-	-	-	-	-	1	-	-
PF03706, Lysylphosphatidylglycerol synthase TM region	-	-	-	-	-	-	-	-	1	-	-
PF12698, ABC-2 family transporter protein	-	-	-	-	-	-	-	-	1	-	-
PF09858, Predicted membrane protein (DUF2085)	-	-	-	-	-	-	-	-	-	1	-
PF01545, Cation efflux family	-	-	-	-	-	-	-	-	-	2	-
PF01384, Phosphate transporter family	-	-	-	-	-	-	-	-	-	1	-
PF01569, PAP2 superfamily	-	-	-	-	-	-	-	-	-	1	-
PF00482, Type II secretion system (T2SS), protein F	-	-	-	-	-	-	-	-	-	1	-
PF01699, Sodium/calcium exchanger protein	-	-	-	-	1	-	-	-	-	1	-
PF00662, NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus	-	-	-	-	-	-	-	-	-	1	-
PF00146: NADH dehydrogenase	-	-	-	-	1	-	-	-	-	-	-
PF06847: Archaeal Peptidase A24 C-terminus Type II	-	-	-	-	1	-	-	-	-	-	-
PF02517: CAAX protease self-immunity	-	-	-	-	1	-	-	-	-	-	-
PF00702: haloacid dehalogenase-like hydrolase	-	-	-	-	-	-	1	-	-	-	-
PF13630: SdpI/YhfL protein family	-	-	-	-	-	-	1	-	-	-	-
Dual membrane anchored											
PF01957: NfeD-like C-terminal, partner-binding	1	-	-	-	-	-	-	-	-	1	-
TIGR02228: signal peptidase I	1	-	-	-	-	-	-	-	-	-	-
PF00801: PKD domain	1	-	-	2	-	-	1	-	-	-	-
PF04519: Polymer-forming cytoskeletal	-	2	-	-	-	-	-	-	-	-	-
PF01841: Transglutaminase-like superfamily	-	2	-	1	-	-	1	1	-	-	1
PF01497: Periplasmic binding protein	-	-	1	-	-	-	-	1	-	1	-
PF07705: CARDB: Cell adhesion related domain found in bacteria	-	-	1	-	-	-	-	-	-	-	-
PF13620: Carboxypeptidase regulatory-like domain	-	-	1	-	1	-	-	-	1	1	-
PF13091: PLD-like domain	-	-	1	-	-	-	-	1	-	-	1
PF01595, Domain of unknown function DUF21	-	-	-	-	-	1	-	-	-	-	-
PF13360: PQQ-like domain	-	-	-	-	1	-	-	-	1	1	-
PF13240, zinc-ribbon domain	-	-	-	-	-	-	-	-	-	1	-
PF00112, Papain family cysteine protease	-	-	-	-	-	-	-	-	-	1	-
PF07705, Cell adhesion related domain found in bacteria.	-	-	-	-	-	-	-	-	-	1	-
PF01032: FecCD transport family	-	-	-	-	-	-	1	-	-	-	-
C-terminal anchored											
TIGR03300 (Role:97): outer membrane assembly lipoprotein YfgL	2		1	-	-	-	-	-	-	-	-
TIGR02167: bacterial surface protein 26-residue repeat	2	-	-	-	-	-	-	-	-	-	-
PF13091: PLD-like domain	-	1	-	-	-	-	-	1	-	-	-
PF01497: Periplasmic binding protein	-	1	-	-	-	-	-	-	-	-	-
PF13472: GDSL-like Lipase/Acylhydrolase family	-	1	-	-	-	-	-	-	-	1	-
PF13620: Carboxypeptidase regulatory-like domain	-	1	-	-	-	-	-	-	-	-	-
PF01841: Transglutaminase-like superfamily	-	-	1	-	-	-	-	-	-	-	-
PF00150, Cellulase (glycosyl hydrolase family 5)	-	-	-	-	-	-	-	-	1	-	-
PF13360, PQQ-like domain	-	-	-	-	-	-	-	-	-	1	-
PF13229, Right handed beta helix region	-	-	-	-	-	-	-	-	-	1	-
PF12849, PBP superfamily domain	-	-	-	-	1	-	-	-	-	-	-
N-terminal anchored											
PF01497: Periplasmic binding protein	6	7	10	3	2	-	1	-	7	2	1
TIGR01390 (Role:122): 2',3'-cyclic-nucleotide 2'- phosphodiesterase	1	-	-	-	-	-	-	-	-	-	-
TIGR03659: heme ABC transporter, heme-binding protein isdE	1	-	-	-	-	-	-	-	-	-	-

PF00963: Cohesin domain: Cohesin domains interact with a complementary domain	1	-	-	-	-	-	-	-	-	-	-	
PF12669: Virus attachment protein p12 family:	2	-	1	-	-	-	-	-	-	-	-	
PF00112: Papain family cysteine protease	1	-	-	-	-	-	-	-	-	-	-	-
PF00932: Lamin Tail Domain: The lamin-tail domain (LTD), which has an immunoglobulin (Ig) fold	-	1	1	1	-	-	1	-	-	-	-	
PF16656: Purple acid Phosphatase, N-terminal domain	-	-	-	-	1	-	-	-	-	-	-	
PF14478, Domain of unknown function (DUF4430)	-	-	-	-	1	-	-	-	-	-	-	
PF07705: CARDB: Cell adhesion related domain found in bacteria	-	-	-	-	1	-	-	-	-	-	-	
PF05048: Periplasmic copper-binding protein (NosD)	-	-	-	-	-	1	-	-	-	-	-	
PF13502: AsmA-like C-terminal region	-	1	-	-	-	-	-	-	-	-	-	
PF14269: Arylsulfotransferase (ASST)	-	1	-	-	-	-	-	-	-	-	-	
PF02872: 5-nucleotidase	-	1	1	-	-	-	-	-	-	-	-	
PF14262: Domain of unknown function (DUF4353)	-	1	-	-	-	-	-	-	-	-	-	
PF03306: Alpha-acetolactate decarboxylase	-	-	1	-	-	-	-	-	-	-	-	
PF13360: PQQ-like domain	-	-	1	-	-	-	-	-	-	-	-	
PF00496: Bacterial extracellular solute-binding proteins	-	-	1	-	-	-	-	-	-	-	-	
PF01145: SPFH domain / Band 7 family	-	-	1	-	-	-	-	-	-	-	-	
PF03721, UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain	-	-	-	1	-	-	-	-	-	-	-	
PF01048, Phosphorylase superfamily	-	-	-	1	-	-	-	-	-	-	-	
PF00801: PKD domain	-	-	-	-	-	-	1	-	-	-	-	
PF13620: Carboxypeptidase regulatory-like domain	-	-	-	-	-	-	1	-	-	-	-	
PF12682, Flavodoxin	-	-	-	-	-	-	1	-	-	-	-	
PF13379, NMT1-like family	-	-	-	-	1	1	-	-	1	2	1	
PF07843, Protein of unknown function (DUF1634)	-	-	-	-	-	-	-	-	-	-	1	
PF00753, Metallo-beta-lactamase superfamily	-	-	-	-	-	-	-	-	-	-	1	
PF00150, Cellulase (glycosyl hydrolase family 5)	-	-	-	-	-	-	-	-	1	-	-	
PF02608, Basic membrane protein	-	-	-	-	-	-	-	-	1	-	-	
PF12849, PBP superfamily domain	-	-	-	-	-	-	1	-	1	1	-	
PF07602, Protein of unknown function (DUF1565)	-	-	-	-	-	-	-	-	1	-	-	
PF11288, Protein of unknown function (DUF3089)	-	-	-	-	-	-	-	-	1	-	-	
PF12740, Chlorophyllase enzyme	-	-	-	-	-	-	-	-	1	-	-	
PF07790, Protein of unknown function (DUF1628)	-	-	-	-	-	-	-	-	-	2	-	
PF01522, Polysaccharide deacetylase	-	-	-	-	-	-	-	-	-	1	-	
PF00254, FKBP-type peptidyl-prolyl cis-trans isomerase	-	-	-	-	-	-	-	-	-	1	-	
PF01841, Transglutaminase-like superfamily	-	-	-	-	1	-	-	-	-	2	-	
PF11611, Domain of unknown function (DUF4352)	-	-	-	-	-	-	-	-	-	1	-	
PF14478, Domain of unknown function (DUF4430)	-	-	-	-	-	-	-	-	-	1	-	
PF01297, Zinc-uptake complex component A periplasmic	-	-	-	-	-	-	-	-	-	2	-	
PF13531, Bacterial extracellular solute-binding protein	-	-	-	-	-	-	-	-	-	1	-	
PF13229, Right handed beta helix region	-	-	-	-	-	-	-	-	-	1	-	
Secreted												
PF00070: Pyridine nucleotide-disulphide oxidoreductase	1	1	1	1	-	-	-	-	1	1	-	
TIGR01766 (Role:154): transposase, IS605 OrfB family	1	-	-	-	-	-	-	-	-	-	-	
PF00155: Aminotransferase class I and II	1	-	-	-	-	-	-	-	-	-	-	
PF06508, Queuosine biosynthesis protein QueC	1	-	-	1	-	1	1	-	-	-	1	
PF05935: Arylsulfotransferase (ASST)	1	-	-	-	-	-	-	-	-	-	-	
PF03749: Sugar fermentation stimulation protein	1	-	-	-	-	-	-	-	-	-	-	
PF00312: Ribosomal protein S15	1	-	-	-	-	-	-	-	-	-	-	
PF00297: Ribosomal protein L3	1	-	-	-	-	-	-	-	1		1	
PF08069: Ribosomal S13/S15 N-terminal domain	-	-	-	-	-	1	-	-	-	-	-	
PF00297: Ribosomal protein L3	-	-	-	-	1	-	-	-	-	-	-	
PF01656: CobQ/CobB/MinD/ParA nucleotide binding domain	-	-	-	-	-	1	-	-	-	-	-	

PF01862: Pyruvoyl-dependent arginine decarboxylase (PvlArgDC)	-	-	-	-	-	1	-	-	-	-	-
PF13091: PLD-like domain	-	-	-	-	1	1	-	-	-	-	-
PF00112: Peptidase_C1	-	-	-	-	1	-	-	-	-	-	-
PF12849: PBP superfamily domain	-	-	-	-	1	-	-	-	-	-	-
PF02525: Flavodoxin-like fold	-	1	-	-	-	-	-	-	-	-	-
PF00764: Arginosuccinate synthase	-	1	-	-	-	-	-	-	-	-	-
PF01035: 6-*O*-methylguanine DNA methyltransferase, DNA binding domai	-	1	-	-	-	-	-	-	-	-	-
PF02153: Prephenate dehydrogenase	-	-	1	1	-	-	1	1	-	-	-
PF00148: Nitrogenase component 1 type Oxidoreductase	-	-	-	-	-	-	1	-	-	-	-
PF01841: Transglutaminase-like superfamily	-	-	1	-	1	-	-	-	-	-	-
PF13229, Right handed beta helix region	-	-	-	-	1	-	-	-	-	-	-
PF05048: Periplasmic copper-binding protein (NosD)	-	-	-	-	1	-	-	-	-	-	-
PF00275, EPSP synthase (3-phosphoshikimate 1-carboxyvinyltransferase)	-	-	-	-	-	-	-	-	-	-	1
PF08757, CotH protein	-	-	-	-	-	-	-	1	-	-	-
PF00254, FKBP-type peptidyl-prolyl cis-trans isomerase	-	-	-	-	-	-	-	1	-	-	-
PF12838, 4Fe-4S dicluster domain	-	-	-	-	-	-	-	-	1	-	-
PF01613, Flavin reductase like domain	-	-	-	-	-	-	-	-	1	1	-
PF12682, Flavodoxin	-	-	-	-	-	-	-	-	1	-	-
PF08445, FR47-like protein	-	-	-	-	-	-	-	-	1	-	-
PF00801, PKD domain	-	-	-	-	-	-	-	-	-	2	-
PF00753, Metallo-beta-lactamase superfamily	-	-	-	-	-	-	-	-	-	1	-
PF02113: D-Ala-D-Ala carboxypeptidase 3 (S13) family	-	-	-	-	1	-	-	-	-	-	-
PF02558: Ketopantoate reductase PanE/ApbA	-	-	-	-	1	-	-	-	-	-	-
PF01593: Flavin containing amine oxidoreductase	-	-	-	-	1	-	-	-	-	-	-
PF01855: Pyruvate flavodoxin/ferredoxin oxidoreductase, thiamine diP-bdg	-	-	-	-	1	-	-	-	-	-	-
PF00037: 4Fe-4S binding domain	-	-	-	-	1	-	-	-	-	-	-
PF08448: PAS fold	-	-	-	-	1	-	-	-	-	-	-
PF01981, Peptidyl-tRNA hydrolase PTH2	-	-	-	-	1	-	-	-	-	1	-
PF14276: Domain of unknown function (DUF4363)	-	-	-	-	1	-	-	-	-	-	-
PF02469: Fasciclin domain	-	-	-	-	1	-	-	-	-	-	-
PF01497, Periplasmic binding protein	-	-	-	-	-	-	-	-	-	3	-
PF08241, Methyltransferase domain	-	-	-	-	-	-	-	-	-	1	-
PF01041, DegT/DnrJ/EryC1/StrS aminotransferase family	-	-	-	-	-	-	-	-	-	1	-

Prediction cutoff: E-05.

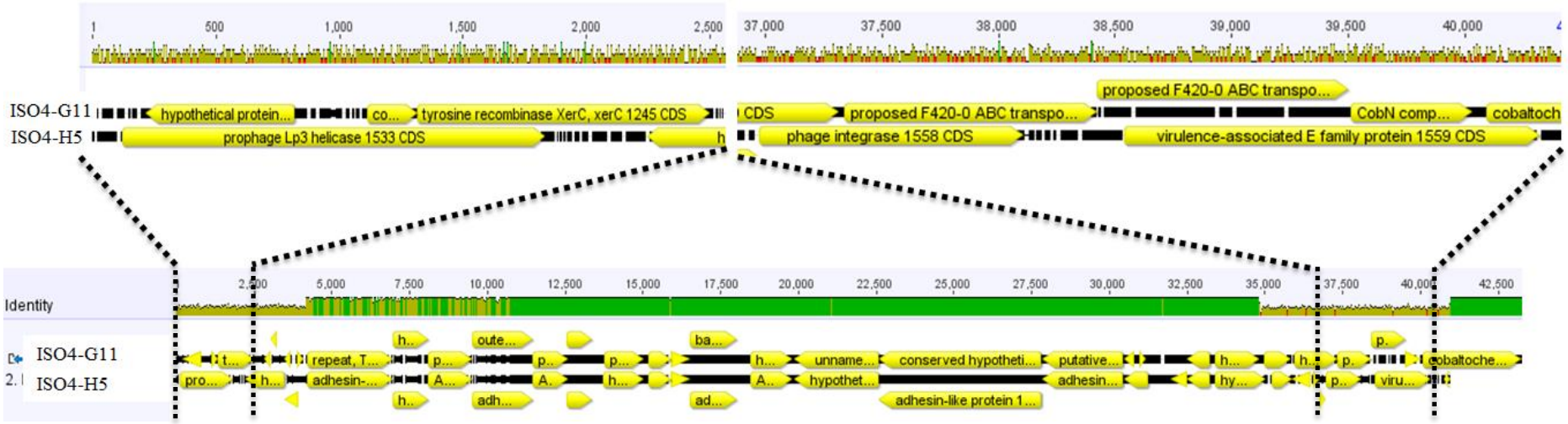


Figure A.4.25. Nucleotide alignment of genome identity region 1 between ISO4-H5 and ISO4-G11. Green area indicates 100% nucleotide identity, olive color indicates 50% nucleotide identity, red color indicates gap in sequence. Yellow block arrow indicates predicted coding sequence. Numbers indicates bases of queried sequence, not base number in genome sequence. Alignment was carried out as described in Section 2.2.26 by Geneious 6.1.5 (Kearse *et al.* 2012).

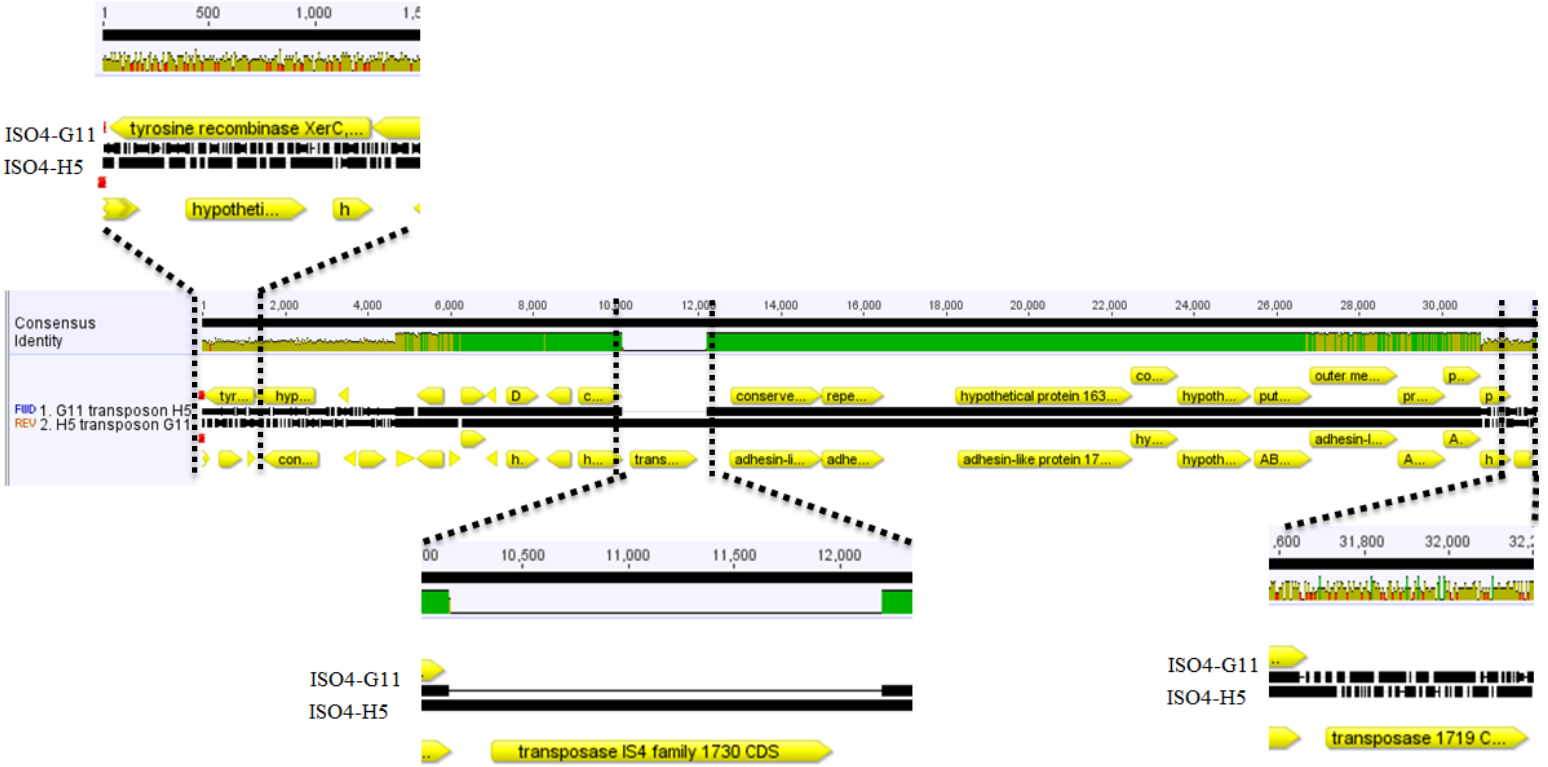


Figure A.4.26 Nucleotide alignment of genome identity region 2 between ISO4-H5 and ISO4-G11. Green area indicates 100% nucleotide identity, olive color indicates 50% nucleotide identity, red color indicates gap in sequence. Yellow block arrow indicates predicted coding sequence. Yellow block arrow indicates predicted coding sequence. Numbers indicates bases of queried sequence, not base number in genome sequence. Alignment carried out in Geneious 6.1.5 (Kearse *et al.* 2012).

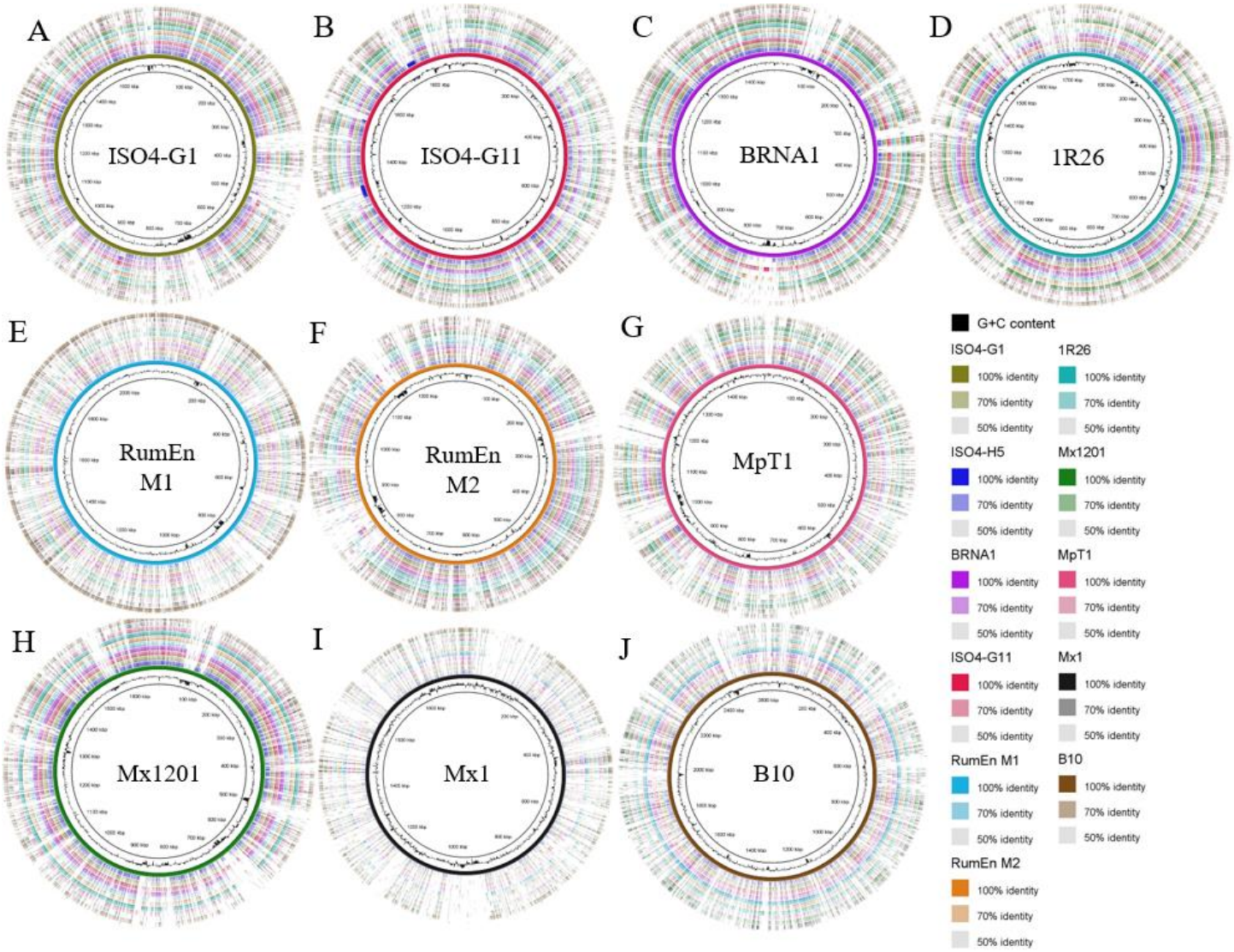


Figure A.4.3. Nucleotide identity comparison of members of the order Methanomassiliicoccales. Methanomassiliicoccales genome nucleotide identity plotted using **A**. ISO4-G1. **B**. ISO4-G11. **C**. BRNA1. **D**. 1R26. **E**. RumEn M1. **F**. RumEn M2. **G**. MpT1. **H.** Mx1201. **I**. Mx1. **J**. B10 as a reference. The colors in each genome ring accord with the identity illustrated in the legend. This image was generated with BRIG (Alikhan *et al.* 2011).

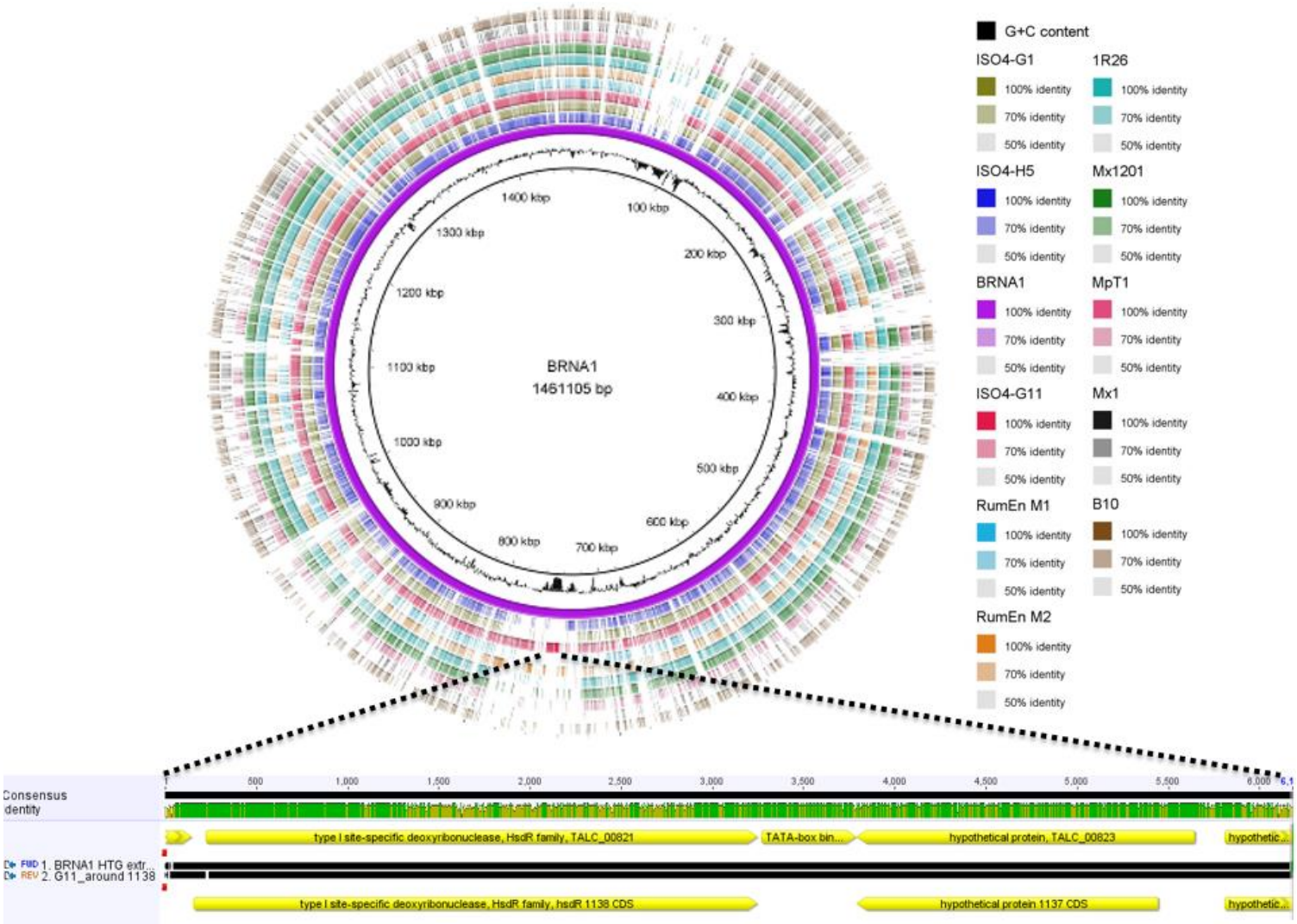


Figure A.4.4. Nucleotide identity region between BRNA1 and ISO4-G11**.** Methanomassiliicoccales genomes plotted using BRNA1 as reference, displaying a region with high nucleotide similarity aligned between BRNA1 and ISO4-G11. The colors in each genome ring accord with the identity illustrated in the legend. This image was generated with BRIG (Alikhan *et al.* 2011).

Table A.4.11. Core genome of ISO4-H5 by functional category

****

Locus_tag	Predicted gene product	COG category
AR505_0001	Orc1/Cdc6 family replication initiation protein	[L]
AR505_0004	acetyl-CoA C-acetyltransferase	[I]
AR505_0006*	signal recognition particle receptor FtsY	[U]
AR505_0013*	ribosomal protein LX	[O]
AR505_0014*	DNA primase small subunit PriA	[G]
AR505_0022*	fumarate hydratase FumA	[C]
AR505_0023*	fumarate hydratase FumB	[C]
AR505_0032	thiamine biosynthesis protein ThiC1	[H]
AR505_0034	DNA gyrase A subunit GyrA	[L]
AR505_0036	hypothetical protein	[S]
AR505_0037*	phosphoribosylformylglycinamidine cyclo-ligase PurM	[F]
AR505_0038*	ribosomal RNA large subunit methyltransferase J RrmJ	[J]
AR505_0039*	ATP-dependent DNA helicase	[R]
AR505_0040	CoB--CoM heterodisulfide reductase subunit D HdrD	[C]
AR505_0041	hypothetical protein	[S]
AR505_0043*	transcriptional regulator ArsR family	[S]
AR505_0044	tRNA 2'-*O*-methylase (pseudo)	[S]
AR505_0047*	thioredoxin TrxA	[O]
AR505_0048	thioredoxin-disulfide reductase TrxB1	[O]
AR505_0049*	XRE family transcriptional regulator	[K]
AR505_0050*	hypothetical protein	[K]
AR505_0051	histone acetyltransferase ELP3 family	[K]
AR505_0052	AAA family ATPase CDC48 subfamily	[O]
AR505_0053*	DNA-directed RNA polymerase subunit H RpoH	[K]
AR505_0054	DNA-directed RNA polymerase subunit B RpoB	[K]
AR505_0055	DNA-directed RNA polymerase subunit A' RpoA1	[K]
AR505_0056	DNA-directed RNA polymerase subunit A" RpoA2	[K]
AR505_0057*	ribosomal protein L30e Rpl30e	[J]
AR505_0058*	transcription elongation factor NusA-like protein	[K]
AR505_0059	radical SAM domain protein	[R]
AR505_0060*	orotidine 5'-phosphate decarboxylase PyrF	[F]
AR505_0066	phosphoribosylamine—glycine ligase PurD	[F]
AR505_0067	hydrogenase accessory protein HypB	[O]
AR505_0068*	glutamyl-tRNA^{Gln} amidotransferase subunit D GatD	[J]
AR505_0069	glutamyl-tRNA^{Gln} amidotransferase subunit E GatE	[J]
AR505_0076	thymidylate synthase ThyX	[F]
AR505_0077*	ribosomal protein S17e Rps17e	[J]
AR505_0080	hydrogenase expression/ formation protein HypE	[O]
AR505_0092	thermosome subunit	[O]
AR505_0094	chaperone protein DnaK	[O]
AR505_0108*	glutamate-5-semialdehyde dehydrogenase ProA	[E]
AR505_0115*	universal archaeal protein Kae1	[O]
AR505_0118*	ribosomal protein S15P Rps15p	[J]
AR505_0120*	fibrillarin	[J]
AR505_0124*	hypothetical protein	Not in COGs
AR505_0129*	ribosomal protein L15e Rpl15e	[J]
AR505_0130	DNA polymerase family B PolB	[L]
AR505_0131	translation-associated GTPase	[J]
AR505_0132*	dimethyladenosine transferase KsgA	[J]
AR505_0133*	conserved hypothetical protein	[J]
AR505_0135*	ribosomal protein L21e Rpl21e	[J]
AR505_0136	hypothetical protein	[J]
AR505_0137	RNA-binding protein	[R]
AR505_0138	serine/threonine protein kinase RIO1 family	[T]
AR505_0139*	translation initiation factor aIF-1A	[J]
AR505_0143*	L-aspartate dehydrogenase	[R]
AR505_0150	2,3-dihydroxy-isovalerate:NADP$^+$ oxidoreductase IlvC	[E]
AR505_0154*	glyceraldehyde-3-phosphate dehydrogenase Gap	[G]
AR505_0155*	phosphoglycerate kinase Pgk	[G]
AR505_0157	LL-diaminopimelate aminotransferase DapC	[E]
AR505_0158*	diaminopimelate epimerase DapF	[E]
AR505_0159	diaminopimelate decarboxylase LysA	[E]
AR505_0161	dihydrodipicolinate synthase DapA	[E]
AR505_0162*	transcriptional regulator AsnC family	[E]
AR505_0167*	conserved hypothetical protein	[S]
AR505_0177	6-phospho 3-hexuloisomerase hxlB	[M]
AR505_0178	endonuclease IV	[L]
AR505_0179*	ribosomal protein L18e Rpl18e	[J]
AR505_0180*	ribosomal protein L13P Rpl13p	[J]
AR505_0181*	ribosomal protein S9P Rps9p	[J]
AR505_0182*	DNA-directed RNA polymerase subunit N RpoN	[K]
AR505_0187*	RdgB/HAM1 family non-canonical purine NTP pyrophosphatase	[F]
AR505_0188*	Kae1-associated kinase Bud32	[T]
AR505_0191*	isopentenyl diphosphate delta-isomerase Fni	[C]
AR505_0194*	hydrogenase assembly chaperone HypC/HupF	[O]
AR505_0195*	alpha-NAC homolog	[K]
AR505_0200*	aspartate--tRNA synthetase AspS1	[J]
AR505_0202	Asp-tRNAAsn/Glu-tRNAGln amidotransferase subunit A GatA	[J]
AR505_0203*	Asp-tRNAAsn/Glu-tRNAGln amidotransferase subunit C GatC	[J]
AR505_0210*	GTP cyclohydrolase RibA	[S]
AR505_0212	CoA-disulfide reductase Cdr	[R]
AR505_0216*	TIGR01210 family protein	[R]
AR505_0218*	ribosomal protein L3P Rpl3p	[J]
AR505_0219*	ribosomal protein L4P Rpl4p	[J]
AR505_0220*	ribosomal protein L23P Rpl23p	[J]
AR505_0221	ribosomal protein L2P Rpl2p	[J]
AR505_0222	ribosomal protein S19P Rps19p	[J]
AR505_0223	ribosomal protein L22P Rpl22p	[J]
AR505_0224	ribosomal protein S3P Rps3p	[J]
AR505_0225*	ribosomal protein L29P Rpl29p	[J]
AR505_0226*	translation initiation factor aSUI1	[J]
AR505_0228*	ribosomal protein S17P Rps17p	[J]
AR505_0229*	ribosomal protein L14e Rpl14e	[J]
AR505_0230*	ribosomal protein L24P Rpl24p	[J]
AR505_0232*	ribosomal protein L5P Rpl5p	[J]
AR505_0233*	ribosomal protein S8P Rps8p	[J]
AR505_0234*	ribosomal protein L6P Rpl6p	[J]
AR505_0235*	ribosomal protein L32e Rpl32e	[J]
AR505_0236*	ribosomal protein L19e Rpl19e	[J]
AR505_0237*	ribosomal protein L18P Rpl18p	[J]
AR505_0238*	ribosomal protein S5P Rps5p	[J]
AR505_0239*	ribosomal protein L30P Rpl30p	[J]
AR505_0240	ribosomal protein L15P Rpl15p	[J]

AR505_0241	preprotein translocase subunit SecY	[U]
AR505_0243*	tRNA pseudouridine synthase B TruB	[J]
AR505_0247	tryptophanyl-tRNA synthetase TrpS	[J]
AR505_0248	phenylalanyl-tRNA synthetase alpha subunit PheS	[J]
AR505_0251	uridylate kinase PyrH	[F]
AR505_0252*	peptide chain release factor aRF1	[J]
AR505_0253*	arginine-tRNA synthetase ArgS	[J]
AR505_0255*	hypothetical transmembrane protein	Not in COGs
AR505_0257*	5-formaminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5'-monophosphate synthetase-like protein	[R]
AR505_0262	adenylate kinase Adk	[F]
AR505_0263*	FAD linked oxidase domain-containing protein	[C]
AR505_0265	phosphoadenosine phosphosulfate reductase	[J]
AR505_0268*	pyruvoyl-dependent arginine decarboxylase PdaD	[S]
AR505_0273*	CoB--CoM heterodisulfide reductase subunit C HdrC	[C]
AR505_0274	CoB--CoM heterodisulfide reductase subunit B HdrB	[C]
AR505_0277*	ribosomal protein S3Ae Rps3ae	[J]
AR505_0283	ribosomal protein L37e Rpl37e	[J]
AR505_0284	amidophosphoribosyltransferase PurF	[F]
AR505_0291	hypothetical protein	[S]
AR505_0292	hypothetical protein	[T]
AR505_0293*	homoserine dehydrogenase MetL	[E]
AR505_0377	NAD synthetase NadE	[H]
AR505_0378	Cobalt-precorrin-5 (C_1)-methyltransferase CbiD	[H]
AR505_0380	precorrin-6x reductase CbiJ	[H]
AR505_0386*	hypothetical protein	[U]
AR505_0387	lysyl-tRNA synthetase LysS	[J]
AR505_0388*	DNA topoisomerase VI subunit A	[L]
AR505_0389	DNA topoisomerase VI subunit B	[L]
AR505_0393	small GTP-binding protein	[R]
AR505_0394	histidyl-tRNA synthetase HisS	[J]
AR505_0405	magnesium chelatase ChlI	[H]
AR505_0412	ABC transporter ATP-binding protein	[V]
AR505_0413	precorrin-2 C20-methyltransferase CbiL	[H]
AR505_0414	precorrin-4 C11-methyltransferase CbiF	[H]
AR505_0415	precorrin-3B C17-methyltransferase CbiH	[H]
AR505_0416	precorrin-8X methylmutase CbiC	[H]
AR505_0418	ABC transporter permease protein	[V]
AR505_0421	aspartate 1-decarboxylase PanD	[H]
AR505_0425*	signal transduction protein with CBS domains	[K]
AR505_0427	phosphoenolpyruvate synthase PpsA1	[G]
AR505_0429	pyruvate Fdx oxidoreductase γ subunit PorC	[C]
AR505_0430*	pyruvate Fdx oxidoreductase Δ subunit PorD	[C]
AR505_0431	pyruvate Fdx oxidoreductase α subunit PorA	[C]
AR505_0432	pyruvate Fdx oxidoreductase β subunit PorB	[C]
AR505_0436*	aspartate carbamoyltransferase regulatory subunit PyrI	[F]
AR505_0437	aspartate carbamoyltransferase PyrB	[F]
AR505_0453*	hypothetical protein	[S]
AR505_0455	isoleucyl-tRNA synthetase IleS	[J]
AR505_0460	flap endonuclease Fen	[L]
AR505_0461	ribulose-phosphate 3-epimerase Rpe	[G]
AR505_0468	ribosomal protein L40e Rpl40e	[J]
AR505_0470	phosphopyruvate hydratase Eno	[G]
AR505_0475*	hydrogenase maturation factor HypF	[O]
AR505_0481*	methylthioadenosine phosphorylase MtnP	[F]
AR505_0485*	PHP domain-containing protein	[R]
AR505_0486*	bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase HisE	[E]
AR505_0487	imidazoleglycerol-phosphate synthase cyclase subunit HisF	[E]
AR505_0488*	imidazoleglycerol-phosphate dehydratase HisB	[E]
AR505_0491	aspartate-semialdehyde dehydrogenase Asd	[E]
AR505_0494*	ATP phosphoribosyltransferase HisG	[E]
AR505_0495*	histidinol-phosphate aminotransferase HisC	[E]
AR505_0496*	imidazole glycerol phosphate synthase glutamine amidotransferase subunit HisH	[E]
AR505_0503*	carbamoyl-phosphate synthase large subunit CarB2	[F]
AR505_0531*	isopropylmalate/isohomocitrate dehydrogenases LeuB	[C]
AR505_0592*	aconitate hydratase 1 AcnA	[C]
AR505_0600*	HTH domain-containing protein	[K]
AR505_0601	hydroxymethylglutaryl-CoA synthase	[I]
AR505_0602	acetyl-CoA acetyltransferase	[I]
AR505_0603*	DNA-binding protein	[R]
AR505_0610*	homoserine kinase ThrB	[E]
AR505_0611*	threonine synthase ThrC	[E]
AR505_0615	ATP-dependent protease S16 family	[O]
AR505_0617	competence/damage-inducible protein CinA	[R]
AR505_0626*	NAD(P)-dependent glycerol-1-phosphate dehydrogenase	[C]
AR505_0637*	archaeosine tRNA-ribosyltransferase TgtA	[J]
AR505_0642*	peptidyl-tRNA hydrolase	[S]
AR505_0643*	thiamine monophosphate synthase ThiE2	[H]
AR505_0671	argininosuccinate lyase ArgH	[E]
AR505_0672*	argininosuccinate synthase ArgG	[E]
AR505_0673*	*N*-acetyl-gamma-glutamyl-phosphate reductase ArgC	[E]
AR505_0674	bifunctional ornithine acetyltransferase/*N*-acetylglutamate synthase protein ArgJ	[E]
AR505_0676*	acetylornithine aminotransferase ArgD	[E]
AR505_0694	homoserine *O*-succinyltransferase MetA	[E]
AR505_0701	ABC transporter ATP-binding protein	[R]
AR505_0709*	methionyl-tRNA synthetase MetG1	[J]
AR505_0740	TATA-box-binding protein Tbp	[K]
AR505_0742	adenylosuccinate lyase PurB	[F]
AR505_0752	nitrate/sulfonate/bicarbonate ABC transporter ATPase NtrD	[P]
AR505_0757*	phosphoserine phosphatase SerB	[S]
AR505_0758	signal recognition particle SRP54 protein	[U]

AR505_0759*	hypothetical protein	[S]
AR505_0767*	phenylalanyl-tRNA synthetase subunit β PheT	[I]
AR505_0768*	hydroxymethylglutaryl-CoA reductase (NADPH) HmgA	[H]
AR505_0776	ribosomal protein S19e Rps19e	[J]
AR505_0777*	DNA-binding protein	[R]
AR505_0778*	ribosomal protein L39e Rpl39e	[J]
AR505_0779*	ribosomal protein L31e Rpl31e	[J]
AR505_0783*	hypothetical protein	[R]
AR505_0786	methanogenesis marker protein 2	[R]
AR505_0799	*O*-acetylserine sulfhydrylase	[E]
AR505_0810*	transcription initiation factor TFIIB Tfb	[K]
AR505_0824*	HPr kinase	[C]
AR505_0826	hypothetical transmembrane protein	[S]
AR505_0827*	wyosine biosynthesis protein TYW1	[C]
AR505_0922*	DNA-directed RNA polymerase subunit K RpoK	[K]
AR505_0927*	SPFH domain / Band 7 family protein	[O]
AR505_0932	fructose-bisphosphate aldolase Fba	[G]
AR505_0947	cobalamin biosynthesis protein CobW	[R]
AR505_0967	Proteasome endopeptidase complex	[O]
AR505_0968	universal archaeal KH-domain/ β-lactamase-domain protein	[R]
AR505_0969	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	[H]
AR505_0972	leucyl-tRNA synthetase LeuS	[J]
AR505_0998	indolepyruvate Fdx oxidoreductase beta subunit IorB	[C]
AR505_0999*	indolepyruvate Fdx oxidoreductase α subunit IorA	[C]
AR505_1009*	phenylacetate-CoA ligase PaaF1	[H]
AR505_1014*	ABC transporter ATP-binding protein	[V]
AR505_1027*	cysteine desulfurase SufS subfamily SufS	[E]
AR505_1029	dihydroorotate dehydrogenase PyrD	[F]
AR505_1055*	site-specific recombinase	[L]
AR505_1073*	histidinol dehydrogenase HisD	[E]
AR505_1120*	aspartokinase beta subunit	[Unclassified]
AR505_1121	glutamyl-tRNA synthetase GltX	[J]
AR505_1142*	DNA primase DnaG	[L]
AR505_1143	ATP-dependent protease	[O]
AR505_1148	triose-phosphate isomerase TpiA	[G]
AR505_1149*	fructose 1,6-bisphosphatase Fbp	[G]
AR505_1150	aspartyl-tRNA synthetase AspS2	[J]
AR505_1151	tryptophan synthase beta subunit TrpB	[R]
AR505_1165*	NTPase	[F]
AR505_1166*	N^2,N^2-dimethylguanosine tRNA methyltransferase Trm1	[J]
AR505_1168	adenylosuccinate synthase PurA	[F]
AR505_1172*	3-octaprenyl-4-hydroxybenzoate decarboxylyase UbiD	[H]
AR505_1180*	3,4-dihydroxy-2-butanone-4-phosphate synthase RibB	[H]
AR505_1181*	FAD synthase	[M]
AR505_1182*	riboflavin synthase RibC	[H]
AR505_1183*	6,7-dimethyl-8-ribityllumazine synthase RibH	[H]
AR505_1191	cysteinyl-tRNA synthetase CysS	[J]
AR505_1199*	alanine aminotransferase	[E]
AR505_1203*	methanogenesis marker protein 8	[S]
AR505_1204*	PP-loop domain-containing protein	[D]
AR505_1209	replication factor C small subunit	[L]
AR505_1217	formate-tetrahydrofolate ligase	[F]
AR505_1222	ferrous iron transport protein B FeoB	[P]
AR505_1233	pseudouridylate synthase	[J]
AR505_1235	peptidyl-prolyl cis-trans isomerase	[O]
AR505_1274	tRNA pseudouridine synthase D TruD	[S]
AR505_1277*	ATP-dependent DNA ligase DnlI	[L]
AR505_1279*	iron-sulfur cluster assembly protein	[S]
AR505_1284	diphthine synthase DphB	[J]
AR505_1287*	deoxycytidine triphosphate deaminase Dcd	[F]
AR505_1299	DNA topoisomerase I TopA	[L]
AR505_1303	asparagine synthase (glutamine-hydrolyzing) AsnB	[E]
AR505_1305*	endonuclease III Nth	[L]
AR505_1307	hydrogenase expression/ formation protein HypD	[O]
AR505_1318	peptidase M18 family	[E]
AR505_1320*	Methyltransferase corrinoid activation protein	[R]
AR505_1321	ATP-dependent RNA helicase	[K]
AR505_1324	pyrrolysine biosynthesis radical SAM protein PylB	[H]
AR505_1325	pyrrolysine--tRNA ligase PylS	[J]
AR505_1328	monomethylamine methyltransferase MtmB	[Unclassified]
AR505_1329	methyltransferase cognate corrinoid proteins	[R]
AR505_1339*	valyl-tRNA synthetase ValS	[J]
AR505_1341*	glycyl-tRNA synthetase GlyS	[J]
AR505_1349	hypothetical protein	[Unclassified]
AR505_1357	gliding motility-associated protein GldE	[R]
AR505_1370*	ribosomal protein S27e Rps27e	[J]
AR505_1373	DNA-directed RNA polymerase subunit E RpoE	[J]
AR505_1374	DNA-directed RNA polymerase RpoE	[K]
AR505_1375*	deoxyhypusine synthase Dys	[O]
AR505_1385*	methanogenesis marker protein 7	[R]
AR505_1387*	methanogenesis marker protein 15	[I]
AR505_1388*	methanogenesis marker protein 5	[S]
AR505_1389*	methanogenesis marker protein 6	[S]
AR505_1391*	methyl-CoM reductase C subunit McrC	[H]
AR505_1392	methyl-CoM reductase component A2 AtwA	[R]
AR505_1396	methyl-CoM reductase α subunit McrA	[H]
AR505_1397	methyl-CoM reductase γ subunit McrG	[H]
AR505_1398*	methyl-CoM reductase operon protein D McrD	[H]
AR505_1399	methyl-CoM reductase β subunit McrB	[H]
AR505_1401*	ornithine carbamoyltransferase ArgF	[E]
AR505_1406	RNA-splicing ligase RtcB	[S]
AR505_1408*	RNA-binding protein	[J]
AR505_1412	cell division ATPase MinD	[D]
AR505_1413	dinitrogenase iron-molybdenum cofactor biosynthesis protein	[S]
AR505_1414	7-cyano-7-deazaguanosine biosynthesis protein QueE	[R]
AR505_1415*	queuosine biosynthesis protein QueC	[H]
AR505_1416	6-pyruvoyl tetrahydropterin synthase QueD	[R]
AR505_1421	ribosomal protein L10e Rpl10e	[J]
AR505_1426*	seryl-tRNA synthetase SerS	[J]
AR505_1431	mevalonate kinase Mvk	[I]
AR505_1433	geranylgeranyl reductase	[C]

AR505_1434*	Fdx	[C]
AR505_1435	molybdenum cofactor biosynthesis protein A MoaA	[R]
AR505_1437	GTPase	[R]
AR505_1438	DNA polymerase II large subunit DP2 PolD2	[L]
AR505_1439*	DNA-directed RNA polymerase subunit D RpoD	[K]
AR505_1440*	ribosomal protein S11P Rps11p	[J]
AR505_1441*	ribosomal protein S4P Rps4p	[J]
AR505_1442*	ribosomal protein S13P Rps13p	[J]
AR505_1447	nitrogenase cofactor biosynthesis protein NifB	[R]
AR505_1448	alanyl-tRNA synthetase AlaS	[J]
AR505_1449	chorismate synthase AroC	[E]
AR505_1453*	ribosomal protein S10P Rps10p	[J]
AR505_1454	translation elongation factor aEF-1 α	[J]
AR505_1455	translation elongation factor aEF-2	[J]
AR505_1456*	ribosomal protein S7P Rps7p	[J]
AR505_1457*	ribosomal protein S12P Rps12p	[J]
AR505_1459*	translation initiation factor aIF-2 α subunit	[J]
AR505_1472*	ribosomal protein S8e Rps8e	[J]
AR505_1475*	phosphoesterase	[R]
AR505_1476*	methyl-viologen-reducing hydrogenase α subunit MvhA	[C]
AR505_1477*	methyl-viologen-reducing hydrogenase γ subunit MvhG	[C]
AR505_1478*	methyl-viologen-reducing hydrogenase Δ subunit	[C]
AR505_1479	CoB--CoM heterodisulfide reductase subunit A HdrA	[C]
AR505_1482*	proteasome alpha subunit PsmA	[O]
AR505_1483	Ribosome maturation protein SBDS	[J]
AR505_1484	exosome complex RNA-binding protein Rrp4	[J]
AR505_1485*	exosome complex exonuclease Rrp41	[J]
AR505_1486	exosome complex RNA-binding protein Rrp42	[J]
AR505_1487*	ribosomal protein L37Ae Rpl37ae	[J]
AR505_1488*	DNA-directed RNA polymerase subunit P RpoP	[K]
AR505_1490*	prefoldin beta subunit PfdB	[O]
AR505_1493*	phosphoesterase DHHA1	[R]
AR505_1494	tyrosyl-tRNA synthetase TyrS	[J]
AR505_1512*	4Fe-4S-binding-domain containing ABC transporter ATP-binding protein	[R]
AR505_1513*	ribose 5-phosphate isomerase A RpiA	[G]
AR505_1514*	ribosomal protein L44e Rpl44e	[J]
AR505_1515*	ribosomal protein S27e Rps27e	[J]
AR505_1516	translation initiation factor aIF-2 α subunit	[J]
AR505_1581	Cytosine deaminase	[F]
AR505_1585*	DNA repair and recombination protein RadA	[L]
AR505_1587	prenyltransferase UbiA	[H]
AR505_1588	geranylgeranylglyceryl phosphate synthase GGGPS	[R]
AR505_1589*	hypothetical protein	[Unclassified]
AR505_1602*	Bifunctional diaminohydroxy-phosphoribosylaminopyrimidine deaminase / 5-amino-6-(5-phosphoribosylamino)uracil reductase RibD	[H]
AR505_1603	translation initiation factor aIF-2 β subunit	[J]
AR505_1605	C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	[J]
AR505_1606	translation initiation factor aIF-5A	[J]
AR505_1608*	phosphoglucomutase	[G]
AR505_1614*	uracil-DNA glycosylase Ung	[L]
AR505_1615	orotate phosphoribosyltransferase PyrE	[F]
AR505_1616*	hypothetical transmembrane protein	[I]
AR505_1617*	DNA/RNA-binding protein	[R]
AR505_1623*	$F_{420}H_2$ dehydrogenase subunit M FpoM	[C]
AR505_1625*	$F_{420}H_2$ dehydrogenase subunit K FpoK	[C]
AR505_1628*	$F_{420}H_2$ dehydrogenase subunit I FpoI	[C]
AR505_1630*	$F_{420}H_2$ dehydrogenase subunit D FpoD	[C]
AR505_1632*	$F_{420}H_2$ dehydrogenase subunit B FpoB	[C]
AR505_1635*	Fe-S oxidoreductase	[C]
AR505_1636*	undecaprenyl pyrophosphate synthetase UppS	[I]
AR505_1637*	methanogenesis marker protein 11	[R]
AR505_1638	RNA 2'-phosphotransferase Tpt1/KptA	[J]
AR505_1639	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase FolD	[H]
AR505_1641	AMP phosphorylase DeoA	[F]
AR505_1642*	ribulose bisphosphate carboxylase RbcL	[G]
AR505_1643*	ribose-1,5-bisphosphate isomerase E2b2	[J]
AR505_1647*	transaldolase	[G]
AR505_1648*	transketolase subunit B	[G]
AR505_1649*	transketolase subunit A	[G]
AR505_1650*	proliferating cell nuclear antigen PcnA	[L]
AR505_1655	hypothetical protein	[S]
AR505_1656	bifunctional phosphoribosylamino imidazolecarboxamide formyltransferase/IMP cyclohydrolase PurH	[F]
AR505_1658	exosome complex RNA-binding protein Csl4	[J]
AR505_1660	thiamine biosynthesis ATP pyrophosphatase ThiI	[H]
AR505_1661*	hypothetical protein	[Unclassified]
AR505_1663	hypothetical protein	[Unclassified]
AR505_1664*	phosphoglycerate dehydrogenase SerA	[E]
AR505_1665	Phosphoserine aminotransferase	[E]
AR505_1666	aspartate aminotransferase	[E]
AR505_1667	5-formaminoimidazole-4-carboxamide-1-β-D- ribofuranosyl 5'-monophosphate-formate ligase PurP	[R]
AR505_1671	ribosomal protein S2P Rps2p	[J]
AR505_1676*	GTPase	[R]
AR505_1679*	riboflavin kinase	[K]
AR505_1680*	phosphoribosylformylglycinamidine synthase I PurQ	[F]
AR505_1681	phosphoribosylformylglycinamidine synthase II PurL	[F]
AR505_1682*	phosphoribosylformylglycinamidine synthase PurS	[F]
AR505_1683*	tRNA methyltransferase subunit	[J]
AR505_1684*	prolyl-tRNA synthetase ProS	[J]
AR505_1685*	ribose-phosphate diphosphokinase Prs	[F]

AR505_1687	hypothetical protein Ta0304	[S]
AR505_1695*	Metal-dependent hydrolases β-lactamase family	[H]
AR505_1751*	hypothetical protein	Not in COGs
AR505_1752*	translation initiation factor aIF-2 γ subunit	[R]
AR505_1753*	ribosomal protein S6e Rps6e	[J]
AR505_1755	translation initiation factor aIF-2	[J]
AR505_1756*	nucleoside diphosphate kinase Ndk	[J]
AR505_1757	ribosomal protein L24e Rpl24e	[J]
AR505_1758*	ribosomal protein S28e Rps28e	[J]
AR505_1759*	ribosomal protein L7Ae Rpl7ae	[J]
AR505_1780	malate dehydrogenase Mdh	[C]
AR505_1781	Fe-S oxidoreductase	[C]
AR505_1783*	CDP-diacylglycerol-serine *O*-phosphatidyltransferase PssA	[I]
AR505_1788	adenosylhomocysteinase AhcY2	[H]
AR505_1789*	hypothetical protein	[S]
AR505_1791	RNA methyltransferase TrmH family	[J]
AR505_1794*	ribosomal protein L12P Rpl12p	[J]
AR505_1795*	acidic ribosomal protein P0 RplPO	[J]
AR505_1796*	ribosomal protein L1P Rpl1p	[J]
AR505_1797*	ribosomal protein L11P Rpl11p	[J]
AR505_1799*	protein translocase subunit Sss1	[U]
AR505_1800	hypothetical protein	[Unclassified]
AR505_1801*	hypothetical protein	[S]
AR505_1802	GMP synthase subunit A GuaA	[F]
AR505_1803	GMP synthase subunit B GuaAb	[F]
AR505_1804*	phosphoribosylaminoimidazole carboxylase PurE	[F]
AR505_1811*	transcriptional regulator	[K]
AR505_1818*	A_1A_0-type ATP synthase subunit D	[C]
AR505_1819*	A_1A_0-type ATP synthase subunit B	[C]
AR505_1820*	A_1A_0-type ATP synthase subunit A	[C]
AR505_1821*	A_1A_0-type ATP synthase subunit F	[C]
AR505_1824*	A_1A_0-type ATP synthase subunit K	[C]
AR505_1827*	CMP/dCMP deaminase	[U]

#only one gene from each gene family is represented. *represents gene families found conserved across 11 genomes analysed, including the incomplete genomes

Table A.4.12. Gene families unique to ISO4-H5

****

Locus_tag	Predicted gene product	COG category
AR505_0002	hypothetical transmembrane protein	Not in COGs
AR505_0005	adhesin-like protein	[S]
AR505_0023	hypothetical transmembrane protein	Not in COGs
AR505_0061	adhesin-like protein	[R]
AR505_0062	hypothetical protein	Not in COGs
AR505_0063	hypothetical protein	Not in COGs
AR505_0083	hypothetical protein	Not in COGs
AR505_0088	hypothetical transmembrane protein	Not in COGs
AR505_0090	hypothetical protein	Not in COGs
AR505_0098	pyridoxamine 5'-phosphate oxidase-related protein	[R]
AR505_0106	transposase	[L]
AR505_0107	transposase	[L]
AR505_0110	methyltransferase cognate corrinoid protein	[R]
AR505_0114	methylase involved in ubiquinone/ menaquinone biosynthesis	[R]
AR505_0116	hypothetical protein	Not in COGs
AR505_0123	phosphoglucosamine mutase GlmM	[G]
AR505_0148	hypothetical protein	Not in COGs
AR505_0156	hypothetical protein	Not in COGs
AR505_0164	peptidase A24B	[O]
AR505_0170	hypothetical transmembrane protein	[M]
AR505_0171	hypothetical transmembrane protein	Not in COGs
AR505_0174	hypothetical protein	Not in COGs
AR505_0175	hypothetical protein	[O]
AR505_0176	hypothetical protein	Not in COGs
AR505_0183	hypothetical protein	Not in COGs
AR505_0185	hypothetical transmembrane protein	Not in COGs
AR505_0186	nitroreductase family protein	[C]
AR505_0207	nitrilase/cyanide hydratase and apolipoprotein *N*-acyltransferase	[R]
AR505_0213	transposase IS116/IS110/IS902 family	[L]
AR505_0214	hypothetical protein	Not in COGs
AR505_0256	hypothetical transmembrane protein	Not in COGs
AR505_0275	TPR repeat-containing protein	[R]
AR505_0279	hypothetical protein	Not in COGs
AR505_0285	Na^+-dependent transporter SNF family	[R]
AR505_0288	hypothetical transmembrane protein	Not in COGs
AR505_0288	NADPH-dependent FMN reductase	[R]
AR505_0296	MatE efflux family protein	[V]
AR505_0297	hypothetical protein	Not in COGs
AR505_0298	Cob/MinD domain containing protein	[D]
AR505_0299	cobalamin biosynthesis protein CbiX	[S]
AR505_0300	hypothetical protein	Not in COGs
AR505_0301	adenine phosphoribosyltransferase Apt	[F]
AR505_0302	hypothetical transmembrane protein	Not in COGs
AR505_0303	ABC transporter solute binding protein	[P]
AR505_0306	hypothetical protein	[P]
AR505_0307	SAM-dependent methyltransferase	[P]
AR505_0310	ABC transporter solute-binding protein	Not in COGs
AR505_0312	ubiquinone/menaquinone biosynthesis methyltransferase UbiE	[R]
AR505_0313	phage integrase	[L]
AR505_0315	hypothetical protein	Not in COGs
AR505_0316	hypothetical protein	Not in COGs
AR505_0318	hypothetical protein	Not in COGs
AR505_0319	hypothetical protein	Not in COGs
AR505_0324	hypothetical protein	Not in COGs
AR505_0325	hypothetical protein	Not in COGs
AR505_0326	hypothetical protein	Not in COGs
AR505_0327	DNA-cytosine methyltransferase Dcm	[L]
AR505_0328	hypothetical protein	[Unclassified]
AR505_0329	DNA mismatch endonuclease Vsr	[L]
AR505_0330	conserved hypothetical protein	[S]
AR505_0331	hypothetical protein	Not in COGs
AR505_0332	MORN repeat-containing protein	[S]
AR505_0333	hypothetical protein	Not in COGs
AR505_0334	hypothetical protein	Not in COGs
AR505_0335	hypothetical protein	Not in COGs
AR505_0336	conserved hypothetical protein	Not in COGs

AR505_0337	hypothetical protein	Not in COGs
AR505_0338	hypothetical protein	Not in COGs
AR505_0339	DNA-cytosine methyltransferase Dcm	[L]
AR505_0340	DNA-cytosine methyltransferase Dcm	[L]
AR505_0341	hypothetical protein	Not in COGs
AR505_0342	hypothetical protein	Not in COGs
AR505_0343	hypothetical protein	Not in COGs
AR505_0344	hypothetical protein	Not in COGs
AR505_0345	hypothetical protein	Not in COGs
AR505_0346	hypothetical protein	Not in COGs
AR505_0347	hypothetical transmembrane protein	Not in COGs
AR505_0348	hypothetical transmembrane protein	Not in COGs
AR505_0349	hypothetical transmembrane protein	Not in COGs
AR505_0350	hypothetical transmembrane protein	Not in COGs
AR505_0351	hypothetical transmembrane protein	Not in COGs
AR505_0352	hypothetical transmembrane protein	[R]
AR505_0355	adhesin-like protein	Not in COGs
AR505_0356	hypothetical protein	Not in COGs
AR505_0357	hypothetical protein	Not in COGs
AR505_0358	hypothetical protein	Not in COGs
AR505_0365	hypothetical transmembrane protein	Not in COGs
AR505_0366	TPR repeat-containing protein	[R]
AR505_0370	TPR repeat-containing protein	[R]
AR505_0374	DNA polymerase III PolC	[L]
AR505_0376	*N*-carbamoyl-D-amino acid amidohydrolase AguB	[R]
AR505_0384	hypothetical protein	Not in COGs
AR505_0390	hypothetical transmembrane protein	Not in COGs
AR505_0392	hypothetical transmembrane protein	[S]
AR505_0395	hypothetical protein	[G]
AR505_0396	hypothetical protein	Not in COGs
AR505_0399	hypothetical protein	[J]
AR505_0400	hypothetical protein	[H]
AR505_0407	adhesin-like protein	Not in COGs
AR505_0419	Sel1 repeat-containing protein	[R]
AR505_0424	hypothetical transmembrane protein	Not in COGs
AR505_0426	ATPase (AAA+ superfamily)	[R]
AR505_0428	HTH domain-containing protein	[K]
AR505_0439	hypothetical transmembrane protein	Not in COGs
AR505_0440	transposase IS4 family	[L]
AR505_0441	hypothetical transmembrane protein	[S]
AR505_0442	hypothetical protein	Not in COGs
AR505_0443	universal stress protein UspA	[T]
AR505_0445	cell division control protein Cdc48	[O]
AR505_0446	flavodoxin	[C]
AR505_0448	6-*O*-methylguanine DNA methyltransferase Ogt	[L]
AR505_0449	aminoacyl-histidine dipeptidase PepD	[E]
AR505_0450	hypothetical protein	Not in COGs
AR505_0456	transporter DMT family	[E]
AR505_0457	transcriptional regulator LysR family	[K]
AR505_0459	DNase TatD family	[L]
AR505_0462	transcriptional regulator ArsR family	[K]
AR505_0469	hypothetical protein	Not in COGs
AR505_0476	ABC transporter solute-binding protein	[P]
AR505_0480	hypothetical transmembrane protein	Not in COGs
AR505_0492	transcription regulator	[Q]
AR505_0493	HipA-like protein	[Unclassified]
AR505_0502	hypothetical protein	Not in COGs
AR505_0514	hypothetical protein	[E]
AR505_0519	hypothetical transmembrane protein	Not in COGs
AR505_0522	ATPase (AAA+ superfamily)	[R]
AR505_0523	transposase IS605 OrfB family	[L]
AR505_0527	Fic family protein	[S]
AR505_0529	hypothetical protein	Not in COGs
AR505_0533	hypothetical protein	[R]
AR505_0534	hypothetical protein	[Unclassified]
AR505_0535	hypothetical protein	Not in COGs
AR505_0537	SAM-dependent methyltransferase	[R]
AR505_0538	hypothetical transmembrane protein	[S]
AR505_0539	polysaccharide biosynthesis protein	[R]
AR505_0544	phosphoglycolate/pyridoxal phosphate phosphatase family	[G]
AR505_0545	4-hydroxy-2-oxovalerate aldolase DmpG	[E]
AR505_0550	glycosyl transferase GT2 family	[R]
AR505_0553	dTDP-glucose 4,6-dehydratase RfbB	[M]
AR505_0554	dTDP-4-dehydrorhamnose 3,5-epimerase RfbC	[M]
AR505_0555	polysaccharide/polyol phosphate ABC transporter permease protein	[M]
AR505_0556	polysaccharide/polyol phosphate ABC transporter ATP-binding protein	[M]
AR505_0557	TPR repeat-containing protein	[M]
AR505_0558	LPS biosynthesis protein LICD family	[M]
AR505_0559	glycosyl transferase GT8 family	[M]
AR505_0561	glycosyl transferase GT2 family	[M]
AR505_0562	hypothetical protein	[Unclassified]
AR505_0563	hypothetical protein	[S]
AR505_0564	hypothetical protein	Not in COGs
AR505_0565	hypothetical protein	Not in COGs
AR505_0566	hypothetical protein	[G]
AR505_0567	hypothetical protein	[Unclassified]
AR505_0568	hypothetical protein	[Unclassified]
AR505_0570	hypothetical protein	Not in COGs
AR505_0571	hypothetical protein	Not in COGs
AR505_0572	hypothetical transmembrane protein	Not in COGs
AR505_0573	hypothetical protein	Not in COGs
AR505_0574	transposase	[L]
AR505_0575	hypothetical transmembrane protein	[Unclassified]
AR505_0578	ribonuclease III Rnc	[T]
AR505_0581	hypothetical protein	Not in COGs
AR505_0582	hypothetical protein	Not in COGs
AR505_0583	helicase SNF2 family	[L]
AR505_0584	hypothetical protein	[S]
AR505_0585	hypothetical protein	Not in COGs
AR505_0586	hypothetical protein	Not in COGs
AR505_0587	ATPase (AAA+ superfamily)	[R]
AR505_0588	hypothetical protein	[Unclassified]
AR505_0589	hypothetical transmembrane protein	Not in COGs
AR505_0590	hypothetical transmembrane protein	Not in COGs
AR505_0591	nitroreductase family protein	[C]
AR505_0593	hypothetical transmembrane protein	Not in COGs
AR505_0594	adhesin-like protein	Not in COGs
AR505_0599	transcriptional regulator	[K]
AR505_0604	phosphoglycolate phosphatase	[R]

AR505_0605	ABC transporter permease protein	[V]
AR505_0606	ABC transporter ATP-binding protein	[V]
AR505_0607	hypothetical transmembrane protein	[R]
AR505_0612	Fic family protein	Not in COGs
AR505_0614	adhesin-like protein	[H]
AR505_0621	hypothetical transmembrane protein	Not in COGs
AR505_0623	hypothetical protein	Not in COGs
AR505_0624	protein-tyrosine phosphatase	[T]
AR505_0627	NADPH-dependent FMN reductase	[R]
AR505_0628	hypothetical protein	Not in COGs
AR505_0636	pap2 family protein	[I]
AR505_0646	adhesin-like protein	Not in COGs
AR505_0647	hypothetical protein	Not in COGs
AR505_0648	hypothetical protein	Not in COGs
AR505_0649	hypothetical protein	Not in COGs
AR505_0654	adhesin-like protein	Not in COGs
AR505_0656	hypothetical transmembrane protein	Not in COGs
AR505_0657	adhesin-like protein	Not in COGs
AR505_0658	adhesin-like protein	Not in COGs
AR505_0665	hypothetical protein	Not in COGs
AR505_0666	adhesin-like protein	[Unclassified]
AR505_0670	adhesin-like protein	Not in COGs
AR505_0677	transposase IS605 OrfB family	[L]
AR505_0682	hypothetical transmembrane protein	Not in COGs
AR505_0683	hypothetical transmembrane protein	[D]
AR505_0684	hypothetical protein	Not in COGs
AR505_0686	hypothetical protein	Not in COGs
AR505_0687	hypothetical protein	[L]
AR505_0702	cysteine proteinase	[O]
AR505_0704	adhesin-like protein	[S]
AR505_0714	ABC transporter permease protein	[P]
AR505_0715	ABC transporter solute-binding protein	[P]
AR505_0716	cobalamin (vitamin B12) biosynthesis protein CbiX	[S]
AR505_0722	von Willebrand factor type A domain protein	[Unclassified]
AR505_0723	hypothetical transmembrane protein	Not in COGs
AR505_0727	hypothetical protein	Not in COGs
AR505_0729	bifunctional protein FolC	[H]
AR505_0730	hypothetical transmembrane protein	[R]
AR505_0731	hypothetical transmembrane protein	Not in COGs
AR505_0739	MFS transporter	[G]
AR505_0741	hypothetical protein	[S]
AR505_0743	hypothetical transmembrane protein	Not in COGs
AR505_0744	adhesin-like protein	Not in COGs
AR505_0745	hypothetical protein	Not in COGs
AR505_0746	mechanosensitive ion channel protein	[M]
AR505_0754	nitrate/sulfonate/bicarbonate ABC transporter substrate-binding protein	[P]
AR505_0775	NADPH-dependent FMN reductase	[R]
AR505_0782	hypothetical transmembrane protein	[S]
AR505_0787	hypothetical protein	Not in COGs
AR505_0788	signal transduction histidine kinase	[T]
AR505_0789	hypothetical protein	Not in COGs
AR505_0790	transcriptional regulator LytR family	[T]
AR505_0791	transcriptional regulator LytS family	[T]
AR505_0792	hypothetical protein	Not in COGs
AR505_0793	TPR repeat-containing protein	[R]
AR505_0794	hypothetical protein	Not in COGs
AR505_0795	TPR repeat-containing protein	[R]
AR505_0796	isochorismatase hydrolase	[Q]
AR505_0797	transcriptional regulator HxlR family	Not in COGs
AR505_0798	ssDNA exonuclease RecJ	[L]
AR505_0803	hypothetical protein	Not in COGs
AR505_0804	hypothetical protein	Not in COGs
AR505_0805	ATP-dependent DNA helicase	[L]
AR505_0807	adhesin-like protein	[R]
AR505_0808	hypothetical protein	Not in COGs
AR505_0811	hypothetical protein	Not in COGs
AR505_0813	hypothetical protein	[L]
AR505_0816	hypothetical transmembrane protein	[R]
AR505_0817	hypothetical protein	[S]
AR505_0818	hypothetical protein	Not in COGs
AR505_0820	hypothetical protein	Not in COGs
AR505_0821	hypothetical protein	Not in COGs
AR505_0823	hypothetical transmembrane protein	Not in COGs
AR505_0829	hypothetical transmembrane protein	[Unclassified]
AR505_0833	hypothetical protein	Not in COGs
AR505_0843	transcriptional regulator	[K]
AR505_0844	GNAT family acetyltransferase	Not in COGs
AR505_0846	hypothetical protein	Not in COGs
AR505_0848	hypothetical protein	Not in COGs
AR505_0849	hypothetical protein	[R]
AR505_0850	hypothetical protein	Not in COGs
AR505_0851	adhesin-like protein	[R]
AR505_0852	hypothetical protein	Not in COGs
AR505_0853	hypothetical protein	Not in COGs
AR505_0854	cell division control protein Cdc48	[O]
AR505_0855	hypothetical protein	Not in COGs
AR505_0856	hypothetical transmembrane protein	Not in COGs
AR505_0858	addiction module toxin, RelE/ StbE family	[R]
AR505_0859	transcriptional regulator TetR family	[K]
AR505_0861	hypothetical protein	Not in COGs
AR505_0862	hypothetical protein	Not in COGs
AR505_0863	ATPase	[R]
AR505_0864	hypothetical protein	Not in COGs
AR505_0865	hypothetical protein	Not in COGs
AR505_0866	hypothetical protein	Not in COGs
AR505_0867	hypothetical protein	Not in COGs
AR505_0868	hypothetical protein	Not in COGs
AR505_0869	hypothetical protein	Not in COGs
AR505_0871	hypothetical protein	Not in COGs
AR505_0872	hypothetical protein	[Unclassified]
AR505_0873	hypothetical protein	[Unclassified]
AR505_0874	adhesin-like protein	[R]
AR505_0876	hypothetical protein	[Unclassified]
AR505_0877	hypothetical protein	Not in COGs
AR505_0878	hypothetical protein	Not in COGs
AR505_0879	hypothetical protein	Not in COGs
AR505_0880	hypothetical protein	Not in COGs
AR505_0882	hypothetical protein	Not in COGs
AR505_0883	hypothetical protein	Not in COGs
AR505_0884	sulfurtransferase DndC	[H]
AR505_0885	DNA sulfur modification protein DndD	[L]
AR505_0886	type III restriction endonuclease	[L]
AR505_0887	hypothetical protein	Not in COGs
AR505_0888	hypothetical protein	[J]
AR505_0889	hypothetical protein	Not in COGs
AR505_0890	SMC domain-containing protein	[R]
AR505_0891	SMC domain-containing protein	[R]

AR505_0892	geranylgeranyl reductase family protein	[C]
AR505_0894	hypothetical protein	[O]
AR505_0895	ATPase	[R]
AR505_0896	hypothetical protein	Not in COGs
AR505_0900	ATPase	[R]
AR505_0904	GNAT family acetyltransferases	[R]
AR505_0905	EamA-like transporter family	[R]
AR505_0906	hypothetical protein	Not in COGs
AR505_0909	hypothetical transmembrane protein	[R]
AR505_0910	hypothetical protein	Not in COGs
AR505_0912	hypothetical transmembrane protein	Not in COGs
AR505_0914	hypothetical protein	Not in COGs
AR505_0915	NADPH-dependent FMN reductase	[R]
AR505_0916	hypothetical protein	Not in COGs
AR505_0917	TPR repeat-containing protein	[R]
AR505_0918	CDP-alcohol phosphatidyltransferase	[I]
AR505_0919	proline-specific peptidases	[R]
AR505_0923	hypothetical transmembrane protein	Not in COGs
AR505_0925	hypothetical protein	Not in COGs
AR505_0928	acyl carrier protein phosphodiesterase AcpD	[I]
AR505_0929	hypothetical protein	Not in COGs
AR505_0931	phage integrase	[L]
AR505_0934	hypothetical transmembrane protein	Not in COGs
AR505_0936	hypothetical protein	Not in COGs
AR505_0940	hypothetical transmembrane protein	Not in COGs
AR505_0945	hypothetical protein	Not in COGs
AR505_0946	hypothetical protein	Not in COGs
AR505_0948	hypothetical protein	Not in COGs
AR505_0954	hypothetical protein	[Unclassified]
AR505_0959	hypothetical protein	Not in COGs
AR505_0961	hypothetical protein	Not in COGs
AR505_0965	hypothetical protein	[Unclassified]
AR505_0971	hypothetical protein	Not in COGs
AR505_0973	SAM-dependent methyltransferase	[R]
AR505_0974	hypothetical protein	Not in COGs
AR505_0977	hypothetical protein	Not in COGs
AR505_0979	HTH domain-containing protein	[K]
AR505_0985	adhesin-like protein	Not in COGs
AR505_0986	signal peptidase I	Not in COGs
AR505_0987	hypothetical protein	Not in COGs
AR505_0988	hypothetical transmembrane protein	Not in COGs
AR505_0989	hypothetical transmembrane protein	Not in COGs
AR505_0990	hypothetical transmembrane protein	Not in COGs
AR505_0991	adhesin-like protein	Not in COGs
AR505_0992	adhesin-like protein	[Unclassified]
AR505_0993	hypothetical transmembrane protein	Not in COGs
AR505_0994	hypothetical protein	[R]
AR505_1002	conserved hypothetical	Not in COGs
AR505_1003	conserved hypothetical	Not in COGs
AR505_1004	hypothetical protein	[S]
AR505_1006	hypothetical transmembrane protein	Not in COGs
AR505_1013	hypothetical transmembrane protein	[Unclassified]
AR505_1015	ABC transporter permease protein	[R]
AR505_1016	hypothetical protein	[R]
AR505_1020	TPR repeat-containing protein	[R]
AR505_1022	hypothetical protein	Not in COGs
AR505_1024	hypothetical transmembrane protein	Not in COGs
AR505_1025	hypothetical transmembrane protein	Not in COGs
AR505_1026	hypothetical transmembrane protein	Not in COGs
AR505_1032	adhesin-like protein	[Unclassified]
AR505_1033	adhesin-like protein	[N]
AR505_1034	MatE efflux family protein	[V]
AR505_1035	methylcobalamin:CoM methyltransferase MtaA	[H]
AR505_1038	hypothetical protein	[Unclassified]
AR505_1040	uroporphyrinogen III synthase HemD	[H]
AR505_1049	hypothetical transmembrane protein	Not in COGs
AR505_1050	hypothetical protein	Not in COGs
AR505_1051	hypothetical protein	Not in COGs
AR505_1054	hypothetical protein	Not in COGs
AR505_1061	hypothetical protein	Not in COGs
AR505_1075	potassium uptake protein TrkH family	[P]
AR505_1076	drug resistance transporter Bcr/ CflA subfamily	[P]
AR505_1078	hypothetical protein	Not in COGs
AR505_1087	hypothetical protein	Not in COGs
AR505_1089	CRISPR-associated endonuclease Cas3-HD	[R]
AR505_1090	CRISPR type I-E/*E .coli*-associated protein CasA/Cse1	[Unclassified]
AR505_1091	CRISPR type I-E/*E .coli* -associated protein CasB/Cse2 complex	[Unclassified]
AR505_1092	CRISPR-associated protein Cas7/ Cse4/CasC	[Unclassified]
AR505_1093	CRISPR-associated protein Cas5/ CasD	[Unclassified]
AR505_1094	CRISPR-associated protein Cas6/ Cse3/CasE	[Unclassified]
AR505_1095	CRISPR-associated endonuclease Cas1	[L]
AR505_1096	hypothetical protein	Not in COGs
AR505_1097	hypothetical protein	Not in COGs
AR505_1098	hypothetical protein	Not in COGs
AR505_1099	transcriptional regulator MarR family	[K]
AR505_1103	hypothetical protein	[I]
AR505_1107	hypothetical protein	Not in COGs
AR505_1109	hypothetical transmembrane protein	Not in COGs
AR505_1110	hypothetical transmembrane protein	Not in COGs
AR505_1112	hypothetical protein	Not in COGs
AR505_1113	hypothetical protein	Not in COGs
AR505_1116	hypothetical protein	Not in COGs
AR505_1124	hypothetical protein	Not in COGs
AR505_1125	hypothetical protein	Not in COGs
AR505_1129	hypothetical protein	Not in COGs
AR505_1130	hypothetical protein	Not in COGs
AR505_1131	hypothetical protein	Not in COGs
AR505_1132	ATP-dependent DNA helicase UvrD/REP family	[L]
AR505_1138	transporter SDF family	[E]
AR505_1140	hypothetical transmembrane protein	[S]
AR505_1141	hypothetical protein	Not in COGs
AR505_1144	hypothetical protein	Not in COGs
AR505_1147	phospholipase D/ transphosphatidylase PlD	[I]
AR505_1151	tryptophan synthase beta subunit TrpB	[R]
AR505_1156	hypothetical transmembrane protein	[R]
AR505_1162	phosphoribosylanthranilate isomerase TrpF	[E]
AR505_1167	hypothetical protein	Not in COGs

AR505_1169	signal transduction histidine kinase	[T]
AR505_1170	signal transduction histidine kinase	[T]
AR505_1171	hypothetical protein	Not in COGs
AR505_1173	adhesin-like protein	Not in COGs
AR505_1174	hypothetical protein	[L]
AR505_1176	transcriptional regulator TetR family	[K]
AR505_1185	hydrogenase nickel insertion protein HypA	[J]
AR505_1186	hypothetical transmembrane protein	Not in COGs
AR505_1187	conserved hypothetical	Not in COGs
AR505_1188	hypothetical transmembrane protein	[Unclassified]
AR505_1190	hypothetical protein	Not in COGs
AR505_1194	SAM-dependent methyltransferase	[R]
AR505_1195	ABC transporter substrate-binding protein	[P]
AR505_1196	SAM-dependent methyltransferase	[R]
AR505_1200	hypothetical transmembrane protein	Not in COGs
AR505_1201	hypothetical protein	Not in COGs
AR505_1202	hypothetical protein	Not in COGs
AR505_1207	hypothetical transmembrane protein	Not in COGs
AR505_1208	hypothetical transmembrane protein	Not in COGs
AR505_1213	TPR repeat-containing protein	[R]
AR505_1215	hypothetical protein	[R]
AR505_1224	hypothetical protein	Not in COGs
AR505_1220	hypothetical transmembrane protein	Not in COGs
AR505_1221	hypothetical transmembrane protein	Not in COGs
AR505_1229	transcriptional regulator	[Unclassified]
AR505_1230	hypothetical protein	[Unclassified]
AR505_1231	hypothetical protein	[O]
AR505_1239	transposase	[L]
AR505_1240	cell wall/surface repeat-containing protein	Not in COGs
AR505_1241	hypothetical protein	Not in COGs
AR505_1244	GNAT family acetyltransferases	[R]
AR505_1246	Na^+-driven multidrug efflux pump	Not in COGs
AR505_1247	anaerobic ribonucleoside-triphosphate reductase NrdD	[O]
AR505_1250	hypothetical protein	Not in COGs
AR505_1251	DNA helicase	[L]
AR505_1252	hypothetical protein	Not in COGs
AR505_1253	hypothetical protein	[K]
AR505_1254	hypothetical protein	Not in COGs
AR505_1255	hypothetical protein	[Unclassified]
AR505_1256	hypothetical protein	Not in COGs
AR505_1258	hypothetical protein	Not in COGs
AR505_1266	hypothetical protein	[R]
AR505_1267	helicase	[L]
AR505_1269	hypothetical transmembrane protein	[Unclassified]
AR505_1270	hypothetical protein	Not in COGs
AR505_1276	hypothetical protein	[Unclassified]
AR505_1283	hypothetical transmembrane protein	[S]
AR505_1286	hypothetical protein	Not in COGs
AR505_1290	adhesin-like protein	Not in COGs
AR505_1292	ABC transporter permease protein	[P]
AR505_1293	ABC transporter substrate-binding protein	[P]
AR505_1297	adenosylcobalamin synthase CobS	[H]
AR505_1300	hypothetical protein	Not in COGs
AR505_1301	hypothetical protein	[Unclassified]
AR505_1309	DNA alkylation repair enzyme	[L]
AR505_1319	hypothetical protein	Not in COGs
AR505_1326	hypothetical protein	Not in COGs
AR505_1330	methyltransferase cognate corrinoid proteins	[R]
AR505_1334	TPR repeat-containing protein	[R]
AR505_1343	universal stress protein UspA	[T]
AR505_1345	hypothetical protein	Not in COGs
AR505_1347	hypothetical transmembrane protein	Not in COGs
AR505_1350	hypothetical transmembrane protein	Not in COGs
AR505_1352	alpha/beta hydrolase fold protein	[I]
AR505_1354	hypothetical transmembrane protein	[Unclassified]
AR505_1356	transposase IS4 family	[Unclassified]
AR505_1359	hypothetical transmembrane protein	Not in COGs
AR505_1360	hypothetical protein	Not in COGs
AR505_1361	hypothetical protein	Not in COGs
AR505_1362	hypothetical protein	Not in COGs
AR505_1364	hypothetical protein	[L]
AR505_1365	hypothetical protein	[L]
AR505_1366	hypothetical protein	Not in COGs
AR505_1367	archaeal ATPase	[R]
AR505_1376	isopropylmalate/isohomocitrate dehydrogenase	[E]
AR505_1380	hypothetical protein	Not in COGs
AR505_1418	hypothetical transmembrane protein	[R]
AR505_1419	hypothetical transmembrane protein	Not in COGs
AR505_1422	haloacid dehalogenase-like hydrolase	[R]
AR505_1428	hypothetical transmembrane protein	Not in COGs
AR505_1436	hypothetical protein	Not in COGs
AR505_1451	hypothetical transmembrane protein	Not in COGs
AR505_1463	TPR repeat-containing protein	[R]
AR505_1466	addiction module antitoxin RelB/ DinJ	[L]
AR505_1467	hypothetical protein	[K]
AR505_1480	hypothetical transmembrane protein	Not in COGs
AR505_1495	Fic family protein	[S]
AR505_1503	cobalamin biosynthesis protein CbiX	[S]
AR505_1506	iron chelate uptake ABC transporter permease	[P]
AR505_1507	iron chelate uptake ABC transporter permease inner membrane subunit	[P]
AR505_1509	adhesin-like protein	[N]
AR505_1528	hypothetical protein	Not in COGs
AR505_1533	hypothetical transmembrane protein	[R]
AR505_1535	hypothetical protein	Not in COGs
AR505_1536	hypothetical transmembrane protein	[Unclassified]
AR505_1539	hypothetical transmembrane protein	Not in COGs
AR505_1541	hypothetical protein	Not in COGs
AR505_1545	hypothetical protein	Not in COGs
AR505_1557	Atpase	[R]
AR505_1558	nucleotidyltransferase	[R]
AR505_1559	adhesin-like protein	[R]
AR505_1560	adhesin-like protein	Not in COGs
AR505_1561	adhesin-like protein	Not in COGs
AR505_1563	transposase	[L]
AR505_1564	ClpP class periplasmic serine protease	[O]
AR505_1565	hypothetical protein	Not in COGs
AR505_1566	death-on-curing family protein	[R]

AR505_1567	hypothetical protein	Not in COGs
AR505_1568	hypothetical protein	Not in COGs
AR505_1569	hypothetical protein	Not in COGs
AR505_1570	phage integrase	[L]
AR505_1571	hypothetical protein	Not in COGs
AR505_1572	hypothetical protein	Not in COGs
AR505_1573	hypothetical protein	[L]
AR505_1580	hypothetical protein	Not in COGs
AR505_1592	hypothetical protein	Not in COGs
AR505_1593	ATPase	[K]
AR505_1600	hypothetical protein	Not in COGs
AR505_1612	signal peptidase I	[Unclassified]
AR505_1652	hypothetical transmembrane protein	[R]
AR505_1659	hypothetical protein	Not in COGs
AR505_1672	hypothetical protein	[S]
AR505_1677	hypothetical transmembrane protein	[S]
AR505_1678	hypothetical transmembrane protein	Not in COGs
AR505_1686	sulfate permease SulP	[P]
AR505_1694	hypothetical protein	[S]
AR505_1697	phage integrase family protein	[L]
AR505_1698	hypothetical protein	[Unclassified]
AR505_1699	hypothetical protein	Not in COGs
AR505_1700	hypothetical protein	Not in COGs
AR505_1701	hypothetical protein	Not in COGs
AR505_1702	hypothetical protein	Not in COGs
AR505_1704	hypothetical protein	[Unclassified]
AR505_1711	hypothetical transmembrane protein	[R]
AR505_1218	haloacid dehalogenase-like hydrolase	[R]
AR505_1713	hypothetical transmembrane protein	[Unclassified]
AR505_1720	hypothetical protein	Not in COGs
AR505_1722	hypothetical protein	[O]
AR505_1724	hypothetical protein	Not in COGs
AR505_1728	hypothetical protein	Not in COGs
AR505_1730	hypothetical protein	Not in COGs
AR505_1731	hypothetical protein	Not in COGs
AR505_1734	hypothetical protein	Not in COGs
AR505_1735	hypothetical protein	[L]
AR505_1736	hypothetical protein	[O]
AR505_1737	hypothetical protein	Not in COGs
AR505_1738	ATPase	[O]
AR505_1739	hypothetical transmembrane protein	Not in COGs
AR505_1741	adhesin-like protein	Not in COGs
AR505_1742	hypothetical protein	Not in COGs
AR505_1744	GNAT family acetyltransferases	[R]
AR505_1747	hypothetical transmembrane protein	Not in COGs
AR505_1749	hypothetical transmembrane protein	[Unclassified]
AR505_1750	hypothetical transmembrane protein	Not in COGs
AR505_1760	hypothetical transmembrane protein	Not in COGs
AR505_1761	adhesin-like protein	Not in COGs
AR505_1762	hypothetical transmembrane protein	[S]
AR505_1776	retron-type reverse transcriptase	[L]
AR505_1777	transposase IS605 OrfB family	[L]
AR505_1778	glycosyl transferase GT2 family	[M]
AR505_1779	hypothetical transmembrane protein	Not in COGs
AR505_1790	hypothetical protein	Not in COGs
AR505_1792	cell division protein FtsZ	[D]
AR505_1793	hypothetical protein	[K]
AR505_1813	CAAX amino terminal protease family protein	Not in COGs

only one gene from each gene family is represented.

Table A.4.13. Gene families unique to ISO4-G1

****

Locus_tag	Predicted gene product	COG category
ISO4G1_0002	signal peptidase I	[U]
ISO4G1_0017	CAAX amino terminal protease family protein	Not in COGs
ISO4G1_0018	hypothetical protein	[X]
ISO4G1_0021	transcriptional regulator	[K]
ISO4G1_0022	hypothetical protein	[S]
ISO4G1_0038	cell division GTPase	[D]
ISO4G1_0048	hypothetical protein	[Unclassified]
ISO4G1_0049	hypothetical protein	[Unclassified]
ISO4G1_0052	prefoldin alpha subunit PfdA	[O]
ISO4G1_0056	metallo-β -lactamase domain-containing protein	[R]
ISO4G1_0067	hypothetical protein	[R]
ISO4G1_0077	geranylgeranyl pyrophosphate synthase	[H]
ISO4G1_0079	NADPH-dependent FMN reductase	[R]
ISO4G1_0086	hypothetical protein	[K]
ISO4G1_0088	hypothetical protein	Not in COGs
ISO4G1_0089	cell surface protein	Not in COGs
ISO4G1_0090	hypothetical protein	Not in COGs
ISO4G1_0091	divergent AAA domain-containing protein	[K]
ISO4G1_0100	hypothetical protein	Not in COGs
ISO4G1_0107	hypothetical protein	Not in COGs
ISO4G1_0108	hypothetical protein	Not in COGs
ISO4G1_0112	hypothetical protein	[S]
ISO4G1_0115	AMP-binding enzyme	[I]
ISO4G1_0116	GDSL-like lipase	[E]
ISO4G1_0133	hypothetical protein	[X]
ISO4G1_0139	geranylgeranyl reductase family protein	[C]
ISO4G1_0142	ribonuclease III Rnc	[K]
ISO4G1_0147	GNAT family acetyltransferase	[K]
ISO4G1_0161	F_{420} dehydrogenase subunit J FpoJ	Not in COGs
ISO4G1_0166	geranylgeranyl pyrophosphate synthase	[H]
ISO4G1_0169	hypothetical protein	Not in COGs
ISO4G1_0170	transposase	Not in COGs
ISO4G1_0178	hypothetical protein	[X]
ISO4G1_0179	cell surface protein	Not in COGs
ISO4G1_0180	phosphomannomutase/ phosphoglucomutase	[G]
ISO4G1_0187	ArsR family transcriptional regulator	[R]
ISO4G1_0190	aldehyde dehydrogenase family protein	[C]
ISO4G1_0191	hypothetical protein	[C]
ISO4G1_0192	hypothetical protein	[C]
ISO4G1_0193	hypothetical protein	[C]
ISO4G1_0195	hypothetical protein	Not in COGs
ISO4G1_0196	hypothetical protein	Not in COGs
ISO4G1_0198	ABC transporter ATP-binding protein	[R]
ISO4G1_0210	mechanosensitive ion channel protein	[M]
ISO4G1_0211	hypothetical protein	[N]
ISO4G1_0219	hypothetical protein	[S]
ISO4G1_0220	hypothetical protein	Not in COGs
ISO4G1_0221	hypothetical protein	Not in COGs
ISO4G1_0241	hypothetical protein	Not in COGs
ISO4G1_0242	hypothetical protein	Not in COGs
ISO4G1_0243	hypothetical protein	[X]
ISO4G1_0247	transporter monovalent cation: proton antiporter-2 family	[P]
ISO4G1_0251	methyltransferase cognate corrinoid protein	[R]
ISO4G1_0252	SPFH domain-band 7 family protein	[S]

ISO4G1_0253	hypothetical protein	[Unclassified]
ISO4G1_0254	hypothetical protein	Not in COGs
ISO4G1_0255	hypothetical protein	Not in COGs
ISO4G1_0256	hypothetical protein	[G]
ISO4G1_0258	ABC transporter permease protein	[R]
ISO4G1_0262	hypothetical protein	Not in COGs
ISO4G1_0266	hypothetical protein	Not in COGs
ISO4G1_0275	hypothetical protein	Not in COGs
ISO4G1_0276	nitroreductase family protein	[C]
ISO4G1_0281	hypothetical protein	[S]
ISO4G1_0285	hypothetical protein	Not in COGs
ISO4G1_0323	hypothetical protein	Not in COGs
ISO4G1_0324	hypothetical protein	Not in COGs
ISO4G1_0328	hypothetical protein	Not in COGs
ISO4G1_0334	hypothetical protein	Not in COGs
ISO4G1_0336	hypothetical protein	Not in COGs
ISO4G1_0338	adenosylcobinamide amidohydrolase CbiZ	[S]
ISO4G1_0339	adenosylcobinamide amidohydrolase CbiZ	[S]
ISO4G1_0342	HAD-superfamily hydrolase	[R]
ISO4G1_0344	hypothetical protein	Not in COGs
ISO4G1_0345	hypothetical protein	Not in COGs
ISO4G1_0347	PAC2 family protein	[R]
ISO4G1_0354	hypothetical protein	[N]
ISO4G1_0356	hypothetical protein	Not in COGs
ISO4G1_0361	MATE efflux family protein	[V]
ISO4G1_0362	iron-containing alcohol dehydrogenase	[C]
ISO4G1_0369	hypothetical protein	[R]
ISO4G1_0371	hypothetical protein	Not in COGs
ISO4G1_0372	ZIP zinc transporter family protein	[P]
ISO4G1_0377	allosteric regulator of homoserine dehydrogenase	[E]
ISO4G1_0385	hypothetical protein	Not in COGs
ISO4G1_0387	hypothetical protein	Not in COGs
ISO4G1_0388	integrase family protein	[L]
ISO4G1_0390	hypothetical protein	Not in COGs
ISO4G1_0392	hypothetical protein	[Unclassified]
ISO4G1_0395	C-5 cytosine-specific DNA methylase	[L]
ISO4G1_0396	hypothetical protein	[Unclassified]
ISO4G1_0397	hypothetical protein	[Unclassified]
ISO4G1_0398	hypothetical protein	Not in COGs
ISO4G1_0399	hypothetical protein	[Unclassified]
ISO4G1_0400	hypothetical protein	[X]
ISO4G1_0402	hypothetical protein	Not in COGs
ISO4G1_0403	hypothetical protein	Not in COGs
ISO4G1_0404	hypothetical protein	Not in COGs
ISO4G1_0405	hypothetical protein	Not in COGs
ISO4G1_0406	hypothetical protein	Not in COGs
ISO4G1_0407	hypothetical protein	Not in COGs
ISO4G1_0408	hypothetical protein	Not in COGs
ISO4G1_0409	hypothetical protein	Not in COGs
ISO4G1_0410	hypothetical protein	[C]
ISO4G1_0411	hypothetical protein	Not in COGs
ISO4G1_0412	hypothetical protein	Not in COGs
ISO4G1_0413	hypothetical protein	Not in COGs
ISO4G1_0414	hypothetical protein	Not in COGs
ISO4G1_0416	hypothetical protein	Not in COGs
ISO4G1_0417	hypothetical protein	Not in COGs
ISO4G1_0418	hypothetical protein	Not in COGs
ISO4G1_0419	hypothetical protein	Not in COGs
ISO4G1_0420	TIGR02543 family repeat-containing cell surface protein	[S]
ISO4G1_0421	TIGR02543 family repeat-containing cell surface protein	Not in COGs
ISO4G1_0423	hypothetical protein	Not in COGs
ISO4G1_0424	hypothetical protein	Not in COGs
ISO4G1_0425	cell surface protein	[R]
ISO4G1_0426	hypothetical protein	Not in COGs
ISO4G1_0427	hypothetical protein	Not in COGs
ISO4G1_0428	hypothetical protein	Not in COGs
ISO4G1_0429	hypothetical protein	Not in COGs
ISO4G1_0432	MFS transporter	[G]
ISO4G1_0435	hypothetical protein	[S]
ISO4G1_0436	hypothetical protein	[L]
ISO4G1_0444	alpha/beta fold family hydrolase	[R]
ISO4G1_0445	transcriptional regulator	[X]
ISO4G1_0456	flavodoxin	Not in COGs
ISO4G1_0458	hypothetical protein	Not in COGs
ISO4G1_0459	XRE family transcriptional regulator	[Unclassified]
ISO4G1_0460	hypothetical protein	[X]
ISO4G1_0466	hypothetical protein	[U]
ISO4G1_0469	EamA-like transporter family protein	[G]
ISO4G1_0473	hypothetical protein	[X]
ISO4G1_0476	hypothetical protein	Not in COGs
ISO4G1_0479	ATPase AAA+ superfamily	[S]
ISO4G1_0480	TIGR02543 family repeat-containing cell surface protein	Not in COGs
ISO4G1_0482	archaeal ATPase	[R]
ISO4G1_0484	hypothetical protein	Not in COGs
ISO4G1_0489	signal peptidase I	[U]
ISO4G1_0490	hypothetical protein	Not in COGs
ISO4G1_0491	hypothetical protein	Not in COGs
ISO4G1_0492	hypothetical protein	Not in COGs
ISO4G1_0493	hypothetical protein	Not in COGs
ISO4G1_0494	hypothetical protein	Not in COGs
ISO4G1_0495	TIGR02543 family repeat-containing cell surface protein	[S]
ISO4G1_0496	hypothetical protein	Not in COGs
ISO4G1_0497	hypothetical protein	Not in COGs
ISO4G1_0498	ATPase	[R]
ISO4G1_0499	hypothetical protein	Not in COGs
ISO4G1_0501	hypothetical protein	Not in COGs
ISO4G1_0510	hypothetical protein	[X]
ISO4G1_0512	hypothetical protein	[Unclassified]
ISO4G1_0513	hypothetical protein	Not in COGs
ISO4G1_0514	hypothetical protein	Not in COGs
ISO4G1_0515	hypothetical protein	[Unclassified]
ISO4G1_0516	hypothetical protein	[X]
ISO4G1_0517	cell surface protein	[C]
ISO4G1_0518	hypothetical protein	Not in COGs
ISO4G1_0524	electron transfer flavoprotein domain-containing protein	[C]
ISO4G1_0527	hypothetical protein	Not in COGs
ISO4G1_0532	hypothetical protein	[L]
ISO4G1_0533	hypothetical protein	Not in COGs
ISO4G1_0537	nitroreductase family protein	[C]
ISO4G1_0539	geranylgeranyl reductase family protein	[C]
ISO4G1_0542	hypothetical protein	Not in COGs
ISO4G1_0543	pyridoxamine 5'-phosphate oxidase-like FMN-binding protein	[R]
ISO4G1_0544	PAS domain-containing protein	[T]
ISO4G1_0551	GNAT family acetyltransferase	[J]
ISO4G1_0553	hypothetical protein	[X]
ISO4G1_0555	SAM-dependent methyltransferase	[Q]
ISO4G1_0558	iron ABC transporter substrate-binding protein	[P]
ISO4G1_0563	DEAD/DEAH box helicase	[L]
ISO4G1_0564	MATE efflux family protein	[V]
ISO4G1_0566	divergent AAA domain-containing protein	[K]
ISO4G1_0568	MMPL family transporter	[R]
ISO4G1_0570	hypothetical protein	[S]
ISO4G1_0587	bifunctional protein FolC	[H]
ISO4G1_0588	alpha-L-glutamate ligase RimK family	[H]
ISO4G1_0589	alpha-L-glutamate ligase RimK family	[H]
ISO4G1_0595	response regulator domain-containing protein	[T]
ISO4G1_0597	xylose isomerase-like TIM barrel domain-containing protein	[Unclassified]

ISO4G1_0604	hydrolase TatD family	[H]
ISO4G1_0606	hypothetical protein	Not in COGs
ISO4G1_0611	hypothetical protein	Not in COGs
ISO4G1_0612	hypothetical protein	Not in COGs
ISO4G1_0635	hypothetical protein	Not in COGs
ISO4G1_0638	cell surface protein	[X]
ISO4G1_0639	HAD-superfamily hydrolase	[R]
ISO4G1_0642	ABC transporter ATP-binding/ permease protein	[V]
ISO4G1_0644	SAM-dependent methyltransferase	[Q]
ISO4G1_0645	iron ABC transporter ATP-binding protein	[P]
ISO4G1_0646	iron ABC transporter permease protein	[P]
ISO4G1_0647	iron ABC transporter substrate-binding protein	[Q]
ISO4G1_0649	SAM-dependent methlyltransferase	[Q]
ISO4G1_0651	hypothetical protein	Not in COGs
ISO4G1_0652	hypothetical protein	Not in COGs
ISO4G1_0656	addiction module antitoxin, RelB/DinJ family	[L]
ISO4G1_0660	pyrrolysine biosynthesis protein PylD	[X]
ISO4G1_0662	small multidrug resistance protein	[P]
ISO4G1_0663	small multidrug resistance protein	[P]
ISO4G1_0670	ThiF family protein	[H]
ISO4G1_0672	hypothetical protein	[S]
ISO4G1_0674	NADPH-dependent FMN reductase	[R]
ISO4G1_0675	hypothetical protein	Not in COGs
ISO4G1_0676	cobaltochelatase CobN	[H]
ISO4G1_0679	hypothetical protein	[P]
ISO4G1_0680	*O*-methyltransferase	[Q]
ISO4G1_0683	hypothetical protein	Not in COGs
ISO4G1_0687	hypothetical protein	Not in COGs
ISO4G1_0692	ATP-dependent DNA helicase	[L]
ISO4G1_0694	hypothetical protein	Not in COGs
ISO4G1_0695	ATP-dependent DNA helicase	[L]
ISO4G1_0696	hypothetical protein	Not in COGs
ISO4G1_0697	hypothetical protein	Not in COGs
ISO4G1_0698	hypothetical protein	Not in COGs
ISO4G1_0699	hypothetical protein	Not in COGs
ISO4G1_0700	type III restriction system endonuclease	[X]
ISO4G1_0701	type III restriction system methylase	[X]
ISO4G1_0702	type III restriction system methylase	[X]
ISO4G1_0703	RloB-like protein	[X]
ISO4G1_0704	ATPase AAA	[R]
ISO4G1_0705	RloA-like protein	[R]
ISO4G1_0706	Sel1 domain-containing protein	[R]
ISO4G1_0707	Sel1 domain-containing protein	[R]
ISO4G1_0708	hypothetical protein	Not in COGs
ISO4G1_0709	PD-(D/E)XK nuclease superfamily protein	Not in COGs
ISO4G1_0710	ATP-dependent DNA helicase	[L]
ISO4G1_0713	hypothetical protein	Not in COGs
ISO4G1_0714	hypothetical protein	Not in COGs
ISO4G1_0715	hypothetical protein	Not in COGs
ISO4G1_0716	SAM-dependent methyltransferase	[Q]
ISO4G1_0717	reverse transcriptase	[L]
ISO4G1_0718	hypothetical protein	Not in COGs
ISO4G1_0719	hypothetical protein	Not in COGs
ISO4G1_0720	Sel1 domain-containing protein	[R]
ISO4G1_0722	hypothetical protein	Not in COGs
ISO4G1_0724	hypothetical protein	[Unclassified]
ISO4G1_0726	hypothetical protein	Not in COGs
ISO4G1_0727	DNA double-strand break repair protein Mre11	[L]
ISO4G1_0728	ATPase AAA	[S]
ISO4G1_0731	cell surface protein	Not in COGs
ISO4G1_0732	ribosomal-protein-alanine acetyltransferase RimI	[R]
ISO4G1_0733	phospholipase-like protein	[I]
ISO4G1_0736	ArsR family transcriptional regulator	[M]
ISO4G1_0740	MATE efflux family protein	[V]
ISO4G1_0741	hypothetical protein	[C]
ISO4G1_0744	hypothetical protein	Not in COGs
ISO4G1_0745	NADPH-dependent FMN reductase	[R]
ISO4G1_0746	transcriptional regulator	[K]
ISO4G1_0747	sodium bile acid symporter family protein	[R]
ISO4G1_0749	TPR repeat-containing protein	[O]
ISO4G1_0752	hypothetical protein	[R]
ISO4G1_0755	cobaltochelatase CobN	[H]
ISO4G1_0760	hypothetical protein	Not in COGs
ISO4G1_0763	Sel1 domain-containing protein	[R]
ISO4G1_0764	hypothetical protein	Not in COGs
ISO4G1_0765	HAD-superfamily hydrolase	[R]
ISO4G1_0766	hypothetical protein	Not in COGs
ISO4G1_0767	hypothetical protein	Not in COGs
ISO4G1_0768	hypothetical protein	Not in COGs
ISO4G1_0769	short-chain dehydrogenase/ reductase	[R]
ISO4G1_0772	hypothetical protein	Not in COGs
ISO4G1_0774	hypothetical protein	Not in COGs
ISO4G1_0775	hypothetical protein	Not in COGs
ISO4G1_0778	MarR family transcriptional regulator	[K]
ISO4G1_0782	potassium transport protein TrkA	[P]
ISO4G1_0783	potassium transport protein TrkH	[P]
ISO4G1_0784	hypothetical protein	Not in COGs
ISO4G1_0785	ATPase AAA	[R]
ISO4G1_0786	hypothetical protein	[Unclassified]
ISO4G1_0789	thioesterase family protein	[Q]
ISO4G1_0790	hypothetical protein	[S]
ISO4G1_0791	PD-(D/E)XK nuclease superfamily protein	Not in COGs
ISO4G1_0792	ATP-dependent DNA helicase	[L]
ISO4G1_0797	CDP-alcohol phosphatidyl-transferase family protein	[I]
ISO4G1_0802	hypothetical protein	[Unclassified]
ISO4G1_0805	hypothetical protein	[X]
ISO4G1_0806	hypothetical protein	Not in COGs
ISO4G1_0808	flavodoxin	[C]
ISO4G1_0812	hypothetical protein	Not in COGs
ISO4G1_0814	hypothetical protein	[K]
ISO4G1_0817	cell surface protein	[X]
ISO4G1_0819	CorA-like $Mg2^{+}$ transporter protein	[P]
ISO4G1_0820	peptidase M50 family	[M]
ISO4G1_0822	hypothetical protein	[S]
ISO4G1_0829	hypothetical protein	Not in COGs
ISO4G1_0830	hypothetical protein	Not in COGs
ISO4G1_0831	hypothetical protein	[Unclassified]
ISO4G1_0834	CopG family transcripitonal regulator	[X]
ISO4G1_0835	cell division protein FtsZ	[C]
ISO4G1_0838	GNAT family acetyltransferase	[K]
ISO4G1_0840	carbamoyl-phosphate synthase small subunit CarA	[E]
ISO4G1_0842	hypothetical protein	Not in COGs
ISO4G1_0847	hypothetical protein	Not in COGs
ISO4G1_0848	SAM-dependent methlyltransferase	[Q]
ISO4G1_0849	hypothetical protein	[Unclassified]
ISO4G1_0851	hypothetical protein	[L]
ISO4G1_0852	hypothetical protein	Not in COGs
ISO4G1_0855	metallo-β-lactamase domain-containing protein	[X]
ISO4G1_0866	hypothetical protein	Not in COGs
ISO4G1_0867	hypothetical protein	Not in COGs

ISO4G1_0870	hypothetical protein	Not in COGs
ISO4G1_0872	type II secretion system protein	[X]
ISO4G1_0874	methyltransferase family protein	[Q]
ISO4G1_0876	cell division protein Fic	[S]
ISO4G1_0878	hypothetical protein	Not in COGs
ISO4G1_0879	trypsin-like peptidase domain-containing protein	[O]
ISO4G1_0880	carbon-nitrogen hydrolase	[R]
ISO4G1_0881	hypothetical protein	[K]
ISO4G1_0883	hypothetical protein	Not in COGs
ISO4G1_0885	hypothetical protein	Not in COGs
ISO4G1_0888	iron ABC transporter substrate-binding protein	[P]
ISO4G1_0889	nitrogenase iron protein NifH	[P]
ISO4G1_0890	hypothetical protein	Not in COGs
ISO4G1_0892	hypothetical protein	[G]
ISO4G1_0894	cobaltochelatase subunit	[H]
ISO4G1_0898	FAD/FMN-containing dehydrogenase	[C]
ISO4G1_0903	ABC transporter ATP-binding/ permease protein	[V]
ISO4G1_0904	ABC transporter ATP-binding/ permease protein	[V]
ISO4G1_0905	hypothetical protein	Not in COGs
ISO4G1_0908	universal stress protein	[T]
ISO4G1_0909	hypothetical protein	[S]
ISO4G1_0910	hypothetical protein	[S]
ISO4G1_0912	hypothetical protein	Not in COGs
ISO4G1_0914	MFS transporter	[G]
ISO4G1_0915	hypothetical protein	Not in COGs
ISO4G1_0916	hypothetical protein	[Unclassified]
ISO4G1_0918	hypothetical protein	[Unclassified]
ISO4G1_0921	hypothetical protein	[N]
ISO4G1_0924	hypothetical protein	Not in COGs
ISO4G1_0925	4Fe-4S binding domain-containing protein	[C]
ISO4G1_0926	metallo-β-lactamase family protein/flavodoxin	[C]
ISO4G1_0927	hypothetical protein	Not in COGs
ISO4G1_0929	hypothetical protein	[L]
ISO4G1_0935	iron dependent repressor	[K]
ISO4G1_0937	MarR family transcriptional regulator	[K]
ISO4G1_0940	hypothetical protein	[S]
ISO4G1_0942	hypothetical protein	[E]
ISO4G1_0943	hypothetical protein	Not in COGs
ISO4G1_0944	TauE family protein	[X]
ISO4G1_0945	trimethylamine:corrinoid methyltransferase MttB	[H]
ISO4G1_0948	MFS transporter	[G]
ISO4G1_0949	hypothetical protein	Not in COGs
ISO4G1_0951	cell surface protein	[R]
ISO4G1_0959	inositol monophosphatase	[G]
ISO4G1_0965	iron transporter FeoA	[P]
ISO4G1_0968	hypothetical protein	Not in COGs
ISO4G1_0969	iron transporter FeoA	[P]
ISO4G1_0972	hypothetical protein	Not in COGs
ISO4G1_0974	Sel1 domain-containing protein	[R]
ISO4G1_0976	hypothetical protein	[Unclassified]
ISO4G1_0978	molecular chaperone GrpE	[O]
ISO4G1_0980	nitrogenase-related protein	[C]
ISO4G1_0981	nitrogenase iron protein NifH	[P]
ISO4G1_0982	universal stress protein	[T]
ISO4G1_0983	4'-phosphopantetheinyl transferase	[H]
ISO4G1_0984	non-ribosomal peptide synthetase	[Q]
ISO4G1_0991	nitrate/sulfonate/bicarbonate ABC transporter substrate-binding protein	[P]
ISO4G1_0993	hypothetical protein	Not in COGs
ISO4G1_0998	cobalamin 5'-phosphate synthase CobS	[H]
ISO4G1_1001	hypothetical protein	Not in COGs
ISO4G1_1003	X-Pro dipeptidyl-peptidase	[X]
ISO4G1_1012	isopropylmalate/isohomocitrate dehydrogenase	[C]
ISO4G1_1014	hypothetical protein	Not in COGs
ISO4G1_1015	cation diffusion facilitator family transporter	[P]
ISO4G1_1023	hypothetical protein	Not in COGs
ISO4G1_1024	4Fe-4S binding domain-containing protein	[C]
ISO4G1_1027	hypothetical protein	[Unclassified]
ISO4G1_1033	LD-carboxypeptidase S66 family	[V]
ISO4G1_1035	hypothetical protein	[X]
ISO4G1_1036	hypothetical protein	[X]
ISO4G1_1039	isochorismatase family protein	[Q]
ISO4G1_1040	hypothetical protein	[S]
ISO4G1_1043	hypothetical protein	Not in COGs
ISO4G1_1046	MFS transporter	[G]
ISO4G1_1048	hypothetical protein	Not in COGs
ISO4G1_1059	hypothetical protein	Not in COGs
ISO4G1_1063	nitrogenase component 1 type oxidoreductase	[C]
ISO4G1_1064	nitrogenase component 1 type oxidoreductase	[C]
ISO4G1_1065	iron ABC transporter ATP-binding protein	[P]
ISO4G1_1066	iron ABC transporter permease protein	[P]
ISO4G1_1067	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1068	TPR repeat-containing protein	[M]
ISO4G1_1072	hypothetical protein	Not in COGs
ISO4G1_1073	hypothetical protein	Not in COGs
ISO4G1_1074	hypothetical protein	Not in COGs
ISO4G1_1075	2'-5' RNA ligase	[J]
ISO4G1_1076	protein-tyrosine phosphatase	[T]
ISO4G1_1083	TPR repeat-containing protein	[R]
ISO4G1_1085	hypothetical protein	[I]
ISO4G1_1095	cobaltochelatase CobN	[H]
ISO4G1_1096	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1100	hypothetical protein	Not in COGs
ISO4G1_1112	nitroreductase family protein	[C]
ISO4G1_1113	FKBP-type peptidyl-prolyl cis-trans isomerase	[O]
ISO4G1_1117	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1118	MarR family transcriptional regulator	[K]
ISO4G1_1119	hypothetical protein	[X]
ISO4G1_1123	archaeosine tRNA-ribosyltransferase	[O]
ISO4G1_1125	hypothetical protein	[X]
ISO4G1_1130	Sel1 domain-containing protein	[R]
ISO4G1_1131	hypothetical protein	Not in COGs
ISO4G1_1139	transglutaminase domain-containing protein	[E]
ISO4G1_1140	DNA alkylation repair enzyme	[X]
ISO4G1_1142	GNAT family acetyltransferase	[K]
ISO4G1_1153	hypothetical protein	[S]
ISO4G1_1154	transposase	Not in COGs
ISO4G1_1155	MATE efflux family protein	[V]
ISO4G1_1156	Na^+/H^+ antiporter NhaC family	[C]
ISO4G1_1165	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1174	nitrogenase-related protein	[C]
ISO4G1_1177	hypothetical protein	[M]
ISO4G1_1179	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1182	4Fe-4S binding domain-containing protein	[X]
ISO4G1_1183	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1193	isochorismatase family protein	[Q]
ISO4G1_1194	hypothetical protein	Not in COGs
ISO4G1_1197	5'-nucleotidase	Not in COGs

ISO4G1_1200	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1201	hypothetical protein	Not in COGs
ISO4G1_1205	hypothetical protein	Not in COGs
ISO4G1_1207	4Fe-4S dicluster domain-containing protein	[C]
ISO4G1_1212	cell surface protein	S]
ISO4G1_1220	hypothetical protein	Not in COGs
ISO4G1_1221	hypothetical protein	Not in COGs
ISO4G1_1223	cobaltochelatase subunit	[H]
ISO4G1_1226	hypothetical protein	Not in COGs
ISO4G1_1227	hypothetical protein	[X]
ISO4G1_1232	hypothetical protein	[X]
ISO4G1_1233	transposase	[L]
ISO4G1_1242	metallo-β-lactamase domain-containing protein	[R]
ISO4G1_1243	hypothetical protein	Not in COGs
ISO4G1_1244	hypothetical protein	[X]
ISO4G1_1256	cobalamin biosynthesis protein CbiX	[S]
ISO4G1_1257	hypothetical protein	Not in COGs
ISO4G1_1264	hypothetical protein	Not in COGs
ISO4G1_1267	flavodoxin	[Unclassified]
ISO4G1_1272	transglutaminase domain-containing protein	[E]
ISO4G1_1277	hypothetical protein	Not in COGs
ISO4G1_1282	hypothetical protein	[J]
ISO4G1_1293	ABC transporter permease protein	[C]
ISO4G1_1307	PAC2 family protein	[R]
ISO4G1_1308	ATPase	[R]
ISO4G1_1310	flavin oxidoreductase	[R]
ISO4G1_1311	hypothetical protein	[X]
ISO4G1_1312	HxlR family transcriptional regulator	[K]
ISO4G1_1313	metallophosphoesterase	[R]
ISO4G1_1314	hypothetical protein	Not in COGs
ISO4G1_1316	hypothetical protein	Not in COGs
ISO4G1_1320	adenine phosphoribosyltransferase	[F]
ISO4G1_1321	hypothetical protein	[S]
ISO4G1_1323	PAP2 superfamily protein	[Unclassified]
ISO4G1_1326	cell surface protein	[X]
ISO4G1_1331	hypothetical protein	Not in COGs
ISO4G1_1333	hypothetical protein	Not in COGs
ISO4G1_1336	hypothetical protein	[S]
ISO4G1_1337	hypothetical protein	[X]
ISO4G1_1348	MORN repeat-containing protein	[S]
ISO4G1_1352	SAM-dependent methyltransferase	[Q]
ISO4G1_1359	hypothetical protein	[S]
ISO4G1_1361	hypothetical protein	Not in COGs
ISO4G1_1363	phosphoglucosamine mutase GlmM	[G]
ISO4G1_1371	ribosomal protein L6P Rpl6p	Not in COGs
ISO4G1_1372	hypothetical protein	Not in COGs
ISO4G1_1377	Fic family protein	[D]
ISO4G1_1379	phenazine biosynthesis-like protein	[R]
ISO4G1_1380	hypothetical protein	[X]
ISO4G1_1389	HPP family protein	[T]
ISO4G1_1390	hypothetical protein	Not in COGs
ISO4G1_1392	TetR family transcriptional regulator	[K]
ISO4G1_1393	heavy metal translocating P-type ATPase	[P]
ISO4G1_1396	chorismate mutase/prephenate dehydrogenase	[J]
ISO4G1_1400	shikimate kinase	[E]
ISO4G1_1404	hypothetical protein	[R]
ISO4G1_1407	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1409	hypothetical protein	Not in COGs
ISO4G1_1411	heavy-metal-associated domain-containing protein	[P]
ISO4G1_1412	ArsR family transcriptional regulator	[K]
ISO4G1_1413	NADPH-dependent FMN reductase	[R]
ISO4G1_1415	hypothetical protein	Not in COGs
ISO4G1_1418	TraB family protein	[S]
ISO4G1_1419	hypothetical protein	Not in COGs
ISO4G1_1423	heavy metal-associated domain-containing protein	[P]
ISO4G1_1426	EamA-like transporter family protein	[G]
ISO4G1_1427	EamA-like transporter family protein	[G]
ISO4G1_1434	hypothetical protein	Not in COGs
ISO4G1_1435	oligosaccharyl transferase	[R]
ISO4G1_1448	MarR family transcriptional regulator	[K]
ISO4G1_1453	GHMP family kinase	[R]
ISO4G1_1458	hypothetical protein	[S]
ISO4G1_1461	EamA-like transporter family protein	[G]
ISO4G1_1466	hypothetical protein	[X]
ISO4G1_1468	hypothetical protein	Not in COGs
ISO4G1_1469	hypothetical protein	Not in COGs
ISO4G1_1471	hypothetical protein	Not in COGs
ISO4G1_1475	iron ABC transporter substrate-binding protein	[P]
ISO4G1_1476	nitrogenase component 1 type oxidoreductase	[C]
ISO4G1_1478	nitrogenase iron protein NifH	[P]
ISO4G1_1482	glycosyl transferase GT8 family	[M]
ISO4G1_1483	hypothetical protein	Not in COGs
ISO4G1_1485	hypothetical protein	Not in COGs
ISO4G1_1495	hypothetical protein	Not in COGs
ISO4G1_1504	bifunctional phosphoglucose/phosphomannose isomerase	[M]
ISO4G1_1505	bifunctional phosphoglucose/phosphomannose isomerase	[M]
ISO4G1_1507	hypothetical protein	[S]
ISO4G1_1508	hypothetical protein	[S]
ISO4G1_1510	hypothetical protein	Not in COGs
ISO4G1_1511	hypothetical protein	Not in COGs
ISO4G1_1512	Sel1 domain-containing protein	[R]
ISO4G1_1519	cobaltochelatase CobN	[H]
ISO4G1_1521	hypothetical protein	[P]
ISO4G1_1522	hypothetical protein	Not in COGs
ISO4G1_1526	hypothetical protein	[S]
ISO4G1_1528	restriction endonuclease	[S]
ISO4G1_1529	helicase	[K]
ISO4G1_1530	PD-(D/E)XK nuclease superfamily protein	[Unclassified]
ISO4G1_1531	type I restriction-modification system M subunit HsdM	[V]
ISO4G1_1533	type I restriction-modification system S subunit HsdS	[V]
ISO4G1_1534	type I restriction-modification system S subunit HsdS	[V]
ISO4G1_1535	hypothetical protein	[R]
ISO4G1_1536	hypothetical protein	Not in COGs
ISO4G1_1537	signal peptidase I	[U]
ISO4G1_1538	hypothetical protein	Not in COGs
ISO4G1_1539	HIRAN domain-containing protein	[X]
ISO4G1_1541	hypothetical protein	Not in COGs
ISO4G1_1544	acetyl-CoA acetyltransferase	Not in COGs

only one gene from each gene family is represented.

Table A.4.14. Gene families unique to ISO4-G11

****

Locus_tag	Predicted gene product	COG category
ISO4G11_0013	TetR family transcriptional regulator	[K]
ISO4G11_0021	porphobilinogen deaminase	[H]
ISO4G11_0024	DNA-formamidopyrimidine glycosylase	[L]
ISO4G11_0025	hypothetical protein	Not in COGs
ISO4G11_0031	hypothtical protein	Not in COGs
ISO4G11_0033	zinc/iron permease	[P]
ISO4G11_0041	radical SAM protein	[R]
ISO4G11_0045	isochorismatase family protein amidases nicotinamidase -like protein	[Q]
ISO4G11_0051	UDP-*N*-acetylmuramoylalanine-D-glutamate ligase	[M]
ISO4G11_0057	hypothetical protein	[X]
ISO4G11_0058	Sel1 domain protein repeat-containing protein	[R]
ISO4G11_0059	nitroreductase	[C]
ISO4G11_0064	cobalamin synthase	[H]
ISO4G11_0074	AMP-binding enzyme	[I]
ISO4G11_0075	hypothetical protein	Not in COGs
ISO4G11_0076	hypothetical protein	Not in COGs
ISO4G11_0078	hypothetical protein	Not in COGs
ISO4G11_0082	hypothetical protein	Not in COGs
ISO4G11_0084	hypothetical protein	Not in COGs
ISO4G11_0085	sirohydrochlorin cobaltochelatase	[S]
ISO4G11_0088	hypothetical protein	Not in COGs
ISO4G11_0089	TetR family transcriptional regulator	[K]
ISO4G11_0090	FMN reductase	[R]
ISO4G11_0091	hypothetical protein	Not in COGs
ISO4G11_0092	hypothetical protein	Not in COGs
ISO4G11_0095	hypothetical protein	[X]
ISO4G11_0098	transposase	Not in COGs
ISO4G11_0104	hypothetical protein	Not in COGs
ISO4G11_0110	hypothetical protein	[N]
ISO4G11_0111	hypothetical protein	[R]
ISO4G11_0114	hypothetical protein	Not in COGs
ISO4G11_0116	pyruvate-formate lyase-activating enzyme	[O]
ISO4G11_0118	hypothetical protein	Not in COGs
ISO4G11_0126	hypothetical protein	Not in COGs
ISO4G11_0129	segregation and condensation protein A	[S]
ISO4G11_0131	hypothetical protein	Not in COGs
ISO4G11_0165	geranylgeranyl reductase family	[C]
ISO4G11_0171	hypothetical protein	[X]
ISO4G11_0187	transporter	[G]
ISO4G11_0200	hypothetical protein	[Unclassified]
ISO4G11_0206	cytosine deaminase	[F]
ISO4G11_0207	hypothetical protein	Not in COGs
ISO4G11_0208	hypothetical protein	Not in COGs
ISO4G11_0209	hypothetical protein	[S]
ISO4G11_0210	hypothetical protein	Not in COGs
ISO4G11_0213	periplasmic binding protein	[P]
ISO4G11_0217	cobaltochelatase subunit CobN	[H]
ISO4G11_0223	cell filamentation protein Fic	[S]
ISO4G11_0229	hypothetical protein	Not in COGs
ISO4G11_0235	ATPase	[R]
ISO4G11_0238	hypothetical protein	Not in COGs
ISO4G11_0239	hypothetical protein	Not in COGs
ISO4G11_0240	transposase	[L]
ISO4G11_0241	hypothetical protein	Not in COGs
ISO4G11_0242	hypothetical protein	Not in COGs
ISO4G11_0243	ribonuclease III	[K]
ISO4G11_0244	hypothetical protein	[X]
ISO4G11_0246	hypothetical protein	Not in COGs
ISO4G11_0248	hypothetical protein	Not in COGs
ISO4G11_0250	hypothetical protein	Not in COGs
ISO4G11_0251	hypothetical protein	Not in COGs
ISO4G11_0252	hypothetical protein	Not in COGs
ISO4G11_0253	hypothetical protein	Not in COGs
ISO4G11_0254	hypothetical protein	Not in COGs
ISO4G11_0256	hypothetical protein	[X]
ISO4G11_0260	toxin of toxin-antitoxin (TA) system	[S]
ISO4G11_0261	antitoxin PHD	[D]
ISO4G11_0262	toxin HipA	[R]
ISO4G11_0263	DNA-binding protein	[K]
ISO4G11_0264	hypothetical protein	Not in COGs
ISO4G11_0265	chromosome segregation protein SMC	[R]
ISO4G11_0266	hypothetical protein	Not in COGs
ISO4G11_0267	hypothetical protein	Not in COGs
ISO4G11_0269	hypothetical protein	Not in COGs
ISO4G11_0271	hypothetical protein	Not in COGs
ISO4G11_0275	hypothetical protein	[R]
ISO4G11_0279	cell filamentation protein Fic	[S]
ISO4G11_0284	hypothetical protein	[R]
ISO4G11_0288	hypothetical protein	[R]
ISO4G11_0289	hypothetical protein	Not in COGs
ISO4G11_0291	periplasmic binding protein	[P]
ISO4G11_0295	cobaltochelatase, CobN subunit	[H]
ISO4G11_0297	transposase	[L]
ISO4G11_0298	hypothetical protein	Not in COGs
ISO4G11_0299	hypothetical protein	[R]
ISO4G11_0300	Na^+-dependent transporter	Not in COGs
ISO4G11_0302	hypothetical protein	Not in COGs
ISO4G11_0306	TPR repeat protein, SEL1 subfamily	[R]
ISO4G11_0310	ATPase	R]
ISO4G11_0311	hypothetical protein	Not in COGs
ISO4G11_0316	hypothetical protein	Not in COGs
ISO4G11_0317	hypothetical protein	Not in COGs
ISO4G11_0318	multidrug transporter MatE	[V]
ISO4G11_0319	geranylgeranyl reductase	[C]
ISO4G11_0325	hypothetical protein	[X]
ISO4G11_0344	electron transfer flavoprotein	[C]
ISO4G11_0349	hypothetical protein	Not in COGs
ISO4G11_0352	cell wall/surface repeat protein	Not in COGs
ISO4G11_0354	hypothetical protein	[X]
ISO4G11_0355	hypothetical protein	[R]
ISO4G11_0361	periplasmic binding protein	[P]
ISO4G11_0365	hypothetical protein	Not in COGs
ISO4G11_0370	hypothetical protein	Not in COGs
ISO4G11_0371	hypothetical protein	[S]
ISO4G11_0373	hypothetical protein	Not in COGs
ISO4G11_0379	hypothetical protein	Not in COGs
ISO4G11_0380	hypothetical protein	Not in COGs
ISO4G11_0382	hypothetical protein	Not in COGs
ISO4G11_0383	hypothetical protein	Not in COGs
ISO4G11_0385	hypothetical protein	[R]
ISO4G11_0386	TPR repeat protein, SEL1 subfamily	[R]
ISO4G11_0390	hypothetical protein	Not in COGs
ISO4G11_0395	hypothetical protein	Not in COGs
ISO4G11_0403	hypothetical protein	[Unclassified]
ISO4G11_0405	transcriptional regulator	[K]
ISO4G11_0406	transporter	[G]
ISO4G11_0407	carbon starvation protein CstA	[T]
ISO4G11_0409	cell wall-binding protein	[S]
ISO4G11_0410	hypothetical protein	Not in COGs
ISO4G11_0413	TetR family transcriptional regulator	[K]
ISO4G11_0416	hypothetical protein	Not in COGs
ISO4G11_0418	hypothetical protein	Not in COGs
ISO4G11_0420	hypothetical protein	Not in COGs
ISO4G11_0424	Adhesin-like protein	[S]
ISO4G11_0432	phosphoglycolate phosphatase	[R]
ISO4G11_0436	hypothetical protein	Not in COGs
ISO4G11_0437	DNA helicase	[L]
ISO4G11_0440	hypothetical transmembrane protein	Not in COGs
ISO4G11_0442	group II intron reverse transcriptase/maturase	[L]

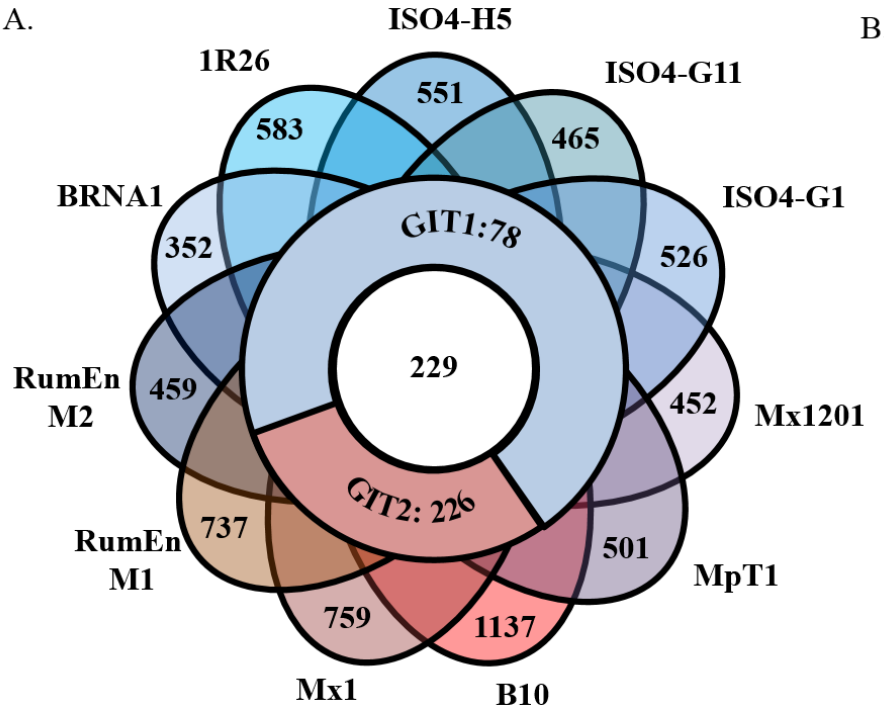
ISO4G11_0457	hypothetical protein	Not in COGs
ISO4G11_0462	ACP phosphodiesterase	[I]
ISO4G11_0465	Na^+/H^+ antiporter	[P]
ISO4G11_0469	multidrug transporter	Not in COGs
ISO4G11_0471	trimethylamine corrinoid protein 2	[R]
ISO4G11_0476	hypothetical protein	[Unclassified]
ISO4G11_0477	Trk system potassium uptake protein TrkA	[P]
ISO4G11_0478	Trk-type K^+ transport systems, membrane component	[P]
ISO4G11_0479	hypothetical protein	Not in COGs
ISO4G11_0480	hypothetical protein	Not in COGs
ISO4G11_0488	hypothetical protein	[R]
ISO4G11_0489	methyltransferase domain protein	[Q]
ISO4G11_0490	hypothetical protein	[Unclassified]
ISO4G11_0492	hypothetical protein	Not in COGs
ISO4G11_0493	hypothetical protein	Not in COGs
ISO4G11_0494	hypothetical protein	Not in COGs
ISO4G11_0495	hypothetical protein	Not in COGs
ISO4G11_0496	hypothetical protein	[S]
ISO4G11_0498	hypothetical protein	Not in COGs
ISO4G11_0501	ATPase AAA	[R]
ISO4G11_0506	regulatory protein TetR	[K]
ISO4G11_0507	hypothetical protein	Not in COGs
ISO4G11_0508	hypothetical protein	[X]
ISO4G11_0513	hypothetical protein	Not in COGs
ISO4G11_0514	hypothetical protein	[X]
ISO4G11_0515	ATPase AAA	[V]
ISO4G11_0529	ubiquinone/menaquinone biosynthesis methylase	[Q]
ISO4G11_0531	hypothetical protein	Not in COGs
ISO4G11_0532	oxidoreductase/nitrogenase component 1	[C]
ISO4G11_0533	peptide ABC transporter ATP-binding protein	[E]
ISO4G11_0534	peptide ABC transporter ATP-binding protein	[E]
ISO4G11_0535	peptide ABC transporter permease	[E]
ISO4G11_0536	peptide ABC transporter permease	[E]
ISO4G11_0537	peptide ABC transporter substrate-binding protein	[E]
ISO4G11_0539	hypothetical protein	Not in COGs
ISO4G11_0541	pyridoxamine 5'-phosphate oxidase	[R]
ISO4G11_0545	hypothetical protein	[T]
ISO4G11_0557	hypothetical protein	Not in COGs
ISO4G11_0559	hypothetical transmembrane protein	Not in COGs
ISO4G11_0560	adhesin-like protein	[M]
ISO4G11_0563	hypothetical protein	[S]
ISO4G11_0578	hypothetical protein	Not in COGs
ISO4G11_0579	hypothetical protein	[R]
ISO4G11_0580	nucleoside-diphosphate-sugar epimerases	[M]
ISO4G11_0583	dTDP-4-dehydrorhamnose 3,5-epimerase	[M]
ISO4G11_0585	glycosyltransferase group 2 family protein	[R]
ISO4G11_0590	ADP-L-glycero-D-manno-heptose-6-epimerase	[M]
ISO4G11_0591	RfaE, domain I	[M]
ISO4G11_0592	hypothetical protein	[M]
ISO4G11_0593	DNA helicase	[L]
ISO4G11_0596	TetR family transcriptional regulator	[K]
ISO4G11_0613	phospholipase D/ transphosphatidylase	[I]
ISO4G11_0624	hypothetical protein	[S]
ISO4G11_0626	ATPase	[R]
ISO4G11_0632	hypothetical transmembrane protein	Not in COGs
ISO4G11_0633	hypothetical transmembrane protein	Not in COGs
ISO4G11_0638	hypothetical transmembrane protein	Not in COGs
ISO4G11_0643	periplasmic binding protein	[P]
ISO4G11_0644	hypothetical protein	Not in COGs
ISO4G11_0649	hypothetical protein	Not in COGs
ISO4G11_0655	hypothetical protein	[X]
ISO4G11_0656	hypothetical protein	Not in COGs
ISO4G11_0657	hypothetical protein	[X]
ISO4G11_0658	hypothetical protein	Not in COGs
ISO4G11_0659	hypothetical protein	Not in COGs
ISO4G11_0660	hypothetical protein	Not in COGs
ISO4G11_0661	hypothetical protein	Not in COGs
ISO4G11_0680	hypothetical protein	Not in COGs
ISO4G11_0681	hypothetical protein	Not in COGs
ISO4G11_0682	hypothetical protein	Not in COGs
ISO4G11_0683	hypothetical protein	Not in COGs
ISO4G11_0684	hypothetical protein	[X]
ISO4G11_0687	chromosome segregation protein SMC	[R]
ISO4G11_0688	hypothetical protein	Not in COGs
ISO4G11_0689	hypothetical protein	Not in COGs
ISO4G11_0695	phosphoglucosamine mutase	[G]
ISO4G11_0705	cobaltochelatase subunit CobN	[H]
ISO4G11_0709	nitrogenase molybdenum-iron protein	[C]
ISO4G11_0710	hypothetical protein	Not in COGs
ISO4G11_0714	hypothetical protein	Not in COGs
ISO4G11_0719	hypothetical protein	Not in COGs
ISO4G11_0720	hypothetical protein	Not in COGs
ISO4G11_0721	hypothetical protein	[Unclassified]
ISO4G11_0722	hypothetical protein	[R]
ISO4G11_0724	hypothetical protein	Not in COGs
ISO4G11_0727	ATPase	[R]
ISO4G11_0732	hypothetical protein	Not in COGs
ISO4G11_0733	hypothetical protein	Not in COGs
ISO4G11_0736	hypothetical transmembrane protein	Not in COGs
ISO4G11_0741	conserved hypothetical	Not in COGs
ISO4G11_0742	hypothetical protein	Not in COGs
ISO4G11_0743	hypothetical protein	Not in COGs
ISO4G11_0745	hypothetical protein	[X]
ISO4G11_0747	hypothetical protein	[S]
ISO4G11_0748	packaged DNA stabilization protein gp26	[M]
ISO4G11_0749	hypothetical transmembrane protein	Not in COGs
ISO4G11_0750	type III restriction protein res subunit	[L]
ISO4G11_0752	hypothetical protein	Not in COGs
ISO4G11_0767	TetR family transcriptional regulator	[K]
ISO4G11_0768	protein tyrosine phosphatase	[T]
ISO4G11_0776	hypothetical protein	Not in COGs
ISO4G11_0785	hypothetical protein	Not in COGs
ISO4G11_0789	hypothetical transmembrane protein	Not in COGs
ISO4G11_0798	hypothetical transmembrane protein	[X]
ISO4G11_0811	hypothetical protein	Not in COGs
ISO4G11_0812	hypothetical protein	[L]
ISO4G11_0826	carbonic anhydrase	[P]
ISO4G11_0830	CAAX amino terminal protease family	Not in COGs
ISO4G11_0835	hypothetical protein	Not in COGs
ISO4G11_0837	DNA-methyltransferase (Dcm)	Not in COGs
ISO4G11_0839	H/ACA RNA-protein complex component Gar1	[J]
ISO4G11_0843	hypothetical protein	Not in COGs
ISO4G11_0851	hypothetical protein	Not in COGs
ISO4G11_0855	cell division GTPase	Not in COGs
ISO4G11_0867	CDP-glycerol:poly (glycerophosphate) glycerophosphotransferase	[M]

ISO4G11_0872	cell filamentation protein Fic	[S]
ISO4G11_0886	hypothetical transmembrane protein	Not in COGs
ISO4G11_0888	LPS biosynthesis protein	[M]
ISO4G11_0894	β-1,4-galactosyltransferase	[S]
ISO4G11_0896	hypothetical transmembrane protein	Not in COGs
ISO4G11_0897	hypothetical protein	[S]
ISO4G11_0902	guanylate kinase	Not in COGs
ISO4G11_0904	hypothetical protein	Not in COGs
ISO4G11_0907	hypothetical protein	Not in COGs
ISO4G11_0911	hypothetical transmembrane protein	Not in COGs
ISO4G11_0916	hypothetical protein	[R]
ISO4G11_0917	hypothetical protein	Not in COGs
ISO4G11_0922	hypothetical protein	Not in COGs
ISO4G11_0925	hypothetical protein	[S]
ISO4G11_0935	hypothetical protein	Not in COGs
ISO4G11_0943	hypothetical protein	Not in COGs
ISO4G11_0945	hypothetical protein	[S]
ISO4G11_0951	hypothetical protein	[S]
ISO4G11_0959	hypothetical protein	Not in COGs
ISO4G11_0961	hypothetical protein	[E]
ISO4G11_0965	ABC-2 type transporter	[R]
ISO4G11_0971	hypothetical transmembrane protein	Not in COGs
ISO4G11_0972	carbohydrate binding protein	[Unclassified]
ISO4G11_0973	nitroreductase	[C]
ISO4G11_0974	hypothetical protein	Not in COGs
ISO4G11_0975	hypothetical protein	Not in COGs
ISO4G11_0978	hypothetical protein	Not in COGs
ISO4G11_0981	GCN5 family acetyltransferase	[K]
ISO4G11_0983	hypothetical protein	Not in COGs
ISO4G11_0998	peptidyl-prolyl cis-trans isomerase	[O]
ISO4G11_1001	Listeria/Bacterioides repeat protein	Not in COGs
ISO4G11_1002	hypothetical protein	[R]
ISO4G11_1013	hypothetical protein	Not in COGs
ISO4G11_1025	hypothetical protein	Not in COGs
ISO4G11_1034	adhesin-like protein	[R]
ISO4G11_1035	hypothetical protein	[X]
ISO4G11_1045	hypothetical protein	Not in COGs
ISO4G11_1046	hypothetical protein	Not in COGs
ISO4G11_1048	hypothetical protein	Not in COGs
ISO4G11_1054	hypothetical protein	Not in COGs
ISO4G11_1061	hypothetical protein	Not in COGs
ISO4G11_1064	hypothetical protein	[R]
ISO4G11_1065	transglutaminase-like superfamily	[E]
ISO4G11_1074	flavodoxin	[E]
ISO4G11_1081	hypothetical transmembrane protein	Not in COGs
ISO4G11_1083	hypothetical transmembrane protein	Not in COGs
ISO4G11_1085	ATPase	[R]
ISO4G11_1086	hypothetical protein	Not in COGs
ISO4G11_1087	hypothetical protein	[U]
ISO4G11_1088	site-specific recombinase XerD	[L]
ISO4G11_1096	hypothetical protein	Not in COGs
ISO4G11_1097	hypothetical protein	Not in COGs
ISO4G11_1103	hypothetical protein	Not in COGs
ISO4G11_1116	hypothetical protein	Not in COGs
ISO4G11_1120	hypothetical protein	[Unclassified]
ISO4G11_1124	adhesin-like protein	Not in COGs
ISO4G11_1125	adhesin-like protein	Not in COGs
ISO4G11_1128	hypothetical protein	[X]
ISO4G11_1129	hypothetical protein	Not in COGs
ISO4G11_1131	hypothetical protein	Not in COGs
ISO4G11_1132	hypothetical protein	Not in COGs
ISO4G11_1133	hypothetical protein	Not in COGs
ISO4G11_1134	hypothetical protein	Not in COGs
ISO4G11_1139	hypothetical protein	[X]
ISO4G11_1140	hypothetical transmembrane protein	Not in COGs
ISO4G11_1142	MarR family transcriptional regulator	[K]
ISO4G11_1145	hypothetical protein	Not in COGs
ISO4G11_1146	hypothetical transmembrane protein	Not in COGs
ISO4G11_1154	molybdopterin biosynthesis MoeB protein	[H]
ISO4G11_1156	hypothetical protein	Not in COGs
ISO4G11_1157	hypothetical protein	Not in COGs
ISO4G11_1163	hypothetical protein	Not in COGs
ISO4G11_1164	hypothetical protein	Not in COGs
ISO4G11_1166	AAA ATPase	[K]
ISO4G11_1167	small-conductance mechanosensitive channel	[M]
ISO4G11_1174	hypothetical protein	Not in COGs
ISO4G11_1175	hypothetical protein	[L]
ISO4G11_1179	hypothetical protein	[X]
ISO4G11_1181	hypothetical protein	Not in COGs
ISO4G11_1185	hypothetical protein	[Unclassified]
ISO4G11_1188	hypothetical protein	Not in COGs
ISO4G11_1189	periplasmic binding protein	[P]
ISO4G11_1209	MATE efflux family protein	[V]
ISO4G11_1212	permease	[R]
ISO4G11_1216	hypothetical protein	Not in COGs
ISO4G11_1217	hypothetical protein	Not in COGs
ISO4G11_1219	site-specific recombinase, DNA invertase	[L]
ISO4G11_1220	hypothetical protein	Not in COGs
ISO4G11_1221	hypothetical protein	Not in COGs
ISO4G11_1222	hypothetical protein	Not in COGs
ISO4G11_1223	hypothetical protein	Not in COGs
ISO4G11_1226	acetyltransferase	[K]
ISO4G11_1228	periplasmic binding protein	[P]
ISO4G11_1231	hypothetical protein	Not in COGs
ISO4G11_1234	transposase	[L]
ISO4G11_1235	ATPase AAA	[O]
ISO4G11_1236	adhesin-like protein	[R]
ISO4G11_1237	hypothetical protein	Not in COGs
ISO4G11_1239	hypothetical protein	[S]
ISO4G11_1240	hypothetical protein	Not in COGs
ISO4G11_1241	hypothetical transmembrane protein	Not in COGs
ISO4G11_1242	hypothetical protein	[R]
ISO4G11_1246	hypothetical protein	Not in COGs
ISO4G11_1248	hypothetical protein	Not in COGs
ISO4G11_1249	hypothetical transmembrane protein	Not in COGs
ISO$G11_1262	hypothetical protein	Not in COGs
ISO4G11_1264	hypothetical protein	Not in COGs
ISO4G11_1265	hypothetical protein	Not in COGs
ISO4G11_1269	periplasmic binding protein	[P]
ISO4G11_1272	hypothetical protein	Not in COGs
ISO4G11_1287	transposase	[X]
ISO4G11_1289	hypothetical transmembrane protein	Not in COGs
ISO4G11_1298	hypothetical transmembrane protein	Not in COGs
ISO4G11_1310	hypothetical protein	Not in COGs
ISO4G11_1311	ABC-type Fe^{3+}-siderophore transport system, permease component	[P]
ISO4G11_1312	vitamin B12 ABC transporter, permease component BtuC	[P]
ISO4G11_1314	periplasmic binding protein	[P]
ISO4G11_1319	hypothetical protein	[E]
ISO4G11_1321	acetyltransferase	[K]
ISO4G11_1331	nitrilase	[R]
ISO4G11_1351	hypothetical transmembrane protein	[S]
ISO4G11_1353	hypothetical protein	[R]
ISO4G11_1354	hypothetical protein	Not in COGs
ISO4G11_1355	Tfp pilus assembly protein PilF	[R]
ISO4G11_1364	hypothetical transmembrane protein	Not in COGs

ISO4G11_1369	Sel1 domain protein repeat-containing protein	[R]
ISO4G11_1372	molecular chaperone	[O]
ISO4G11_1373	hypothetical protein	[Unclassified]
ISO4G11_1374	hypothetical protein	Not in COGs
ISO4G11_1375	hypothetical protein	[O]
ISO4G11_1376	molecular chaperone DnaK	[O]
ISO4G11_1382	adhesin-like protein	[S]
ISO4G11_1383	hypothetical protein	Not in COGs
ISO4G11_1384	hypothetical protein	[X]
ISO4G11_1385	site-specific recombinase XerD	[L]
ISO4G11_1387	Fic family protein	[D]
ISO4G11_1389	ABC transporter related protein	[P]
ISO4G11_1391	periplasmic binding protein	[P]
ISO4G11_1398	CobN component of cobalt chelatase involved in B12 biosynthesis	[H]
ISO4G11_1399	hypothetical transmembrane protein	[P]
ISO4G11_1414	4-oxalocrotonate tautomerase	[R]
ISO4G11_1423	helicase	[L]
ISO4G11_1436	hypothetical protein	Not in COGs
ISO4G11_1437	hypothetical protein	Not in COGs
ISO4G11_1438	glycosyl transferase family	[M]
ISO4G11_1439	glycosyltransferase WbsX family protein	[M]
ISO4G11_1440	ADP-heptose:LPS heptosyltransferase	[M]
ISO4G11_1443	hypothetical protein	Not in COGs
ISO4G11_1444	hypothetical protein	[Q]
ISO4G11_1445	hypothetical protein	Not in COGs
ISO4G11_1446	hypothetical protein	Not in COGs
ISO4G11_1447	hypothetical protein	Not in COGs
ISO4G11_1453	cell filamentation protein Fic	[D]
ISO4G11_1454	hypothetical protein	Not in COGs
ISO4G11_1455	hypothetical protein	Not in COGs
ISO4G11_1458	exonuclease SbcC	[L]
ISO4G11_1459	exonuclease SbcD	[L]
ISO4G11_1470	ABC transporter substrate-binding protein	[P]
ISO4G11_1480	alcohol dehydrogenase, class IV	[C]
ISO4G11_1488	nicotinate phosphoribosyltransferase	[H]
ISO4G11_1492	transcriptional regulator	[T]
ISO4G11_1494	hypothetical transmembrane protein	[S]
ISO4G11_1495	hypothetical transmembrane protein	[Unclassified]
ISO4G11_1504	Mg-dependent DNase	[L]
ISO4G11_1509	hypothetical protein	[R]
ISO4G11_1518	hypothetical protein	Not in COGs
ISO4G11_1522	hypothetical protein	[Unclassified]
ISO4G11_1523	hypothetical protein	[X]
ISO4G11_1524	hypothetical protein	Not in COGs
ISO4G11_1539	amino acid transporter	[E]
ISO4G11_1540	hypothetical protein	[X]
ISO4G11_1541	TetR family transcriptional regulator	[K]
ISO4G11_1542	adhesin-like protein	[R]
ISO4G11_1543	hypothetical protein	Not in COGs
ISO4G11_1545	IclR family transcriptional regulator	[K]
ISO4G11_1552	hypothetical protein	[X]
ISO4G11_1559	ATPase	Not in COGs
ISO4G11_1602	metal-binding protein	[S]
ISO4G11_1609	adhesin-like protein	[R]
ISO4G11_1623	hypothetical transmembrane protein	Not in COGs
ISO4G11_1628	site-specific recombinase XerD	[L]
ISO4G11_1630	hypothetical protein	Not in COGs
ISO4G11_1637	hypothetical protein	Not in COGs
ISO4G11_1639	hypothetical protein	Not in COGs
ISO4G11_1646	hypothetical protein	[X]
ISO4G11_1647	hypothetical protein	Not in COGs
ISO4G11_1654	isocitrate/isopropylmalate dehydrogenase	[C]
ISO4G11_1659	ATPase (AAA+ superfamily)	[R]
ISO4G11_1662	Rrf2 family transcriptional regulator	[K]
ISO4G11_1666	hypothetical transmembrane protein	[X]
ISO4G11_1667	hypothetical protein	[X]
ISO4G11_1681	hypothetical transmembrane protein	[S]
ISO4G11_1695	hypothetical protein	Not in COGs
ISO4G11_1697	hypothetical protein	Not in COGs
ISO4G11_1698	hypothetical protein	[J]
ISO4G11_1700	hypothetical protein	Not in COGs
ISO4G11_1701	hypothetical protein	Not in COGs
ISO4G11_1702	hypothetical protein	Not in COGs
ISO4G11_1703	hypothetical transmembrane protein	Not in COGs
ISO4G11_1707	hypothetical protein	Not in COGs
ISO4G11_1709	nitrogenase	[C]
ISO4G11_1710	hypothetical protein	Not in COGs
ISO4G11_1711	type 11 methyltransferase	[Q]
ISO4G11_1712	periplasmic binding protein	[P]
ISO4G11_1715	hypothetical protein	Not in COGs
ISO4G11_1717	ATPase	[R]
ISO4G11_1719	adhesin-like protein	[U]
ISO4G11_1720	hypothetical protein	[S]
ISO4G11_1723	hypothetical protein	Not in COGs
ISO4G11_1724	hypothetical protein	[Unclassified]
ISO4G11_1726	hypothetical protein	Not in COGs
ISO4G11_1727	hypothetical protein	Not in COGs
ISO4G11_1729	hypothetical protein	[S]
ISO4G11_1730	hypothetical protein	Not in COGs
ISO4G11_1731	hypothetical protein	Not in COGs
ISO4G11_1732	hypothetical protein	Not in COGs
ISO4G11_1734	hypothetical protein	[Unclassified]
ISO4G11_1736	hypothetical protein	Not in COGs
ISO4G11_1749	hypothetical protein	Not in COGs
ISO4G11_1750	hypothetical protein	[X]
ISO4G11_1754	hypothetical protein	Not in COGs

only one gene from each gene family is represented.

A.



B.

Genome	Gene families
ISO4-H5	1,732
ISO4-G1	1,448
ISO4-G11	1,666
1R26	1,432
BRNA1	1,508
Mx1201	1,611
MpT1	1,364
RumEnM1	1,846
RumEnM2	1,166
Mx1	1,767
B10	2,424

Figure A.4.5. Conserved and novel gene families among the 11 Methanomassiliicoccales genomes analysed. **A**. Venn diagram indicating the number of conserved gene families between completed genomes based on BLASTp analysis, using a 50% identity and 50% coverage cutoff. Regions that do not overlap with other genomes depict the number of unique gene families under the same criteria. **B**. Table listing the total number of gene families in each of the Methanomassiliicoccales genomes.

Table A.4.295. Genes involved in methanogenesis and energy generation in Methanomassiliicoccales

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Methyl CoM reductase											
mrtB	AR505_1399	ISO4G1_0619	ISO4G11_1686	TALC_00469	AOA80_07775	AOA81_01030	AUQ37_08455 (pseudo)	MMALV_03970	MMINT_15670	WP_019176771.1	Mpt1_c01980
mrtD	AR505_1398	ISO4G1_0620	ISO4G11_1685	TALC_00470	AOA80_07770	AOA81_01025	AUQ37_08450	MMALV_03980	MMINT_15660	WP_019176772.1	Mpt1_c01990
mrtG	AR505_1397	ISO4G1_0621	ISO4G11_1684	TALC_00471	AOA80_07765	AOA81_01020 (pseudo)	Not annotated, but present between _08450 and _08445	MMALV_03990	MMINT_15650	WP_019176773.1	Mpt1_c02000
mrtA	AR505_1396	ISO4G1_0622	ISO4G11_1683	TALC_00472	AOA80_07760 (pseudo)	AOA81_01015	AUQ37_08445	MMALV_04000	MMINT_15640	WP_019176774.1	Mpt1_c02010
mrt?	AR505_1395	ISO4G1_0623	ISO4G11_1682	TALC_00473	AOA80_07755	AOA81_01010	AUQ37_08440	MMALV_04010	MMINT_15640	WP_019176775.1	Mpt1_c02020
atwA	AR505_1392	ISO4G1_0626	ISO4G11_1678	TALC_00478	AOA80_07670	AOA81_00980	AUQ37_08400 (pseudo)	MMALV_04060	MMINT_06320	WP_019176782.1	Mpt1_c06090
mrtC	AR505_1391	ISO4G1_0627	ISO4G11_1677	TALC_00479	AOA80_07665	AOA81_00975	AUQ37_08395	MMALV_04070	MMINT_06330	WP_019176783.1	Mpt1_c06080
Methyl-compound: corrinoid methyltransferases											
mtaB1	AR505_0949	ISO4G1_1057	ISO4G11_0445	-	AOA80_00420	-	AUQ37_01190	MMALV_09470	MMINT_02820	WP_019176360.1	Mpt1_c05090
mtaB2	AR505_0951	-	ISO4G11_0654	-	AOA80_03340	-	-	-	MMINT_03160	WP_019177726.1	-
mtaB3	-	-	-	-	AOA80_00090 (pseudo)	-	-	-	MMINT_05920	WP_019178088.1	-
mtmB1	AR505_1327	ISO4G1_0654	ISO4G11_0009	TALC_01093	AOA80_07720 + AOA80_07725	-	AUQ37_06000 + AUQ37_06005	MMALV_11290	MMINT_08950	WP_019176305.1 +WP_019176306.1	Mpt1_c05730
mtmB2	AR505_1328	ISO4G1_1238	ISO4G11_0011	TALC_01094	AOA80_07680 + AOA80_07675	-	AUQ37_06010 + AUQ37_06015	MMALV_11300	MMINT_08870	WP_019176314.1 + WP_019176315.1	Mpt1_c05760
mtmB3	-	-	-	-	-	-	-	-	-	WP_019178553.1 +WP_019178554.1	-
mtmB4	-	-	-	-	-	-	-	-	-	WP_019178517.1 + WP_019178516.1	-
mtbB1	AR505_1332	ISO4G1_0478	ISO4G11_0551	TALC_00301	-	-	AUQ37_04330 + AUQ37_04335	MMALV_11370	MMINT_09020	WP_019178526.1 +WP_019178527.1	-
mtbB2	-	ISO4G1_0939	ISO4G11_1513	TALC_01103	-	-	AUQ37_05965 + AUQ37_05970	-	-	-	-
mtbB3	-	ISO4G1_1235	-	-	-	-	-	-	-	-	-
mttB1	AR505_0772	ISO4G1_0471	ISO4G11_0003	TALC_00305	-	-	AUQ37_05975 (pseudo), AUQ37_05980	MMALV_11360	MMINT_09010	WP_019178519.1 + WP_019178520.1	-
mttB2	-	ISO4G1_0945	-	TALC_00306	-	-	-	-	-	-	-
mttB3	-	ISO4G1_0947	-	TALC_01102	-	-	-	-	-	-	-
mttB4	-	ISO4G1_1231	-	-	-	-	-	-	-	-	-
mttB5	-	ISO4G1_1444	-	-	-	-	-	-	-	-	-

Methyltransferase corrinoid activation protein											
ramA	AR505_1320	ISO4G1_0653, ISO4G1_1056	ISO4G11_0018, ISO4G11_0448, ISO4G11_0650	TALC_01088	AOA80_07690 (pseudo)	AOA81_01050	AUQ37_06035, AUQ37_06040 (interrupted)	MMALV_11240, MMALV_09430, MMALV_13490	MMINT_08890, MMINT_05880, MMINT_15750	WP_019178502.1, WP_019176361.1, WP_019176312.1, WP_019176767.1, WP_026068898.1	Mpt1_c05140, Mpt1_c05740
Methylamine transporter											
mtmP	-	-	-	-	-	-	-	-	-	-	-
mtbP	AR505_1242, AR505_1331	ISO4G1_1219	ISO4G11_0753	TALC_01005	-	-	AUQ37_04890	MMALV_05040, MMALV_11390	MMINT_08970	-	-
mttP	AR505_0749	ISO4G1_1228	-	TALC_01100	-	-	AUQ37_05990	MMALV_11340	MMINT_08990	-	-
Methyl-compound corrinoid ptroteins											
mtaC1	AR505_0950	ISO4G1_1058	ISO4G11_0444	-	AOA80_00415	-	AUQ37_01185	MMALV_09460	MMINT_02810	WP_019176359.1	Mpt1_c05100
mtaC2	AR505_0952	-	ISO4G11_0653	-	AOA80_03335	-	-	-	MMINT_03150	WP_019177725.1	-
mtaC3	-	-	-	-	AOA80_00085	-	-	-	MMINT_05930	WP_019178087.1	-
mtmC1	AR505_1329	ISO4G1_1237	ISO4G11_0008	TALC_01095	-	-	AUQ37_05995	MMALV_11310	MMINT_08940	WP_019176307.1	Mpt1_c05770
mtmC2	-	-	-	-	-	-	-	MMALV_04350	MMINT_08880	WP_019176313.1	-
mtmC3	-	-	-	-	-	-	-	-	-	WP_019178555.1	-
mtmC4	-	-	-	-	-	-	-	-	-	WP_019178518.1	-
mtbC	AR505_1333	ISO4G1_0742	ISO4G11_0550	TALC_00300	-	-	AUQ37_04325	MMALV_11380	MMINT_09030	WP_019178525.1	-
mtbC2	-	ISO4G1_1236	ISO4G11_1514	TALC_01104	-	-	AUQ37_05960	-	-	-	-
mttC	AR505_0773	ISO4G1_1229	ISO4G11_0001	TALC_01101	-	-	AUQ37_05985 (pseudo)	MMALV_11350	MMINT_09000	WP_019178521.1	-
mtsB1	AR505_1067	-	ISO4G11_0447	-	AOA80_02370		AUQ37_08035	MMALV_13480	-	WP_019177198.1	-
mtsB2	-	-	-	-	-	-	-	-	-	WP_019177442.1	-
mtxC1	AR505_1330	ISO4G1_0251	-	-	-	-	-	MMALV_04220	MMINT_15770	WP_019176763.1	-
mtxC2	-	-	-	-	-	-	-	-	MMINT_04600	WP_019178649.1	-
mtxC3	-	-	-	-	-	-	-	-		WP_019177907.1	-
Methylcobalamin: and methylthiol: CoM methyltransferases											
mta/mtbA1	AR505_1035	ISO4G1_0538	ISO4G11_0086	TALC_00463	AOA80_07810	AOA81_01055	AUQ37_01300	MMALV_05210	MMINT_07210	WP_019176765.1	-
mta/mtbA2	AR505_1063	ISO4G1_0617	ISO4G11_0224	-	AOA80_08780	-	AUQ37_02825 + AUQ37_02830	MMALV_09620	MMINT_15760	WP_019177502.1	Mpt1_c01960
mta/mtbA3	AR505_1404	ISO4G1_1013	ISO4G11_1691	-	AOA80_03175	-	-	MMALV_13470	-	-	-
mta/mtbA4	AR505_1575	-	-	-	-	-	-	-	-	-	-
mtsA1	AR505_1066	-	ISO4G11_0446	-	AOA80_02375	-	AUQ37_08040	-	-	WP_019177199.1	-
mtsA2	-	-	-	-	-	-	-	MMALV_11470	-	WP_019177441.1	-
Methanogenesis Marker proteins											
mmp1	AR505_1405	ISO4G1_0640	ISO4G11_1392	TALC_00474	AOA80_07785 (pseudo)	AOA81_01035	AUQ37_08420 (pseudo)	MMALV_04020	MMINT_15690	WP_019176769.1	Mpt1_c06130

mmp 2	AR505_0715	ISO4G1_0818	ISO4G11_0329	TALC_00899	AOA80_06395	-	-	MMALV_08110	MMINT_11220	-	Mpt1_c04680
mmp 3	AR505_1390	ISO4G1_0628	ISO4G11_1676	TALC_00480	AOA80_07655 (pseudo)	AOA81_00970 (pseudo)	AUQ37_08380	MMALV_04080	MMINT_06350	WP_019176786.1	Mpt1_c06070
mmp 4	AR505_1417	ISO4G1_0453	ISO4G11_1281	TALC_00447	AOA80_03390 (pseudo)	AOA81_01180	AUQ37_02910	MMALV_03720	MMINT_13280	-	Mpt1_c01760
mmp 5	AR505_1388	ISO4G1_0630	ISO4G11_1674	TALC_00482	AOA80_07645	AOA81_00960	AUQ37_08370	MMALV_04100	MMINT_06370	WP_019176787.1	Mpt1_c06050
mmp 6	AR505_1389	ISO4G1_0629	ISO4G11_1675	TALC_00481	AOA80_07650	AOA81_00965	AUQ37_08375	MMALV_04090	MMINT_06360	WP_019176786.1	Mpt1_c06060
mmp 7	AR505_1385	ISO4G1_0633	ISO4G11_1671	TALC_00485	AOA80_07630	AOA81_00945	AUQ37_08355	MMALV_04130	MMINT_06400	-	Mpt1_c06020
mmp 8	AR505_1203	ISO4G1_1489	ISO4G11_0905	TALC_00121	AOA80_05695	AOA81_01625	AUQ37_03590	MMALV_01130	MMINT_05050	WP_019178690.1	Mpt1_c11370
mmp 9	-	-	-	-	-	-	-	-	-	-	-
mmp 10	-	-	-	-	-	-	-	-	-	-	-
mmp 11	AR505_1637	ISO4G1_0150	ISO4G11_0156	TALC_00205	AOA80_02825	AOA81_00780	AUQ37_00775	MMALV_01920	MMINT_01950	WP_019176187.1	Mpt1_c12700
mmp 12	-	-	-	-	-	-	-	-	-	-	-
mmp 13	AR505_0362	ISO4G1_0364	ISO4G11_0399	TALC_01251	AOA80_08455	-	AUQ37_06445 (pseudo)	MMALV_13210	MMINT_17230	WP_019176684.1	Mpt1_c08020
mmp 14	-	-	-	-	-	-	-	-	-	-	-
mmp 15	AR505_1387	ISO4G1_0631	ISO4G11_1673	TALC_00483	AOA80_07640	AOA81_00955	AUQ37_08365 (pseudo)	MMALV_04110	MMINT_06380	WP_019176788.1	Mpt1_c06040
mmp 16	AR505_0724	ISO4G1_0520	ISO4G11_0782	-	AOA80_07845	AOA81_03185	-	-	MMINT_14480	WP_019176502.1	-
mmp 16-2	-	-	-	-	-	-	-	-	-	WP_019178086.1	-
mmp 17	AR505_1386	ISO4G1_0632	ISO4G11_1672	TALC_00484	AOA80_07635	AOA81_00950	AUQ37_08360	MMALV_04120	MMINT_06390	WP_019176789.1	Mpt1_c06030
CoM biosynthesis											
comA	-	-	-	-	-	-	-	-	-	-	-
comB	-	-	-	-	-	-	-	-	-	-	-
comC	-	-	-	-	-	-	-	-	-	-	-
comD	-	-	-	-	-	-	-	-	MMINT_18030	WP_019178319.1	-
comE	-	-	-	-	-	-	-	-	MMINT_18040	WP_019178320.1	-
comDE	-	-	-	-	-	-	-	MMALV_07790	-	-	-
cysteate synthase	-	-	-	-	-	-	-	MMALV_07780	MMINT_17990	WP_019178318.1	-
aspartate aminotransferase	AR505_1666	ISO4G1_0120	ISO4G11_1529	TALC_00175	AOA80_03975	AOA81_01815	AUQ37_02955	MMALV_01630	MMINT_01710	WP_019178142.1	Mpt1_c13100
F_{430} biosynthesis											
corA	AR505_1041	-	-	TALC_01211	AOA80_08495	-	AUQ37_03770	MMALV_12560	MMINT_17270	WP_019176692.1	Mpt1_c07030
Heterodisulfide reductase											

hdrA	AR505_1479	ISO4G1_1281	ISO4G11_1403	TALC_00332	AOA80_02455	AOA81_00385	AUQ37_03270	MMALV_03030	MMINT_04810	WP_019177460.1	Mpt1_c10580
hdrB1	AR505_0274	ISO4G1_0352	ISO4G11_0773	TALC_01276	AOA80_00670	AOA81_04325	AUQ37_05955	MMALV_13660	MMINT_03230	WP_019177711.1	Mpt1_c08830
hdrB2	AR505_0679	-	ISO4G11_0083	TALC_01035	-	-	-	MMALV_07360	MMINT_05370	WP_019176125.1	-
hdrC1	AR505_0273	ISO4G1_0351	ISO4G11_0772	TALC_01277	AOA80_00665	AOA81_04320	AUQ37_05950	MMALV_13670	MMINT_03240	WP_019177712.1	Mpt1_c08840
hdrC2	-	-	-	-	-	-	-	-	-	WP_019176126.1	Mpt1_c02960
hdrD1	AR505_0040	ISO4G1_0085	ISO4G11_1012	TALC_01537	AOA80_10440	AOA81_02315	AUQ37_07515	MMALV_16610	MMINT_01020	WP_019178460.1	Mpt1_c00310
hdrD2	AR505_0168	ISO4G1_0217	ISO4G11_1159	TALC_01391	AOA80_05385	AOA81_02260	-	MMALV_14890	MMINT_02470	WP_019177852.1	Mpt1_c10580
hdrD3	-	-	-	-	-	-	-	-	-	WP_019177557.1	-
Methyl viologen hydrogenase											
mvhD1	AR505_1478	ISO4G1_1280	ISO4G11_1402	TALC_00333	AOA80_02450	AOA81_00390	AUQ37_03265	MMALV_03040	MMINT_04820	WP_019177459	Mpt1_c10570
mvhD2	-	-	-	-	AOA80_05200	-	-	-	MMINT_05200	WP_019176201.1	-
mvhD3	-	-	-	-	-	-	-	-	-	WP_019176130.1	-
mvhG	AR505_1477	ISO4G1_1279	ISO4G11_1401	TALC_00334	AOA80_02445	AOA81_00395	AUQ37_03260	MMALV_03050	MMINT_04830	WP_019177458.1	Mpt1_c10560
mvhA	AR505_1476	ISO4G1_1278	ISO4G11_1400	TALC_00335	AOA80_02440	AOA81_00400	AUQ37_03255	MMALV_03060	MMINT_04840	WP_019177457.1	Mpt1_c10550
Fpo-like complex											
fpoA	AR505_1633	ISO4G1_0154	ISO4G11_0152	TALC_00209	AOA80_02800	AOA81_00760	AUQ37_00795	MMALV_01955	MMINT_01985	WP_019176183.1	Mpt1_c12660
fpoB	AR505_1632	ISO4G1_0155	ISO4G11_0151	TALC_00210	AOA80_02795	AOA81_00755	AUQ37_00800	MMALV_01960	MMINT_02000	WP_019176182.1	Mpt1_c12650
fpoC	AR505_1631	ISO4G1_0156	ISO4G11_0150	TALC_00211	AOA80_02790	AOA81_00750	AUQ37_00805	MMALV_01970	MMINT_02010	WP_019176181.1	Mpt1_c12640
fpoD	AR505_1630	ISO4G1_0157	ISO4G11_0149	TALC_00212	AOA80_02785	AOA81_00745	AUQ37_00810	MMALV_01980	MMINT_02020	WP_019176180.1	Mpt1_c12630
fpoH	AR505_1629	ISO4G1_0158	ISO4G11_0148	TALC_00213	AOA80_02780	AOA81_00740 (pseudo)	AUQ37_00815 (pseudo)	MMALV_01990	MMINT_02030	WP_019176179.1	Mpt1_c12620
fpoI	AR505_1628	ISO4G1_0159	ISO4G11_0147	TALC_00214	AOA80_02775	AOA81_00735	AUQ37_00820	MMALV_02000	MMINT_02040	WP_019176178.1	Mpt1_c12610
fpoJ	AR505_1627	ISO4G1_0160	ISO4G11_0146	TALC_00215	AOA80_02770	AOA81_00730	AUQ37_00825	MMALV_02010	MMINT_02050	WP_019176177.1	Mpt1_c12600
Hypothetical protein	AR505_1626	ISO4G1_0161	ISO4G11_0145	TALC_00216	AOA80_02765	AOA81_00725	AUQ37_00830	MMALV_02020	MMINT_02060	WP_019176176.1	Mpt1_c12590
fpoK	AR505_1625	ISO4G1_0162	ISO4G11_0144	TALC_00217	AOA80_02760	AOA81_00720	AUQ37_00835	MMALV_02030	MMINT_02070	WP_019176175.1	Mpt1_c12580
fpoL	AR505_1624	ISO4G1_0163	ISO4G11_0143	TALC_00218	AOA80_02755	AOA81_00715	AUQ37_00840	MMALV_02040	MMINT_02080	WP_019176174.1	Mpt1_c12570
fpoM	AR505_1623	ISO4G1_0164	ISO4G11_0142	TALC_00219	AOA80_02750	AOA81_00710	AUQ37_00845	MMALV_02050	MMINT_02090	WP_019176173.1	Mpt1_c12560
fpoN	AR505_1622	ISO4G1_0165	ISO4G11_0141	TALC_00220	AOA80_02745	AOA81_00705	AUQ37_00850	MMALV_02060	MMINT_02100	WP_019176172.1	Mpt1_c12550
fpoFO	-	-	-	-	-	-	-	-	-	-	-
Energy conserving hydrogenase											
echA1	-	-	-	-	AOA80_10430	-	-	-	MMINT_01030	WP_019178471.1	
echA2	-	-	-	-	-	-	-	-	MMINT_17050	WP_019176386.1	
echB1	-	-	-	-	AOA80_11085	-	-	-	MMINT_01040	WP_019178472.1	
echB2	-	-	-	-	-	-	-	-	MMINT_17040	WP_019176385.1	
echC1	-	-	-	-	AOA80_11080	-	-	-	MMINT_01050	WP_019178473.1	
echC2	-	-	-	-	-	-	-	-	MMINT_17030	WP_019176384.1	
echD1	-	-	-	-	AOA80_11075	-	-	-	MMINT_01060	WP_019178474.1	
echD2	-	-	-	-	-	-	-	-	MMINT_17020	WP_019176383.1	
echE1	-	-	-	-	AOA80_11070	-	-	-	MMINT_01070	WP_019178475.1	
echE2	-	-	-	-	-	-	-	-	MMINT_17010	WP_019176382.1	
echF1	-	-	-	-	AOA80_11065 (pseudo)	-	-	-	MMINT_01080	WP_019178476.1	
echF2	-	-	-	-	-	-	-	-	MMINT_17000	-	
Hyf-like hydrogenase											
HyfB, D,F-	-	-	-	-	-	-	-	-	-	WP_026069091.1	Mpt1_c08810

like protein HyfB, D,F-like protein	-	-	-	-	-	AOA81_04545	-	-	-	WP_026068660.1	-
HyfC	-	-	-	-	AOA80_03040	AOA81_04550 (pseudo)	-	-	-	WP_019178464.1	Mpt1_c08800
HyfE-like protein	-	-	-	-	AOA80_03045	AOA81_04555	-	-	-	WP_019178465.1	Mpt1_c08790
HyfF	-	-	-	-	AOA80_03050	AOA81_04560	-	-	-	WP_019178466.1	Mpt1_c08780
HyfG	-	-	-	-	-	AOA81_04565 (pseudo)	-	-	-	WP_019178467.1	Mpt1_c08770
HyfH	-	-	-	-	-		-	-	-	-	-
HyfI	-	-	-	-	-	AOA81_04570	-	-	-	WP_019178469.1	Mpt1_c08760
HyfJ	-	-	-	-	-	-	-	-	-		-
Hydrogenase maturation											
hypA	AR505_1306	ISO4G1_0585	ISO4G11_1337	TALC_01114	-	-	AUQ37_00515	MMALV_11500	MMINT_08370	WP_026068853.1	Mpt1_c04050
hypB	AR505_0067	ISO4G1_1431	ISO4G11_1039	TALC_01509	AOA80_09555	AOA81_03535	AUQ37_04800 (pseudo)	MMALV_16320	MMINT_01720	WP_019178273.1	Mpt1_c13730
hypC	AR505_0194	ISO4G1_0271	ISO4G11_1611	TALC_01355	AOA80_01330	AOA81_03930	AUQ37_09040	MMALV_14480	MMINT_03830	WP_019177568.1	Mpt1_c00660
hypD	AR505_1307	ISO4G1_0584	ISO4G11_1336	TALC_01113	AOA80_04875	AOA81_05590 (pseudo)	-	MMALV_11490	MMINT_08360	WP_019177228.1	Mpt1_c04060
hypE	AR505_0080	ISO4G1_1394	ISO4G11_0716	TALC_01492	AOA80_04125	AOA81_03625	AUQ37_07370 (pseudo)	MMALV_16340	MMINT_01350	WP_026069035.1	Mpt1_c00120
hypF hydrogenase maturation protease	AR505_0475 AR505_0195	ISO4G1_0994 ISO4G1_0272	ISO4G11_1483 ISO4G11_1610	TALC_00534 TALC_01354	AOA80_07940 AOA80_01325	AOA81_03115 AOA81_03935	AUQ37_07745 AUQ37_09035	MMALV_04840 MMALV_14470	MMINT_06110 MMINT_03840	WP_019176478.1 WP_026068929.1	Mpt1_c02320 Mpt1_c00670
Na^+/H^+ transporter											
Na^+/H^+ antiporter family	AR505_0385, AR505_0507, AR505_0835	ISO4G1_0239, ISO4G1_0247, ISO4G1_0483, ISO4G1_1091, ISO4G1_1156, ISO4G1_1268	ISO4G11_0460, ISO4G11_0465, ISO4G11_1315, ISO4G11_1457	TALC_00558, TALC_00770, TALC_01225	AOA80_11175, AOA80_04995	AOA81_00030, AOA81_00125, AOA81_02805 (pseudo), AOA81_02810 (pseudo)	AUQ37_06680, AUQ37_03820	MMALV_05090, MMALV_08820, MMALV_12700	MMINT_07080, MMINT_08500, MMINT_13230, MMINT_15840	WP_019177045.1, WP_019176441.1, WP_019176331.1	Mpt1_c02540, Mpt1_c04220, Mpt1_c07860, Mpt1_c13630
A_1A_0-ATP synthase											
ahaH	AR505_1826	ISO4G1_0004	ISO4G11_0815	TALC_00004	AOA80_00805	AOA81_02535	AUQ37_01965	MMALV_00050	MMINT_00050	WP_019178382.1	Mpt1_c12340

ahaI	AR505_1825	ISO4G1_0005	ISO4G11_0816	TALC_00005	AOA80_00810	AOA81_02540	AUQ37_01960 (pseudo)	MMALV_00060	MMINT_00060	WP_019178381.1	Mpt1_c12330
ahaK	AR505_1824	ISO4G1_0006	ISO4G11_0817	TALC_00006	AOA80_00815	AOA81_02545	AUQ37_01955	MMALV_00070	MMINT_00070	WP_019178380.1	Mpt1_c12320
ahaE	AR505_1823	ISO4G1_0007	ISO4G11_0818	TALC_00007	AOA80_00820	AOA81_02550	AUQ37_01950	MMALV_00080	MMINT_00080	WP_019178379.1	Mpt1_c12310
ahaC	AR505_1822	ISO4G1_0008	ISO4G11_0819	TALC_00008	AOA80_00825	AOA81_02555	AUQ37_01945	MMALV_00090	MMINT_00090	WP_019178378.1	Mpt1_c12300
ahaF	AR505_1821	ISO4G1_0009	ISO4G11_0820	TALC_00009	AOA80_00830	AOA81_02560	AUQ37_01940	MMALV_00100	MMINT_00100	WP_019178377.1	Mpt1_c12290
ahaA	AR505_1820	ISO4G1_0010	ISO4G11_0821	TALC_00010	AOA80_00835	AOA81_02565 (pseudo)	AUQ37_01935	MMALV_00110	MMINT_00110	WP_019178376.1	Mpt1_c12280
ahaB	AR505_1819	ISO4G1_0011	ISO4G11_0822	TALC_00011	AOA80_00840	AOA81_02570	AUQ37_01930	MMALV_00120	MMINT_00120	WP_019178375.1	Mpt1_c12270
ahaD	AR505_1818	ISO4G1_0012	ISO4G11_0823	TALC_00012	AOA80_00845	AOA81_02575	AUQ37_01925	MMALV_00130	MMINT_00130	WP_019178374.1	Mpt1_c12260

*NCBI accession number is shown instead of locus_tag due to draft genome. # Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes.+indicates functional protein is translated from two genes merged by amber codon read-through. Methyl:CoM reductase subunit X (*mrtX*), methanol:corrinoid methyltransferase (*mtaB*), monomethylamine:corrinoid methyltransferase (*mtmB*), dimethylamine:corrinoid methyltransferase (*mtbB*), trimethylamine:corrinoid methyltransferase (*mttB*), methylamine methyltranferase corrinoid activation protein (*ramA*), monomethylamine permease (*mtmP*), dimethylamine permease (*mtbP*), trimethylamine permease (*mttP*), methanol corrinoid protein (*mtaC*), monomethylamine corrinoid protein (*mtmC*), dimethylamine corrinoid protein (*mtbC*), trimethylamine corrinoid protein (*mttC*), methylthiol corrinoid protein (*mtsB*), corrinoid protein with unknown substrate (*mtxC*), methyl:CoM methyltransferase (*mtaA*), bifunctional methylthiol:corrinoid methyltransferase (*mtsA*), methanogenesis marker protein (*mmp1-17*), (2*R*)-phospho-3-sulfolactate synthase (*comA*), 2-phosphosulfolactate phosphohydrolase (*comB*), (2*R*)3-sulfolactate dehydrogenase (*comC*), sulfopyruvate decarboxylase (*comDE*), uroporphyrinogen-III *C*-methyltransferase (*corA*), heterodisulfide reductase (*hdr*), methylviologen hydrogenase (*mvh*), $F_{420}H_2$ dehydrogenase-like complex subunit (*fpoA-O*), energy-conserving hydrogenase (*echA-F*), Hyf-like hydrogenase (*hyf*), A_1A_0 ATP synthase subunits are present (*ahaA-K*).

Table A.4.306. Genes involved in central carbon metabolism in Methanomassiliicoccales

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Gluconeogenesis											
pgm	AR505_1608	ISO4G1_1503	ISO4G11_0127	TALC_00241	AOA80_09650	AOA81_01905	AUQ37_00975 (pseudo)	MMALV_02250	MMINT_16490	WP_019178100.1	Mpt1_c11280
pgi	AR505_0560, AR505_1769	ISO4G1_1504, ISO4G1_1505	ISO4G11_0584	TALC_00585	AOA80_10780	AOA81_01605 (pseudo), AOA81_05095	AUQ37_06535 (pseudo)	MMALV_01010, MMALV_05310	MMINT_07950	WP_019177252.1	-
tpiA	AR505_1148	ISO4G1_0734	ISO4G11_0602	TALC_00680	AOA80_09265	AOA81_04980	AUQ37_01925 (pseudo)	MMALV_09930	MMINT_11880	WP_026068833.1	Mpt1_c07300
fbp	AR505_1149	ISO4G1_0735	ISO4G11_0603	TALC_00681	AOA80_09260	AOA81_04985	AUQ37_07155	MMALV_09940	MMINT_11890	WP_026068834.1	Mpt1_c07310
pfkB	-	-	-	-	-	-	-	-	-	WP_019176285.1	
fba	AR505_0932, AR505_0508	ISO4G1_1397	ISO4G11_1452, ISO4H11_0601	TALC_00560, TALC_00670	AOA80_11330, AOA80_10200	AOA81_06970	AUQ37_06665 (pseudo)	MMALV_05100	MMINT_14400	WP_019176284.1	Mpt1_c09850, Mpt1_c10430
gap2	AR505_0154	ISO4G1_0260	ISO4G11_1178	TALC_01406	AOA80_06285	AOA81_02175	AUQ37_09135	MMALV_15110	MMINT_02260	WP_019177881.1	Mpt1_c00900
pgk	AR505_0155	ISO4G1_0261	ISO4G11_1177	TALC_01405	AOA80_06265, AOA80_11340	AOA81_02180	AUQ37_09140	MMALV_15100	MMINT_02310	WP_019177877.1	Mpt1_c00890
gpmA	AR505_0942	ISO4G1_0737	ISO4G11_0453	TALC_00943	AOA80_11335	-	AUQ37_01310	MMALV_09640			
apgM	AR505_0474	ISO4G1_0995	ISO4G11_1484	TALC_00533	AOA80_07955	AOA81_03110, AOA81_05970	AUQ37_07750	MMALV_04820	MMINT_12550	WP_019176885.1	Mpt1_c02310
eno	AR505_0470	ISO4G1_0594	ISO4G11_1491	TALC_00528	AOA80_10720	AOA81_05650	AUQ37_07770 (pseudo)	MMALV_04790	MMINT_08060	WP_026068854.1	Mpt1_c03940
ppsA	AR505_0472, AR505_1104	ISO4G1_0443, ISO4G1_0509	ISO4G11_0575	TALC_01169	AOA80_11320	AOA81_02890 (pseudo)	AUQ37_00215	MMALV_12140	MMINT_13110	WP_019177575.1	Mpt1_c07690
porA	AR505_0431	ISO4G1_0440	ISO4G11_0518	TALC_01166	AOA80_06560, AOA80_05885, AOA80_01355 (pseudo)	AOA81_02905	AUQ37_00235	MMALV_12090	MMINT_03660	WP_019177572.1	Mpt1_c07640
porB	AR505_0432	ISO4G1_0439	ISO4G11_0519	TALC_01165	AOA80_05880, AOA80_01350, AOA80_06555	AOA81_02910 (pseudo)	AUQ37_00240 (pseudo)	MMALV_12080	MMINT_03670	WP_019177571.1	Mpt1_c07630
porC	AR505_0429	ISO4G1_0442	ISO4G11_0516	TALC_01168	AOA80_06570 AOA80_01365 (pseudo)	AOA81_02895	AUQ37_00225	MMALV_12110	MMINT_03640	WP_026068932.1	Mpt1_c07660
porD	AR505_0430	ISO4G1_0441	ISO4G11_0517	TALC_01167	AOA80_01360, AOA80_06565	AOA81_02900	AUQ37_00230	MMALV_12100	MMINT_03650	WP_019177573.1	Mpt1_c07650
acs/acdA	AR505_1282	ISO4G1_1025	ISO4G11_0077	TALC_01130	AOA80_00190	AOA81_03020	AUQ37_00410	MMALV_11710	MMINT_06050	WP_019176444.1	Mpt1_c02450
ppcA	-	-	-	-	AOA80_04255	-	-	-	MMINT_00470	WP_019176213.1	

Incomplete TCA cycle											
gltA	AR505_0499, AR505_0678	ISO4G1_0688	ISO4G11_0693, ISO4G11_1468	-	AOA80_05790	-	-	-	MMINT_18640	WP_019178659.1	-
acnA	AR505_0592	ISO4G1_0573	ISO4G11_0272	TALC_00589	-	AOA81_06915	AUQ37_04625	MMALV_05560	MMINT_07370	-	Mpt1_c02920
Idh (NAD^+)	AR505_1376	ISO4G1_1012	ISO4G11_1654	TALC_00501	AOA80_06985	AOA81_04485	AUQ37_08765	MMALV_04330	-	WP_019176816.1	Mpt1_c03670
Idh ($NADP^+$)	AR505_0531	ISO4G1_0567	ISO4G11_0277	TALC_00573	-	AOA81_03260	AUQ37_00567	-	MMINT_05170	WP_026068719.1	Mpt1_c02660, Mpt1_c03670
korA	-	ISO4G1_0862	ISO4G11_0987	TALC_00345	AOA80_07230	AOA81_05180	-	-	MMINT_08420	WP_019176337.1	-
korB	-	ISO4G1_0863	ISO4G11_0986	TALC_00344	AOA80_07225	AOA81_05175	-	-	MMINT_08430	WP_019176336.1	-
korG	-	ISO4G1_0864	ISO4G11_0985	TALC_00343	-	AOA81_05170 (pseudo)	-	-	MMINT_08440	-	-
korD	-	ISO4G1_0861	-	TALC_00346	AOA80_07235	AOA81_05185	-	-	MMINT_08410	WP_019176338.1	-
sucC	-	-	-	-	-	-	-	-	-	WP_019177398.1	-
sucD	-	-	-	-	-	-	-	-	-	WP_019177399.1	-
sdhA	-	-	-	-	-	-	-	-	-	-	-
sdhB	-	-	-	-	-	-	-	-	-	-	-
sdhC	-	-	-	-	-	-	-	-	-	-	-
sdhD	-	-	-	-	-	-	-	-	-	-	-
tfrA	-	-	-	-	-	-	-	-	-	WP_019176936.1	-
tfrB	-	-	-	-	-	-	-	-	-	WP_019178460.1	-
nfrA	-	-	-	-	-	-	-	-	-	-	-
nfrB	-	-	-	-	-	-	-	-	-	-	-
fumA	AR505_0022	ISO4G1_0065	ISO4G11_0790	TALC_01553	AOA80_10615	AOA81_02405	AUQ37_02080 (pseudo)	MMALV_16760	MMINT_19270	WP_019178423.1	Mpt1_c14310
fumB	AR505_0023	ISO4G1_0066	ISO4G11_0791	TALC_01554	AOA80_10610	AOA81_02400	AUQ37_02085 (pseudo)	MMALV_16770	MMINT_19280	WP_019178422.1	Mpt1_c14320
mae1	AR505_1780	ISO4G1_0047	ISO4G11_0866	TALC_00057		AOA81_02790	AUQ37_08155	MMALV_00650	MMINT_19680	WP_019178295.1	Mpt1_c11760
purA	AR505_1168	ISO4G1_0684	ISO4G11_0612	TALC_00689	AOA80_09160 (pseudo)	AOA81_05040 (pseudo)	AUQ37_07105 (pseudo)	MMALV_10030	MMINT_12990	WP_019177117.1	Mpt1_c06950
purB	AR505_0742	ISO4G1_0911	ISO4G11_1050	TALC_00921	AOA80_11375 (pseudo)	AOA81_05825	AUQ37_01480	MMALV_07730	MMINT_10940	WP_019177003.1	Mpt1_c04990
serC	AR505_1665	ISO4G1_0121	ISO4G11_1530	TALC_00176	AOA80_03930	AOA81_01820	AUQ37_00625 (pseudo)	MMALV_01640	MMINT_01870	WP_026069026.1	Mpt1_c13090
Pentose phosphate pathway											
tktA N'	AR505_1649	ISO4G1_0136	ISO4G11_0168	TALC_00193	AOA80_06125	AOA81_03775	AUQ37_00715	MMALV_01800	MMINT_17780	WP_019176117.1	Mpt1_c12820
tktA C'	AR505_1648	ISO4G1_0137	ISO4G11_0167	TALC_00194	AOA80_06130	AOA81_03780	AUQ37_00720	MMALV_01810	MMINT_17770	WP_019176116.1	Mpt1_c12810
rpe	AR505_0461	ISO4G1_0602	ISO4G11_1502	TALC_00519	AOA80_06835	AOA81_05705	AUQ37_08010 (pseudo)	MMALV_04660	MMINT_07420	WP_019176846.1	Mpt1_c03860
rpiA	AR505_1513	ISO4G1_1303	ISO4G11_0890, ISO4G11_1427	TALC_00307	AOA80_00160	AOA81_03850	AUQ37_06850	MMALV_00820, MMALV_02800	MMINT_05570	WP_019176373.1	Mpt1_c10780
prs	AR505_1685	ISO4G1_1451	ISO4G11_0932	TALC_00147	AOA80_05650	AOA81_01710	AUQ37_02260	MMALV_01410	MMINT_18540	WP_019178734.1	Mpt1_c13340
Reductive acetyl-CoA pathway											
fdhAB	-	-	-	-	-	-	-	-	-	-	-

fhs	AR505_1217	ISO4G1_0964	ISO4G11_1362	TALC_00725	AOA80_02845 (pseudo)	AOA81_06540	AUQ37_05415 (pseudo)	MMALV_10510	MMINT_00330	WP_019176194.1	Mpt1_c06480
folD	AR505_1639	ISO4G1_0148	ISO4G11_0158	TALC_00203	AOA80_02840 (pseudo)	AOA81_00790	AUQ37_00765	MMALV_01900	MMINT_16400	WP_019176192.1	Mpt1_c12720
metV	-	-	-	-	AOA80_04305 (pseudo)	-	-	-	MMINT_00380	WP_019176202.1	-
metF	-	-	-	-	AOA80_04300	AOA81_00805	-	-	MMINT_00390	WP_019176203.1	-
Methylene tetrahydro folate reductase (Fdx)	-	-	-	-	-	-	-	-	-	-	-
acsE	-	-	-	-	AOA80_04280	AOA81_00810	-	-	MMINT_00400	WP_019176206.1	-
acsA	-	-	-	-	AOA80_10155 (pseudo)	AOA81_00590	-	-	MMINT_19140	WP_019176239.1	-
acsB	-	-	-	-	AOA80_09460	AOA81_06000	-	-	-	-	Mpt1_c02980
cdhA	-	-	-	-	-	-	-	-	-	-	-
cdhB	-	-	-	-	-	-	-	-	-	-	-
cdhC	-	-	-	-	-	-	-	-	-	-	-
cdhD	-	-	-	-	AOA80_10150 (pseudo)	AOA81_00595	-	-	MMINT_19130	WP_019176238.1	-
cdhE	-	-	-	-	AOA80_10145 (pseudo)	AOA81_00600	-	-	MMINT_19120	WP_019176237.1	-
Other enzymes of interest											
adh	AR505_0483	ISO4G1_0362	ISO4G11_1480	TALC_00541	-	-	AUQ37_07725	MMALV_04880	-	-	-
ldhA	-	-	-	TALC_01217	AOA80_03925	AOA81_01825	AUQ37_02340	MMALV_05640	-	-	-
ald	AR505_1599	ISO4G1_0190	ISO4G11_0117	TALC_00254	AOA80_03010	AOA81_04480	-		-	-	-
deoA	AR505_1641 , AR505_0705	ISO4G1_0145 , ISO4G1_0953	ISO4G11_0103 , ISO4G11_0160	TALC_00201	AOA80_01145	AOA81_00815 (pseudo)	AUQ37_00755 (pseudo)	MMALV_01880	MMINT_17690	WP_019176103.1	Mpt1_c12740
rbcL	AR505_1642	ISO4G1_0144	ISO4G11_0161	TALC_00200	AOA80_01150 (pseudo)	AOA81_00820	AUQ37_00750	MMALV_01870	MMINT_17700	WP_019176104.1	Mpt1_c12750
e2b2	AR505_1643	ISO4G1_0143	ISO4G11_0162	TALC_00199	AOA80_01155	AOA81_00825	AUQ37_00745 (pseudo)	MMALV_01860	MMINT_17710	WP_019176105.1	Mpt1_c12760

*NCBI accession number is shown instead of locus_tag due to draft genome. # Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes. Phosphoglucomutase (*pgm*), phosphoglucose isomerase (*pgi*), triose-phosphate isomerase (TpiA), fructose-1,6-bisphosphatatase (*fbp*), phosphofructokinase (*pfkB*), fructose-bisphosphate aldolase (*fba*), glyceraldehyde-3-phosphate dehydrogenase (*gap2*), phosphoglycerate kinase (*pgk*), phosphoglycerate mutase (*apgM*), phosphopyruvate hydratase (*eno*), phosphoenolpyruvate synthase (*ppsA*), pyruvate:Fdx oxidoreductase (*porABCD*), acetyl CoA synthetase (*acs/acdA*), phosphoenolpyruvate carboxylase (*ppcA*), aconitate hydratase (*acnA*), isocitrate dehydrogenase (*idh*), 2-oxoglutarate synthase (*korABDG*), succinyl-CoA synthetase (*sucCD*), succinate:quinone oxidoreductase (*sdhABCD*), thiol:fumarate reductase (*tfrAB*), NADPH-FMN oxidoreductase (*nfrAB*), fumarate hydratase (*fumAB*), NAD-dependent malic enzyme (*mae1*), adenylosuccinate synthetase (*purA*), adenylosuccinate lyase (*purB*), phosphoserine aminotransferase/aspartate aminotransferase (*serC/aat*), transketolase (*tktA*), ribulose-phosphate 3-epimerase (*rpe*), ribose-5-phosphate isomerase (*rpiA*), ribose-phosphate diphosphokinase (*prs*), NADP-dependent formate dehydrogenase (*fdhAB*), formate-tetrahydrofolate ligase (*fhs*), NADP-dependent methylene tetrahydrofolate dehydrogenase (*folD*), 5,10-methylenetetrahydrofolate reductase (*metVF*), methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (*acsE*), CO dehydrogenase (*cdhABCDE*), acetyl-CoA decarbonylase (*acsAE*), alcohol dehydrogenase (*adh*), lactate dehydrogenase (*ldhA*), aldehyde dehydrogenase (*ald*), AMP phosphorylase (*deoA*), ribose-1,5-bisphosphate isomerase (*e2b2*), ribulose 1,5-bisphosphate carboxylase (*rbcL*).

Table A.4.317. Genes involved in amino acid metabolism in Methanomassiliicoccales

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Glutamate											
gdhA	AR505_0145	ISO4G1_0129	ISO4G11_1197	TALC_01422	AOA80_04200 (pseudo)	-	AUQ37_02735	MMALV_06300	-	WP_026068644.1	-
Glutamine											
glnA	AR505_0091	ISO4G1_1386	ISO4G11_1073	TALC_01482	AOA80_00710	AOA81_00865	AUQ37_07285	MMALV_16060	MMINT_03850	WP_026068923.1	Mpt1_c00250
amtB	AR505_0205	ISO4G1_1253	ISO4G11_1600	TALC_00998	AOA80_01865	-	AUQ37_05600	MMALV_14360	MMINT_08960	WP_019178560.1	Mpt1_c05750
Arginine											
argJ	AR505_0674	ISO4G1_1106	ISO4G11_0428	TALC_00630	AOA80_03500 (pseudo)	AOA81_05420	AUQ37_02785	MMALV_06150	MMINT_07650	WP_019177288.1	Mpt1_c03210
argB	AR505_0675	ISO4G1_1105	ISO4G11_0429	TALC_00631	AOA80_03505	AOA81_05415 (pseudo)	AUQ37_02780	MMALV_06160	MMINT_07660	-	Mpt1_c03220
argC	AR505_0673	ISO4G1_1107	ISO4G11_0427	TALC_00629	AOA80_03495	AOA81_05425	AUQ37_02790	MMALV_06140	MMINT_07640	WP_019177289.1	Mpt1_c03200
arg5,6	-	-	-	-	-	-	-	-	-	-	-
argD	AR505_0676	ISO4G1_1104	ISO4G11_0430	TALC_00632	AOA80_03510	AOA81_05410	AUQ37_02775	MMALV_06170	MMINT_07670	WP_019177286.1	Mpt1_c03230
argF	AR505_1401	ISO4G1_0613	ISO4G11_1688	TALC_00466	AOA80_02905	AOA81_01080	AUQ37_02810	MMALV_03950	MMINT_15810	WP_019176758.1	Mpt1_c01930
argG	AR505_0672	ISO4G1_1108	ISO4G11_0423	TALC_00628	AOA80_03490	AOA81_05430	AUQ37_02795	MMALV_06120	MMINT_07630	WP_019177290.1	Mpt1_c03190
argH	AR505_0671	ISO4G1_1109	ISO4G11_0422	TALC_00627	AOA80_03485	AOA81_05435	AUQ37_02800 (pseudo)	MMALV_06110	MMINT_07620	WP_019177291.1	Mpt1_c03180
Proline											
proA	AR505_0108	ISO4G1_1376	ISO4G11_0919	TALC_01468	AOA80_06140	AOA81_00650	AUQ37_04015	MMALV_15640	MMINT_03600	WP_019177916.1	Mpt1_c10170
proB	AR505_0109	ISO4G1_1375	ISO4G11_0920	TALC_01467	AOA80_06145	AOA81_00645	AUQ37_04010	MMALV_15630	MMINT_03590	WP_019177915.1	Mpt1_c10160
proC	AR505_1634	ISO4G1_0153	ISO4G11_0153	TALC_00208	AOA80_02810	AOA81_00765	AUQ37_00790	MMALV_01950	MMINT_01980	WP_019176184.1	Mpt1_c12670
putP	AR505_0287	ISO4G1_0569	ISO4G11_1006	TALC_01357	AOA80_01965	AOA81_06055	AUQ37_06365	MMALV_13380	MMINT_01840	WP_019178133.1	Mpt1_c04660
Asparagine/ alanine											
ast	AR505_1666	ISO4G1_0120	ISO4G11_1529	TALC_00175, TALC_00438	AOA80_03280	AOA81_01815	AUQ37_02955	MMALV_01630, MMALV_03620	MMINT_01710, MMINT_07580	WP_019178142.1, WP_019177332.1	Mpt1_c01680, Mpt1_c13100
aat	AR505_1199	ISO4G1_0877	ISO4G11_0227	TALC_00711	AOA80_10820	AOA81_05120	AUQ37_06995	MMALV_10320	MMINT_07890	WP_026068856.1	Mpt1_c06640
asnB	AR505_0736, AR505_1303	ISO4G1_0511, ISO4G1_0590	ISO4G11_1076, ISO4G11_1341	TALC_00744, TALC_01117	-	AOA81_05605 AOA81_00025	AUQ37_00495	MMALV_03160, MMALV_11530	MMINT_07720, MMINT_08200	WP_019177232.1	Mpt1_c03990, Mpt1_c03330
Histidine											
hisA	AR505_0497	ISO4G1_0667	ISO4G11_1464	TALC_00552	AOA80_07885	AOA81_03155	AUQ37_07665	MMALV_05000	MMINT_14550	WP_026068714.1	Mpt1_c02390
hisB	AR505_0488	ISO4G1_1137	ISO4G11_1476	TALC_00545	AOA80_07890	AOA81_03145	AUQ37_07705	MMALV_04920	MMINT_14560	WP_019176491.1	Mpt1_c02380
hisC	AR505_0495	ISO4G1_0665	ISO4G11_1466	TALC_00550	AOA80_07875	AOA81_03165	AUQ37_07675	MMALV_04980, MMALV_00840	MMINT_14530	WP_019176494.1	Mpt1_c02410
hisF	AR505_0487	ISO4G1_1136	ISO4G11_1477	TALC_00544	AOA80_07895 (pseudo)	AOA81_03140	AUQ37_07710	MMALV_04910	MMINT_14570	WP_026068713.1	Mpt1_c02370
hisG	AR505_0494	ISO4G1_0664	ISO4G11_1467	TALC_00549	AOA80_07870	AOA81_03170	AUQ37_07685	MMALV_04970	MMINT_14520	WP_019176495.1	Mpt1_c02420
hisH	AR505_0496	ISO4G1_0666	ISO4G11_1465	TALC_00551	AOA80_07880	AOA81_03160	AUQ37_07670	MMALV_04990	MMINT_14540	WP_026068715.1	Mpt1_c02400
hisI	AR505_0486	ISO4G1_1135	ISO4G11_1478	TALC_00543	AOA80_07900	AOA81_03135	AUQ37_07715	MMALV_04900	MMINT_14580	WP_019176489.1	Mpt1_c02360
hisJ	-	-	ISO4G11_0196	-	-	-	-	-	-		
hisD	AR505_1073	ISO4G1_0781	ISO4G11_1475	TALC_00752	AOA80_02095	AOA81_00230	AUQ37_05105 (pseudo)	MMALV_07480	MMINT_11150	WP_019176957.1	Mpt1_c02110

Serine											
serA	AR505_1664	ISO4G1_0122	ISO4G11_1531	TALC_00177	AOA80_03925	AOA81_01825	AUQ37_00630	MMALV_01650	MMINT_01880	-	Mpt1_c13080
serB	AR505_0073	ISO4G1_1420	ISO4G11_0706	TALC_01501	AOA80_11345	AOA81_03570 , AOA81_05830	AUQ37_01485	MMALV_16250	MMINT_01210	-	Mpt1_c13640
serC	AR505_1665	ISO4G1_0121	ISO4G11_1530	TALC_00176	AOA80_03930	AOA81_01820	AUQ37_00625 (pseudo)	MMALV_01640	MMINT_01870	WP_026069026.1	Mpt1_c13090
Cysteine											
cysE	AR505_1192	ISO4G1_0954	ISO4G11_0222	TALC_00713	AOA80_09095	AOA81_06490	AUQ37_06985	MMALV_10360	MMINT_08710	WP_026068840.1	Mpt1_c04500
cysK	AR505_0800	ISO4G1_0383	ISO4G11_0186	TALC_01462	AOA80_01530	AOA81_06485	AUQ37_08265	MMALV_06500	MMINT_11280	WP_019178685.1	Mpt1_c03020
cysM	AR505_0695	ISO4G1_1045	-	TALC_00651	AOA80_01540	-	-	-	-	WP_019177668.1	-
cys4	-	-	-	-	-	-	-	MMALV_09490	-	-	-
cys3	-	-	-	-	-	-	-	MMALV_09480	-	-	-
sepS	-	-	-	-	-	-	-	-	-	-	-
pscS	-	-	-	-	-	-	-	-	-	-	-
Glycine											
glyA	-	ISO4G1_0934	ISO4G11_0236	TALC_00697	AOA80_10770	AOA81_05090	-	MMALV_10170	MMINT_07960	WP_019177251.1	Mpt1_c07180
Chrorismate											
aroA	AR505_0512	ISO4G1_1401	ISO4G11_1448	TALC_00564	AOA80_03190	AOA81_06975	AUQ37_06645	MMALV_05140	MMINT_14440	WP_019176506.1	Mpt1_c09820
fba	AR505_0508 , AR505_0932	ISO4G1_1397 , ISO4G1_1494	ISO4G11_1452 , ISO4G11_0601	TALC_00560	AOA80_10200	AOA81_06970	AUQ37_06665 (pseudo)	MMALV_05100	MMINT_14400	WP_019176511.1	Mpt1_c09850 , Mpt1_c10430
aroB	AR505_0509	ISO4G1_1398	ISO4G11_1451	TALC_00561	AOA80_10205	AOA81_01375	AUQ37_06660	MMALV_05110	MMINT_14410	WP_019176509.1	Mpt1_c10440
aroF	AR505_1449	ISO4G1_1402	ISO4G11_0665	TALC_00394	AOA80_00970	AOA81_06965	AUQ37_03115 (pseudo)	MMALV_03310	MMINT_14450	WP_019176505.1	Mpt1_c09810
aroE	AR505_0510	ISO4G1_1399	ISO4G11_1450	TALC_00562	AOA80_02175	AOA81_06955	AUQ37_06655	MMALV_05120	MMINT_14420	-	Mpt1_c09840
aroL	-	-	-	-				-	-	-	-
aroK	AR505_0511	ISO4G1_1400	ISO4G11_1449	TALC_00563	AOA80_02170	AOA81_06960	AUQ37_06650	MMALV_05130	MMINT_14430	WP_019176507.1	Mpt1_c09830
Phenylalanine											
pheA	AR505_0515	ISO4G1_1403	ISO4G11_0281	TALC_00571	AOA80_03190	AOA81_01370	AUQ37_06640	MMALV_05150	MMINT_14600	WP_026068711.1	Mpt1_c07490
Tyrosine											
tyrA	AR505_1450	ISO4G1_1396	ISO4G11_0666	TALC_00393	AOA80_00975	AOA81_06980	AUQ37_03600	MMALV_03300	MMINT_14460	-	Mpt1_c10450
aro8	AR505_1429	ISO4G1_1204	ISO4G11_1173	TALC_00438			AUQ37_02955			-	Mpt1_c01680
Tryptophan											
trpA	AR505_1164	-	ISO4G11_1072	TALC_01440	-	-	AUQ37_08270	MMALV_09260	-	WP_019176707.1	Mpt1_c00530
trpB	AR505_1163	-	ISO4G11_1071	TALC_01439	AOA80_07575 (pseudo)	-	AUQ37_08265	MMALV_09270	MMINT_06500	WP_019176708.1	Mpt1_c00520
trpC (bifunctional)	-	-	-	-	-	-	-	-	-	WP_019176709.1	-
trpC	AR505_1161	-	ISO4G11_1069	TALC_01437	-	-	AUQ37_08255	MMALV_09290	-	-	Mpt1_c00500
trpF	AR505_1162	-	ISO4G11_1070	TALC_01438	-	-	AUQ37_08260	MMALV_09280	-	-	Mpt1_c00510
trpD	AR505_1160 , AR505_0997	-	ISO4G11_1068	TALC_01436	AOA80_05365	-	AUQ37_08250	MMALV_09300	MMINT_02540	WP_019176712.1	Mpt1_c00490
trpE	AR505_1158	-	ISO4G11_1066	TALC_01434	-	-	AUQ37_08240	MMALV_09320	-	WP_026068755.1	Mpt1_c00470
trpG	AR505_1159	-	ISO4G11_1067	TALC_01435	-	-	AUQ37_08245	MMALV_09310	-	WP_019177873.1	Mpt1_c00480
Lysine											

lysC	AR505_0160	ISO4G1_0207	ISO4G11_1170	TALC_01399	AOA80_06245	AOA81_02200	AUQ37_07580 (pseudo)	MMALV_15030	MMINT_02370	WP_019177873.1	Mpt1_c00850
metL	AR505_0293	-	-	-	AOA80_01575 (pseudo)	AOA81_04425	AUQ37_06415	-	-	WP_019176487.1	
asd	AR505_0491	ISO4G1_0661	ISO4G11_1473	TALC_00548	AOA80_00980	AOA81_01390	AUQ37_07690 (pseudo)	MMALV_04960	MMINT_14470	WP_019176503.1	Mpt1_c02440
dapA	AR505_0161	ISO4G1_0208	ISO4G11_1169	TALC_01398	AOA80_06235	AOA81_02210	AUQ37_07570 (pseudo)	MMALV_15000	MMINT_02390	WP_019177871.1	Mpt1_c00830
dapB	AR505_0728	ISO4G1_0796	ISO4G11_0081	TALC_01007	AOA80_06240 (pseudo)	AOA81_04815	AUQ37_07575	MMALV_08600	MMINT_02380	WP_019177872.1	Mpt1_c00840
dapL	AR505_0157	ISO4G1_0204	ISO4G11_1173	TALC_01402	AOA80_06260	AOA81_02185 (pseudo)	AUQ37_07595	MMALV_15060	MMINT_02340	WP_019177876.1	Mpt1_c00880
dapF	AR505_0158	ISO4G1_0205	ISO4G11_1172	TALC_01401	AOA80_06255	AOA81_02190	AUQ37_07590	MMALV_15050	MMINT_02350	WP_019177875.1	Mpt1_c00870
lysA	AR505_0159	ISO4G1_0206	ISO4G11_1171	TALC_01400	AOA80_06250	AOA81_02195	AUQ37_07585	MMALV_15040	MMINT_02360	WP_019177874.1	Mpt1_c00860
Pyrrolysine											
pylB	AR505_1324	ISO4G1_0658	ISO4G11_0015	TALC_00497	AOA80_07705	-	AUQ37_08740	MMALV_04280	MMINT_08920	WP_019178530.1, WP_019176309.1	Mpt1_c05690, Mpt1_c03070
pylC	AR505_1323	ISO4G1_0659	ISO4G11_0016	TALC_01090	AOA80_07700	-	AUQ37_06025 (pseudo)	MMALV_11270	MMINT_08910	WP_019178531.1, WP_026069110.1	Mpt1_c05680
pylD	AR505_1322	ISO4G1_0660	ISO4G11_0017	TALC_01089	AOA80_07695	-	AUQ37_06030	MMALV_11260	MMINT_08900	WP_019178556.1	Mpt1_c05670
Methionine											
metA	AR505_0694	ISO4G1_0381	ISO4G11_0185	TALC_00650			AUQ37_02535	MMALV_06490	MMINT_11300	-	Mpt1_c03450
metX					AOA80_01535 (pseudo)	-	-	-	-	WP_026069128.1	-
metZ	-	-	-	-		-	-	-	-	-	-
oah	AR505_0738, AR505_0799	ISO4G1_0382	ISO4G11_0049	TALC_00626, TALC_01463	AOA80_01530	-	AUQ37_04355	MMALV_06100	MMINT_11290	WP_019178666.1	Mpt1_c03170
metB	-	-	-	-	AOA80_01430	-	-	MMALV_09480	-	WP_026069129.1, WP_019178065.1	-
mdeA	-	-	-	-	-	-	-	MMALV_09500	-	-	-
metC	-	-	-	-		-	-	-	-	WP_019178701.1	-
metH	-	-	-	-	AOA80_05425, AOA80_05440	AOA81_01065	-	MMALV_04220	-	-	-
metE	-	-	-	-		-	-	-	-	-	-
mmuM	-	-	-	-	-	-	-	-	-	-	-
SAM cycle											
metK	AR505_0579	-	-	-	AOA80_03765		-	-	MMINT_07810	WP_019177266.1	-
dcm	AR505_0327, AR505_0339, AR505_0340	ISO4G1_0395	ISO4G11_1725	-	-	AOA81_00110, AOA81_04845	AUQ37_04255	-	MMINT_10170	-	Mpt1_c08470

ahcy	AR505_0547, AR505_1788	ISO4G1_0041	ISO4G11_0858	TALC_00076	AOA80_07050	AOA81_02745 (pseudo)	AUQ37_08195	MMALV_00570	MMINT_02630	WP_026069069.1	Mpt1_c11910
Threonine											
thrB	AR505_0610	ISO4G1_0580	ISO4G11_0957	TALC_00598	AOA80_11020	AOA81_06850	AUQ37_04275	MMALV_05660	MMINT_15310	WP_019176554.1	Mpt1_c02830
thrC	AR505_0611	ISO4G1_0581	ISO4G11_0958	TALC_00599	AOA80_11015	AOA81_06845	AUQ37_04270	MMALV_05670	MMINT_15300	WP_019176555.1	Mpt1_c02820
Branched chain leucine, valine, isoleucine											
ilvB	AR505_0152, AR505_0153, AR505_0542	-	ISO4G11_1180, ISO4G11_1182	TALC_01407, TALC_01409	AOA80_06290	AOA81_02165	AUQ37_01135, AUQ37_09130	MMALV_15130, MMALV_15170	MMINT_12080, MMINT_02270	WP_026068987.1	Mpt1_c00920
ilvN	AR505_1768	-	ISO4G11_0876	TALC_00065		AOA81_01530	AUQ37_08115	MMALV_00740	MMINT_12090	-	Mpt1_c11510
ilvC	AR505_0150	ISO4G1_1327	ISO4G11_1183	TALC_01411	AOA80_06300 (pseudo)	AOA81_02160 (pseudo)	AUQ37_01120	MMALV_15180	MMINT_02280	WP_019177885.1	Mpt1_c00940
ilvD	AR505_1462	-	ISO4G11_0667	TALC_00391	AOA80_02245	AOA81_07005	AUQ37_03125 (pseudo)	MMALV_03290	MMINT_14670	WP_019177408.1	Mpt1_c00760
ilvE	AR505_1767	ISO4G1_1502	ISO4G11_0877	TALC_00066	AOA80_05900	AOA81_01535	AUQ37_08110	MMALV_00750	MMINT_00560	WP_019176466.1	Mpt1_c11520
leuA	AR505_0631	ISO4G1_0375	ISO4G11_0979	TALC_00615	AOA80_09055	AOA81_04500	AUQ37_04460	MMALV_05860	MMINT_13330	WP_026068862.1	Mpt1_c02690
leuC	AR505_0632	-	-	-	AOA80_04060	AOA81_05500	AUQ37_04625	MMALV_05870	MMINT_13340	WP_019178159.1, WP_019177309.1	Mpt1_c02680
leuD	AR505_0633	-	-	-	AOA80_04055	AOA81_05495	AUQ37_04450	MMALV_05880	MMINT_13350	WP_019178158.1, WP_019177308.1	Mpt1_c02670
leuB	AR505_0634	ISO4G1_0567, ISO4G1_1012	ISO4G11_0277	TALC_00573, TALC_00501	AOA80_03430, AOA80_11525 (pseudo)	AOA81_03260, AOA81_05490	AUQ37_04445	MMALV_05890	MMINT_13360	WP_019177307.1	Mpt1_c02660, Mpt1_c03670
cimA	-	-	-	-	AOA80_03415	AOA81_03245	AUQ37_07620, AUQ37_06715	-	-	WP_026068862.1	Mpt1_c02550
sbat1	AR505_0285	ISO4G1_0485	ISO4G11_0755	-	-	-	-	-	-	-	-
Agmatine											
speA	AR505_0268	ISO4G1_0346	ISO4G11_0762	TALC_01282	AOA80_00635	AOA81_04280	AUQ37_05920	MMALV_13720	MMINT_03300	WP_019177718.1	Mpt1_c08900
Methionine salvage/spermidine biosynthesis											
speD	-	-	ISO4G11_0389	TALC_00987	AOA80_11220	AOA81_06205	AUQ37_02310	-	-	WP_026068816.1	Mpt1_c05830
speE	-	-	-	-				-	-		-
potA	-	-	-	TALC_00237, TALC_00239	AOA80_09640, AOA80_09635	AOA81_01895, AOA81_01890	AUQ37_00965, AUQ37_00960	MMALV_02230, MMALV_02220	MMINT_10520	WP_019178102.1, WP_019178103.1	-
potB	-	-	-	TALC_00236	-	AOA81_01885	AUQ37_00955	MMALV_02210	MMINT_10530	-	-
potC	-	-	-	-	-	-	-	-	-	-	-
potD	-	-	-	-	-	-	-	-	-	-	-
Methionine salvage											
mtnN	AR505_1118	ISO4G1_1493	ISO4G11_0050	TALC_00995	-	-	-	-	-	-	-
mtnK	-	-	-	-	-	-	-	-	-	-	-
mtnA	-	ISO4G1_0186	ISO4G11_0121	TALC_00247	AOA80_09670	AOA81_01930	-	MMALV_02310	-	-	Mpt1_c11230

mtnB	-	ISO4G1_0185	ISO4G11_0122	TALC_00246	AOA80_09665	AOA81_01925	AUQ37_07845	MMALV_02300	-	-	-
mtnC	-	-	-	-	-	-	-	-	-	-	-
mtnW	-	-	-	-	-	-	-	-	-	-	-
mtnX	-	-	-	-	-	-	-	-	-	-	-
mtnE	-	-	-	-	-	-	-	-	-	-	-
mtnP	AR505_0481	ISO4G1_1133	ISO4G11_1482	TALC_00535	AOA80_07930	AOA81_03125	AUQ37_07735	MMALV_04860	MMINT_14620	-	Mpt1_c02330
Aryl pyruvates salvage											
iorA	AR505_0999	ISO4G1_0837	ISO4G11_0640	TALC_00960	AOA80_06550	AOA81_06165	-	MMALV_09390	MMINT_12480	WP_019176899.1	Mpt1_c05230
iorB	AR505_0998	ISO4G1_0836	ISO4G11_0639	TALC_00961	AOA80_06545	AOA81_06160	-	MMALV_09380	MMINT_12490	WP_019176900.1	Mpt1_c05240
Amino acid permease											
Amino acid permease	AR505_0830	ISO4G1_0565	ISO4G11_1511	TALC_01421	-	-	AUQ37_07635	MMALV_06350	MMINT_03580	WP_049796369.1	-

*NCBI accession number is shown instead of locus_tag due to draft genome. # Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes. Glutamate dehydrogenase (*gdhA*), glutamine synthetase (*glnA*), ammonium transporter (*atmB*), ornithine acetyltransferase (*argJ*), acetylglutamate kinase (*argB*), *N*-acetylglutamylphosphate reductase (*argC*), acetylglutamate kinase/*N*-acetyl-γ-glutamyl-phosphate reductase (*arg5,6*), acetylornithine aminotransferase (*argD*), ornithine carbamoyltransferase (*argF*), argininosuccinate synthase (*argG*), argininosuccinase (*argH*), glutamate-5-semialdehyde dehydrogenase (*proA*), glutamate 5-kinase (*proB*), pyrroline-5-carboxylate reductase (*proC*), proline:Na^+ symporter (*putP*), aspartate amino transferase (*ast*), alanine aminotransferase (*aat*), asparagine synthase (*asnB*), [*N*-(5-phosphoribosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (*hisA*), IGP dehydratase (*hisB*), histidinol-phosphate aminotransferase (*hisC*), IGP synthase cycloligase subunit (*hisF*), ATP phosphoribosyl transferase (*hisG*), IGP synthase glutamine amidotransferase subunit (*hisH*), phosphoribosyl-ATP pyrophosphatase/phosphoribosyl-AMP cyclohydrolase (*hisI*), histidine-binding periplasmic protein precursor (*hisJ*), histidinal dehydrogenase (*hisD*), phosphoglycerate dehydrogenase (*serA*), phosphoserine phosphatase (*serB*), phosphoserine aminotransferase (*serC*), serine acetyltransferase (*cysE*), cysteine synthase subunit A (*cysK*), *O*-acetylserine sulfhydrylase (*cysM*), cystathionine β-synthase (*cys4*), cystathionine γ-lyase (*cys3*), *O*-phosphoseryl-tRNA ligase (*sepS*), Sep-tRNA:Cys-tRNA synthase (*pscS*), serine hydroxymethyltransferase (*glyA*), 3-phosphoshikimate-1-carboxyvinyltransferase (*aroA*), fructose 1,6-bisphosphate aldolase (*fba*), dehydroquinate synthase (*aroB*), chorismate synthase (*aroF/aroC*), shikimate dehydrogenase (*aroE*), shikimate kinase (*aroL/aroK*), prephenate dehydratase (*pheA*), prephenate dehydrogenase (*tyrA*), aromatic amino acid aminotransferase (*aro8*), tryptophan synthase (*trpAB*), indole-3-glycerol phosphate synthase (*trpC*), phosphoribosyl anthranilate isomerase (*trpF*), anthranilate phosphoribosyltransferase (*trpD*), anthranilate synthase (*trpEG*), aspartate kinase (*lysC*), bifunctional aspartate kianse/homoserine dehydroegnase (*metL*), dihydrodipicolinate synthase (*dapA*), 4-hydroxy-tetrahydrodipicolinate reductase (*dapB*), L,L-diaminopimelate aminotransferase (*dapL*), diaminopimelate epimerase (*dapF*), diaminopimelate decarboxylase (*lysA*), methylornithine synthase (*pylB*), (2*R*,3*R*)-3-methylornithyl-N^6-lysine synthase (*pylC*), pyrrolysine synthase (*pylD*), homoserine *O*-succinyltransferase (*metA*), homoserine *O*-acetyltransferase (*metX*), *O*-succinyl-L-homoserine sulfhydrylase (*metZ*), *O*-acetylhomoserine aminocarboxypropyltransferase (*oah*), *O*-succinylhomoserine(thiol)-lyase/*O*-succinylhomoserine lyase (*metB*), methionine γ-lyase (*mdeA*), cystathionine β-lyase (*metC*), cobalamin-independent methionine synthase (*metE*), methionine synthase (*metH*), homocysteine *S*-methyltransferase (*mmuM*), *S*-adenosylmethionine synthetase (*metK*), DNA-cytosine methyltransferase (*dcm*), *S*-adenosyl-L-homocysteinase (*ahcY*), homoserine kinase (*thrB*), threonine synthase (*thrC*), acetolactate synthase large subunit (*ilvB*), acetolactate synthase small subunit (*ilvN*), 2,3-dihydroxy-isovalerate:$NADP^+$ oxidoreductase (*ilvC*), dihydroxy-isovalerate dehydratase (*ilvD*), branched-chain amino acid aminotransferase (*ilvE*), 2-isopropylmalate synthase (*leuA*), isopropylmalate dehydratase (*leuCD*), 3-isoprolypmalate dehydrogenase (*leuB*), *R*-citramalate synthase (*cimA*), branched-chain amino acid transporter (*sbat1*), arginine decarboxylase (*speA*), agmatinase (*speB*), adenosylmethionine decarboxylase (*speD*), spermidine synthase (*speE*), putrescine/spermidine ABC transporter (*potABCD*), 5'-methylthioadenosine/*S*-adenosylhomocysteine nucleosidease (*mtnN*), 5'-methylthioribose kinase (*mtnK*), 5'-methylthioribose 1-phosphate isomerase (*mtnA*), methylthioribulose-1-phosphate dehydratase (*mtnB*), acireductone synthase (*mtnC*), 2,3-diketo-5-methylthiopentyl-1-phosphate enolase (*mtnW*), 2-hydroxy-3-keto-5-methylthiopentenyl-1-phosphate phosphatase (*mtnX*), 2-oxo-4-methylthiobutanoate-glutamine aminotransferase (*mtnE*), *S*-methyl-5'-thioadenosine phosphorylase (*mtnP*), indolepyruvate oxidoreductase (*iorAB*).

Table A.4.328. Genes involved in nucleotide biosynthesis in Methanomassiliicoccales

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Purine biosynthesis											
purB	AR505_0742	ISO4G1_0911	ISO4G11_1050	TALC_00921	AOA80_11375 (pseudo)	AOA81_05825	AUQ37_01480	MMALV_07730	MMINT_10940	WP_019177003.1	Mpt1_c04990
purF	AR505_0284	ISO4G1_0359	ISO4G11_0781	TALC_01266	AOA80_00720	AOA81_04370	AUQ37_06300 (pseudo)	MMALV_13560	MMINT_03370	WP_019177701.1	Mpt1_c08300
purP domain protein	AR505_0257, AR505_1667	ISO4G1_0335, ISO4G1_0119	ISO4G11_1551	TALC_00174, TALC_01295	AOA80_00545	AOA81_01810, AOA81_04225	AUQ37_05855	MMALV_01620, MMALV_13860	MMINT_02910, MMINT_01670	WP_019178145.1, WP_019177740.1	Mpt1_c09180, Mpt1_c13110
ade2	AR505_1804	ISO4G1_1515	ISO4G11_0841	TALC_00033	AOA80_05525	AOA81_02650	AUQ37_01825	MMALV_00360	MMINT_02590	WP_026069067.1	Mpt1_c12090
purD	AR505_0066	ISO4G1_1432	ISO4G11_1038	TALC_01510	AOA80_09550	AOA81_03530	AUQ37_04805 (pseudo)	MMALV_16330	MMINT_19510	WP_019178274.1	Mpt1_c13740
purH	AR505_1656	ISO4G1_0130	ISO4G11_0175	TALC_00185	AOA80_04270	AOA81_03735	AUQ37_00665 (pseudo)	MMALV_01730	MMINT_00450	WP_019176210.1	Mpt1_c12920
purC	AR505_0622	ISO4G1_0671	ISO4G11_0967	TALC_00471	AOA80_05080	AOA81_05535 (pseudo)	AUQ37_04510	MMALV_05760	MMINT_15090	WP_019176588.1	Mpt1_c11990
purM	AR505_0037	ISO4G1_0078	ISO4G11_1009	TALC_01540	AOA80_10465	AOA81_02340	AUQ37_02170	MMALV_16640	MMINT_19240	WP_026069089.1	Mpt1_c14130
purQ	AR505_1680	ISO4G1_0105	ISO4G11_0941	TALC_00152	AOA80_05685	AOA81_01740	AUQ37_02295	MMALV_01470	MMINT_18470	WP_019178718.1	Mpt1_c13280
purL	AR505_1681	ISO4G1_0104	ISO4G11_0940	TALC_00151	AOA80_05670 (pseudo)	AOA81_01735	AUQ37_02290	MMALV_01460	MMINT_18480	WP_019178719.1	Mpt1_c13290, Mpt1_c13310
purS	AR505_1682	ISO4G1_0103	ISO4G11_0939	TALC_00150	AOA80_05675	AOA81_01730	AUQ37_02285	MMALV_01450	MMINT_18490	WP_019178720.1	Mpt1_c13300
apt	AR505_0084, AR505_0301, AR505_1496	ISO4G1_1391, ISO4G1_1320	ISO4G11_0744	TALC_01490	-	AOA81_03635	AUQ37_07350, AUQ37_06900	MMALV_16130	MMINT_04670	-	Mpt1_c00150
purN NCAIR mutase (PurE)-related protein	AR505_0141	ISO4G1_1335	ISO4G11_1195	TALC_01418	AOA80_10665	AOA81_02135	AUQ37_01070	MMALV_15280	MMINT_03630	WP_019177929.1	Mpt1_c00990
	AR505_1595	ISO4G1_1354	ISO4G11_0112	TALC_00266	AOA80_09320 (pseudo)	AOA81_00205	AUQ37_06110	MMALV_02500	MMINT_10890	WP_026068828.1	Mpt1_c07930
Purine interconversion											
adk	AR505_0262	ISO4G1_0337	ISO4G11_0754	TALC_01289	AOA80_00570	AOA81_04250	AUQ37_05780	MMALV_13800	MMINT_02950	WP_019177734.1	Mpt1_c09070
purA	AR505_1168	ISO4G1_0684	ISO4G11_0612	TALC_00689	AOA80_09160 (pseudo)	AOA81_05040 (pseudo)	AUQ37_07105 (pseudo)	MMALV_10030	MMINT_12290	WP_019177117.1	Mpt1_c06950
add guaB	AR505_1340	ISO4G1_0550	ISO4G11_0525	TALC_01078	AOA80_08965	AOA81_06755	AUQ37_06110	MMALV_10930	MMINT_10450	WP_026068845.1	Mpt1_c06190
Pyrimidine biosynthesis											
pyrB	AR505_0437	ISO4G1_1149	ISO4G11_1349	TALC_01157	AOA80_00335	AOA81_02935	AUQ37_00280 (pseudo)	MMALV_12010	MMINT_05510	WP_026068695.1	Mpt1_c07580
pyrI	AR505_0436	ISO4G1_1150	ISO4G11_1348	TALC_01158	AOA80_00330	AOA81_02930	AUQ37_00275	MMALV_12020	MMINT_05500	WP_019176408.1	Mpt1_c07590
pyrC	AR505_0281	ISO4G1_0357	ISO4G11_0778	TALC_01269	AOA80_00705	AOA81_04355	AUQ37_06285 (pseudo)	MMALV_13600	MMINT_03400	WP_026068955.1	Mpt1_c08340
pyrD	AR505_1029	ISO4G1_1070	ISO4G11_0378	TALC_01017	AOA80_09045	AOA81_06435	AUQ37_02360 (pseudo)	MMALV_07220	MMINT_08570	WP_019177150.1	Mpt1_c05890

pyrK	AR505_1030	ISO4G1_1069	ISO4G11_0377	TALC_01016	AOA80_09050	AOA81_06440	AUQ37_02365	MMALV_07210	MMINT_08580	WP_019177149.1	Mpt1_c05900
pyrE	AR505_1615	ISO4G1_0172	ISO4G11_0135	TALC_00226	AOA80_03860	AOA81_01845	AUQ37_00910 (pseudo)	MMALV_02120	MMINT_01930	WP_019178120.1	Mpt1_c12480
pyrF	AR505_0060	ISO4G1_1436	ISO4G11_1033	TALC_01515	AOA80_07505	AOA81_03495	AUQ37_04830	MMALV_16380	MMINT_19560	WP_019178279.1	Mpt1_c13790
carB	AR505_0500 , AR505_0503	ISO4G1_0839	ISO4G11_1462	TALC_00553	AOA80_11600	AOA81_03230	AUQ37_07660	MMALV_05010	MMINT_06210	WP_019176516.1	Mpt1_c02580
carA	AR505_0501 , AR505_0504	ISO4G1_0840	ISO4G11_1461	TALC_00554	AOA80_11590	AOA81_03235	AUQ37_07655	MMALV_05020	MMINT_06200	WP_019176517.1	Mpt1_c02570
Pyrimidine interconversion											
pyrH	AR505_0251	ISO4G1_0330	ISO4G11_1557	TALC_01301	AOA80_00520	-	AUQ37_05825	MMALV_13920	MMINT_02860	WP_019177745.1	Mpt1_c09230
tmk	AR505_1102 , AR505_1690	ISO4G1_1455 , ISO4G1_1086	ISO4G11_0928	TALC_00143	AOA80_05625	AOA81_01685	AUQ37_02245	MMALV_01370	MMINT_18610	WP_026069141.1	Mpt1_c13380
thyX	AR505_0076	ISO4G1_1417	ISO4G11_0708	TALC_01498	AOA80_07340	AOA81_03585 (pseudo)	-	MMALV_16220	MMINT_01230	WP_026069046.1	Mpt1_c00060
trxA	AR505_0047	ISO4G1_0096	ISO4G11_1019	TALC_01529	AOA80_03085	AOA81_03405 , AOA81_03410	AUQ37_04905	MMALV_16520	MMINT_19530 , MMINT_17530	WP_019178276.1 , WP_019178572.1	Mpt1_c14010
trxB	AR505_0048 , AR505_1346	ISO4G1_0097 , ISO4G1_0547	ISO4G11_0538 , ISO4G11_1020	TALC_01069 , TALC_01528	AOA80_09000 , AOA80_03080	AOA81_06725 , AOA81_03415 (pseudo)	AUQ37_04900 (pseudo)	MMALV_16510 , MMALV_10860	MMINT_17540	WP_019178573.1 , WP_019177165.1	Mpt1_c14020
dcd	AR505_1287	ISO4G1_1032	ISO4G11_0071	TALC_01127	AOA80_08040	AOA81_03040	AUQ37_00445	MMALV_11680	MMINT_14300	WP_019176464.1	Mpt1_c02190
cmk	-	ISO4G1_0321	-	TALC_01309	AOA80_08760	AOA81_04145			MMINT_04580	WP_019177505.1	Mpt1_c09370
Others											
guaA	AR505_0706	ISO4G1_1061	ISO4G11_0198	TALC_01003	AOA80_05520	AOA81_02645	AUQ37_01810	-	MMINT_02570	-	-
guaAa	AR505_1803	ISO4G1_1514	ISO4G11_0842	TALC_00034	-	-	-	MMALV_00370	MMINT_02580	WP_019178350.1	Mpt1_c12080
guaAb	AR505_1802	ISO4G1_1513	ISO4G11_0844	TALC_00036	-	-	-	MMALV_00390	MMINT_02560	WP_019178347.1	Mpt1_c12060
ndk	AR505_1756	ISO4G1_1497	ISO4G11_0884	TALC_00115	AOA80_05850	AOA81_01595	AUQ37_03475	MMALV_01070	MMINT_19180	WP_026069122.1	Mpt1_c11420
gmk	-	-	-	-	-	-	-	-	-	-	-
prsA	AR505_1685	ISO4G1_1451	ISO4G11_0932	TALC_00147	AOA80_05650	AOA81_01710	AUQ37_02260	MMALV_01410	MMINT_18540	WP_019178734.1	Mpt1_c13340
rdgB	AR505_0187	ISO4G1_0263	ISO4G11_1618	TALC_01364	AOA80_00245	AOA81_03885	AUQ37_05555	MMALV_14540	MMINT_05400	WP_026068692.1	Mpt1_c00560
udk	-	-	-	-	-	-	-	-	-	-	-
mtnN	AR505_1118	ISO4G1_1493	ISO4G11_0050	TALC_00995	-	-	-	-	-	-	-
nrdJ	AR505_0964	ISO4G1_0875	ISO4G11_0491	TALC_00972	-	-	AUQ37_08310	MMALV_09160	MMINT_04250	-	Mpt1_c05370
comEB	AR505_1827	ISO4G1_0003	ISO4G11_0814	TALC_00003	AOA80_00800	AOA81_02530	AUQ37_01970	MMALV_00040	MMINT_00040	WP_019178383.1	Mpt1_c12350

*NCBI accession number is shown instead of locus_tag due to draft genome. # Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes. Bifunctional enzyme adenosuccinate lyase (*purB*), amidophosphoribosyl transferase (*purF*), 5-formaminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5'-monophosphate synthetase (*purP*), phosphoribosylaminoimidazole carboxylase (*ade2*), phosphoribosylamineglycine ligase (*purD*), bifunctional AICAR transformylase/IMP cyclohydrolase (*purH*), phosphoribosylaminoimidazole-succinocarboxamide synthase (*purC*), phosphoribosylformylglycinamide cyclo-ligase (*purM*), phosphoribosylformylglycinamidine synthetase subunit Q (*purQ*), phosphoribosylformylglycinamidine synthetase subunit L (*purL*), phosphoribosylformylglycinamidine synthetase subunit S (*purS*), adenine phosphoribosyltransferase (*apt*), phosphoribosylglycinamide formyltransferase (*purN*), N^5-carboxyaminoimidazole ribonucleotide mutase (*purE*), adenylate kinase (*adk*), adenylosuccinate synthetase (*purA*), adenosine deaminase (*add*), IMP dehydrogenase (*guaB*), aspartate transcarbamylase (*pyrB*), aspartate transcarbamylase regulatory subunit (*pyrI*), dihydroorotase (*pyrC*), dihydroorotate dehydrogenase (*pyrD*), dihydroorotate dehydrogenase electron transfer subunit (*pyrK*), orotate phosphoribosyltransferase (*pyrE*), orotidine-5'-phosphate decarboxylase (*pyrF*), carbamoyl phosphate synthase (*carAB*), UMP kinase (*pyrH*), thymidylate kinase (*tmk*), flavin-dependent thymidylate synthase (*thyX*), thioredoxin (*trxA*), thioredoxin-disulfide reductase (*trxB*), CMP/dCMP deaminase (*dcd*), cytidylate kinase (*cmk*), GMP synthase (*guaA/guaAab*), broad substrate specificity nucleoside diphosphate kinase (*ndk*), guanylate kinase (*gmk*), ribose-phosphate diphosphokinase (*prsA*), dITP/XTP pyrophosphoatase (*rdgB*), uridine kinase/cytidine kinase (*udk*), MTA/SAH nucleosidase (*mtnN*), ribonucleoside-diphosphate reductase (*nrdJ*), dCMP deaminase (*comEB*).

Table A.4.339. Genes involved in cell replication in Methanomassiliicoccales

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Chromosome replication											
Replicative DNA helicase Mcm	AR505_0018	ISO4G1_0062	ISO4G11_0794	TALC_01557	AOA80_08280 (pseudo)	AOA81_02425 (pseudo), AOA81_02420 (pseudo)	AUQ37_02065	MMALV_16800	MMINT_09980, MMINT_19320	-	Mpt1_c14350
Orc1/cdc6 family replication initiation protein	AR505_0001, AR505_1205	ISO4G1_0001, ISO4G1_0957	ISO4G11_0219, ISO4G11_0812,	TALC_00001, TALC_00716	AOA80_10380, AOA80_00790	AOA81_02520, AOA81_06510	AUQ37_01980, AUQ37_04975	MMALV_00010, MMALV_10400	MMINT_00010, MMINT_18840	WP_019178385.1, WP_019178317.1	Mpt1_c00010, Mpt1_c06580
Replication factor A	AR505_1652	ISO4G1_0118	ISO4G11_1527	TALC_00172	AOA80_04020	AOA81_01805	AUQ37_00610	MMALV_01600	MMINT_01650	WP_019178149.1	Mpt1_c13130
Replication factor C large subunit	AR505_0801	ISO4G1_0811	ISO4G11_0324	TALC_00887	AOA80_06710 (pseudo)	AOA81_04590 (pseudo)	AUQ37_01645	MMALV_08210	MMINT_12650	-	Mpt1_c04630
Replication factor C small subunit	AR505_1202	ISO4G1_0958	ISO4G11_0212	TALC_00718	AOA80_10735	AOA81_06520	AUQ37_04990 (pseudo)	MMALV_10430	MMINT_08020	WP_019177244.1	Mpt1_c06560
DNA polymerase family B	AR505_0130	ISO4G1_1347	ISO4G11_1207	TALC_01432	AOA80_09935 (pseudo)	AOA81_02070	AUQ37_07865 (pseudo)	MMALV_15420	MMINT_13800	WP_019177962.1	Mpt1_c01120
DNA polymerase II large subunit	AR505_1438	ISO4G1_1214	ISO4G11_1300	TALC_00430	AOA80_10405 (pseudo)	AOA81_01285 (pseudo), AOA81_01280 (pseudo)	AUQ37_03010 (pseudo)	MMALV_03540	MMINT_14000	WP_026068874.1	Mpt1_c01600
DNA polymerase II small subunit	AR505_1816	ISO4G1_0014	ISO4G11_0825	TALC_00014	AOA80_00850	AOA81_02585	AUQ37_01920 (pseudo)	MMALV_00150	MMINT_17980	WP_026069074.1	Mpt1_c12240
DNA polymerase sliding clamp subunit PCNA family Pcn	AR505_1650	ISO4G1_0135	ISO4G11_0169	TALC_00192	AOA80_06120	AOA81_03770	AUQ37_00710	MMALV_01790	MMINT_17790	WP_019176118.1	Mpt1_c12830
DNA primase DnaG	AR505_1142	ISO4G1_0729	ISO4G11_0623	TALC_00677	AOA80_09290	AOA81_04960	AUQ37_05285	MMALV_09870	MMINT_11650	WP_019177096.1	Mpt1_c07250
DNA primase large subunit PriB	AR505_1782	ISO4G1_0045	ISO4G11_0864	TALC_00054, TALC_00052	AOA80_08580 (pseudo)	AOA81_02770	AUQ37_08165 (pseudo)	MMALV_00630	MMINT_19050	-	Mpt1_c11830
DNA primase small subunit PriA	AR505_0009	ISO4G1_0054	ISO4G11_0804	TALC_01569	AOA80_08210	AOA81_02475	AUQ37_02020	MMALV_16880	MMINT_19390	WP_019178400.1	Mpt1_c14490
Ribonuclease HII RnhB	AR505_0209	ISO4G1_0287	ISO4G11_1597	TALC_01342	AOA80_01230 (pseudo)	AOA81_03980	AUQ37_05620	MMALV_14320	MMINT_04050	WP_019177553.1	Mpt1_c09760
Flap structure-specific endonuclease Fen	AR505_0460	ISO4G1_0603	ISO4G11_1503	TALC_00518	AOA80_06840	AOA81_05715 (pseudo)	AUQ37_08015 (pseudo)	MMALV_04650	MMINT_12970	WP_019176843.1	Mpt1_c03850
ATP-dependent DNA ligase	AR505_1277	ISO4G1_1019	ISO4G11_1358	TALC_01146	AOA80_08855	AOA81_02995	AUQ37_00360	MMALV_11800	MMINT_05290	WP_019176428.1	Mpt1_c07340
Genome segregation											

DNA topoisomerase I TopA	AR505_1299	ISO4G1_0592	ISO4G11_0062	TALC_01120	AOA80_10690	AOA81_05625	AUQ37_00475 (pseudo)	MMALV_11600	MMINT_08160	WP_019177234.1	Mpt1_c03970
DNA topoisomerase VI subunit A	AR505_0388	ISO4G1_0433	ISO4G11_0027	TALC_01216	AOA80_01455	AOA81_02825	AUQ37_03805	MMALV_12670	MMINT_14250	WP_019177592.1	Mpt1_c07830
DNA topoisomerase VI subunit B	AR505_0389	ISO4G1_0434	ISO4G11_0026	TALC_01215	AOA80_01450	AOA81_02830	AUQ37_03800 (pseudo)	MMALV_12660	MMINT_14240	WP_019177591.1	Mpt1_c07820
DNA gyrase subunit B GyrB	AR505_0033	ISO4G1_0073	ISO4G11_1004	-	AOA80_10545 (pseudo)	AOA81_02360	AUQ37_02145 (pseudo)	MMALV_16690	MMINT_18860	WP_019178436.1	Mpt1_c14200
DNA gyrase subunit A GyrA	AR505_0034	ISO4G1_0074	ISO4G11_1005	TALC_01544	AOA80_10555	AOA81_02365	AUQ37_02140 (pseudo)	MMALV_16680	MMINT_18850	-	Mpt1_c14190
Chromosome segregation and condensation protein ScpA	AR505_1610	ISO4G1_0176	ISO4G11_0129	TALC_00234	AOA80_09605	AOA81_01870	AUQ37_00940 (pseudo)	MMALV_02180	MMINT_16440	WP_019178109.1	Mpt1_c12420
Chromosome segregation and condensation protein ScpB	AR505_1609	ISO4G1_0177	ISO4G11_0128	TALC_00235	AOA80_09615	AOA81_01875	AUQ37_00950	MMALV_02190	MMINT_16450	WP_026069020.1	Mpt1_c12410
Chromosome segregation protein SMC	AR505_1611	ISO4G1_0175	ISO4G11_0130	TALC_00233	AOA80_09600 (pseudo)	AOA81_01860 , AOA81_01865	AUQ37_00935 (pseudo)	MMALV_02170	MMINT_16430	WP_026069021.1	Mpt1_c12430
Cell division											
Cell division protein FtsZ	AR505_0975	ISO4G1_0835	ISO4G11_0498	TALC_00964	AOA80_01585 , AOA80_01615	AOA81_06145 (pseudo)	AUQ37_08280	MMALV_09250	MMINT_17410 , MMINT_17400	WP 019176461.1	Mpt1_c05250 , Mpt1_c11940
CBS/parB domain-containing protein	AR505_0425	ISO4G1_0447	ISO4G11_0574	TALC_01172	AOA80_04760 (pseudo)	AOA81_01160 (pseudo)	AUQ37_02885	MMALV_12160	MMINT_13100	WP_019177577.1	-
ParA/MinD ATPase-like protein	AR505_1412	ISO4G1_0448	ISO4G11_1276	TALC_00452	AOA80_01375	AOA81_02885	AUQ37_00205	MMALV_03800	MMINT_13900	WP_019176627.1	-
Cell division GTPase	AR505_1792	ISO4G1_0038	ISO4G11_0855	TALC_00046	-	-	-	-	MMINT_11260	-	-
AAA family ATPase CDC48 subfamily protein	AR505_0052 , AR505_1810	ISO4G1_0020 , ISO4G1_0102	ISO4G11_0833 , ISO4G11_1024	TALC_00022 , TALC_01524	AOA80_05965 , AOA80_05565	AOA81_02690 , AOA81_03445	AUQ37_07920	MMALV_00240 , MMALV_00250 , MMALV_16470	MMINT_17110 , MMINT_14960 , MMINT_17620	WP_019176301.1 , WP_019178362.1 , WP_026069115.1	Mpt1_c14070 , Mpt1_c12160

*NCBI accession number is shown instead of locus_tag due to draft genome. # Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes.

Table A.4.340. Genes involved in cofactor biosynthesis in Methanomassiliicoccales

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Tetrapyrrole											
gltX	AR505_1121	ISO4G1_0936	ISO4G11_0692	TALC_00694	AOA80_09140	AOA81_05065	AUQ37_07065 (pseudo)	MMALV_10120	MMINT_12410	WP_019177132.1	Mpt1_c06970
hemA	AR505_1045	-	-	TALC_01207	AOA80_08475	-	AUQ37_03750	MMALV_12520	MMINT_17310	WP_026068749.1	Mpt1_c07070
hemB	AR505_1044	-	-	TALC_01208	AOA80_08480	-	AUQ37_03755	MMALV_12530	MMINT_17300	WP_026068750.1	Mpt1_c07060
hemC	AR505_1042	-	ISO4G11_0021	TALC_01210	AOA80_08490 (pseudo)	-	AUQ37_03765 (pseudo)	MMALV_12550	MMINT_17280	WP_026068751.1	Mpt1_c07040
hemD	AR505_1040	-	-	TALC_01212	AOA80_08500 (pseudo)	-	AUQ37_03775	MMALV_12570	MMINT_17260	WP_019176693.1	Mpt1_c07020
hemL	AR505_1043	-	-	TALC_01209	AOA80_08485	-	AUQ37_03760	MMALV_12540	MMINT_17290	WP_019176690.1	Mpt1_c07050
corA	AR505_1041	-	-	TALC_01211	AOA80_08495	-	AUQ37_03770	MMALV_12560	MMINT_17270	WP_019176692.1	Mpt1_c07030
cysG	AR505_1046	-	-	TALC_01249	AOA80_08470	AOA81_04390	AUQ37_06460	-	MMINT_17315	WP_019176687.1	Mpt1_c07080
Adenylosylcobalamin											
cbiX	AR505_0597	ISO4G1_0224	ISO4G11_0617	-	-	-		-	-	-	-
cbiL	AR505_0413	ISO4G1_0228	ISO4G11_0564	TALC_01192	-	-	AUQ37_00095	MMALV_12390	MMINT_01790	-	Mpt1_c07140
cbiH	AR505_0415	ISO4G1_0226	ISO4G11_0566	TALC_01190	-	-	AUQ37_00105	MMALV_12360	MMINT_01770	WP_026068753.1	Mpt1_c07120
cbiF	AR505_0414	ISO4G1_0227	ISO4G11_0545	TALC_01191	-	-	AUQ37_00100 (pseudo)	MMALV_12380	MMINT_01780	WP_019176698.1	Mpt1_c07130
cbiG	AR505_0379	ISO4G1_0230	ISO4G11_0036	TALC_01232	-	-	AUQ37_03850	MMALV_12760	MMINT_01810	WP_019176699.1	Mpt1_c07160
cbiD	AR505_0378	ISO4G1_0229	ISO4G11_0037	TALC_01233	-	-		MMALV_12770	MMINT_01800	WP_019176701.1	Mpt1_c07150
cbiJ	AR505_0380	ISO4G1_0231	ISO4G11_0035	TALC_01231	-	-	AUQ37_03845	MMALV_12750	MMINT_01820	-	Mpt1_c07170
cobL	-	-	-	-	-	-	-	-	-		-
cbiT	-	-	-	-	-	-	-	-	-	WP_026068752.1	-
cbiE	-	-	-	-	-	-	-	-	-	WP_019176697.1	-
cbiC	AR505_0416	ISO4G1_0225	ISO4G11_0567	TALC_01189	-	-	AUQ37_00110	MMALV_12350	MMINT_01760	WP_019177037.1	Mpt1_c07110
cbiA	AR505_0381, AR505_0363	ISO4G1_0232 , ISO4G1_0363	ISO4G11_0034 , ISO4G11_0400	TALC_01230, TALC_01250	AOA80_08510	AOA81_06795, AOA81_03105	-	MMALV_12740, MMALV_13200	MMINT_10700 , MMINT_17215	WP_019177213.1, WP_026068748.1	Mpt1_c05110, Mpt1_c08010
fldA	-	-	-	-	-	-	-	-	-	-	-
cobA	AR505_0444	ISO4G1_1062	ISO4G11_0512	TALC_00413	AOA80_09520	-	AUQ37_03105	MMALV_03350	-	WP_019176716.1	Mpt1_c03460
cbiP	AR505_1136, AR505_1137	ISO4G1_0690 , ISO4G1_0691	ISO4G11_0628 , ISO4G11_0629	TALC_01049, TALC_01050	AOA80_04815, AOA80_04820	AOA81_03100	-	MMALV_09820, MMALV_09830	MMINT_08250 , MMINT_08260	WP_019177214.1	Mpt1_c06670, Mpt1_c06680
cbiB	AR505_1295	ISO4G1_0999	ISO4G11_0067	TALC_01124	AOA80_08005	AOA81_03075	AUQ37_00455	MMALV_11640	MMINT_06180	WP_026068706.1	Mpt1_c02260
cobU	-	-	-	-	-	-	-	-	-	-	-
cobS/cobV	AR505_1297	-	ISO4G11_0064	TALC_01122	AOA80_07990	AOA81_03085	-	MMALV_11620	MMINT_06160	WP_019176471.1	Mpt1_c02280
cobC	-	-	-	-	AOA80_01000 (pseudo)	-	-	-	-	-	-
NAD^+ and $NADP^+$											
nadB	-	-	-	-	-	-	-	-	-	WP_019176936.1	-
nadA	AR505_1155	-	ISO4G11_0606	TALC_00684	AOA80_09185	AOA81_05000	AUQ37_07135	MMALV_09980	MMINT_12200	WP_026068837.1	Mpt1_c06910
pncB	AR505_0471	ISO4G1_0593	ISO4G11_1490	TALC_00529	AOA80_10715	AOA81_05645	AUQ37_06315	MMALV_13500	MMINT_08070	WP_019177240.1	Mpt1_c03950
nadC	AR505_1654	-	ISO4G11_0173	TALC_00188	AOA80_06090	AOA81_03755	AUQ37_00685 (pseudo)	MMALV_01750	MMINT_17820	WP_019176124.1	Mpt1_c12890
nadK	AR505_0121	ISO4G1_1365	ISO4G11_0697	TALC_01451	AOA80_09810	AOA81_02005	AUQ37_07960	MMALV_15550	MMINT_02690	WP_019177988.1	Mpt1_c01280

nadE	AR505_0377	ISO4G1_0386	ISO4G11_0377	TALC_01235	AOA80_01555	AOA81_04445	AUQ37_04040 (pseudo	MMALV_12810	MMINT_14860	WP_019177639.1	Mpt1_c07990
nadM	AR505_0616	ISO4G1_0899	ISO4G11_0619	TALC_00604	AOA80_10995	AOA81_05550 (pseudo)	-	MMALV_05730	MMINT_15265	WP_019176559.1	Mpt1_c02780
Riboflavin											
ribK	AR505_1679	ISO4G1_0106	ISO4G11_0942	TALC_00153	AOA80_05690	AOA81_01745	AUQ37_00550	MMALV_01480	MMINT_18460	WP_019178717.1	Mpt1_c13270
ribB	AR505_1180	ISO4G1_0933	ISO4G11_0234	TALC_00701	AOA80_10790	AOA81_05100	AUQ37_07045	MMALV_10180	MMINT_07940	WP_026068855.1	Mpt1_c07200
ribL	AR505_1181	ISO4G1_0932	ISO4G11_0233	TALC_00702	AOA80_10795	AOA81_05105	AUQ37_07040	MMALV_10190	MMINT_07930	WP_019177254.1	Mpt1_c07210
ribC	AR505_1182	ISO4G1_0931	ISO4G11_0232	TALC_00703	AOA80_10800	AOA81_05110	AUQ37_07035	MMALV_10200	MMINT_07920	WP_019177255.1	Mpt1_c07220
ribH	AR505_1183	ISO4G1_0930	ISO4G11_0231	TALC_00704	AOA80_10805	AOA81_05115	AUQ37_07030	MMALV_10210	MMINT_07910	WP_019177256.1	Mpt1_c07230
ribD	AR505_1602	ISO4G1_0188	ISO4G11_0120	TALC_00248	AOA80_01755	AOA81_01940	AUQ37_07835	MMALV_02330	MMINT_16300	WP_019178342.1	Mpt1_c11210
ribA	AR505_0210	ISO4G1_0288	ISO4G11_1596	TALC_01341	AOA80_01240	AOA81_03975	AUQ37_05625 (pseudo)	MMALV_14310	MMINT_04030	WP_019177554.1	Mpt1_c09750
Molybdenum cofactor											
moaA 1	AR505_1263	-	ISO4G11_1699	TALC_00802	-	-	-	MMALV_15830	-	-	-
moaA 2	AR505_1435	ISO4G1_1211	ISO4G11_1297	TALC_00433	AOA80_07265	AOA81_01255	AUQ37_02980	MMALV_03570	MMINT_13190	WP_026068867.1	Mpt1_c01630

*NCBI accession number is shown instead of locus_tag due to draft genome. [#] Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes. Glutamyl-tRNA synthetase (*gltX*). Glutamyl-tRNA reductase (*hemA*). Delta-aminolevulinic acid dehydratase (*hemB*). Porphobilinogen deaminase (*hemC*). Uroporphyrinogen III synthase (*hemD*). Glutamate-1-semialdehyde-2,1-aminomutase (*hemL*). Uroporphyrin-III *C*-methyltransferase (*corA*). Siroheme synthase (*cysG*). Sirohydrochlorin cobaltochelatase (*cbiX*). Cobalt-sirohydrochlorin (C^{20})-methyltransferase (*cbiL*). Cobalt-precorrin-3 (C^{17})-methyltransferase (*cbiH*). Cobalt precorrin-4 (C^{11})-methyltransferase (*cbiF*). Cobalt-precorrin 5A hydrolase (*cbiG*). Cobalt-precorrin-6A synthase (*cbiD*). Cobalt-precorrin-6A reductase (*cbiJ*). Precorrin-6B $C^{5,15}$-methyltransferase (*cobL*). Cobalt-precorrin-6B (C^{15})-methyltransferase (*cbiT*). Cobalt-precorrin-7 (C^{5})-methyltransferase (*cbiE*). Cobalt-precorrin-8 methylmutase (*cbiC*). Cobyrinate *a,c*-diamide synthase (*cbiA*). Flavodoxin A (*fldA*). Cobalamin adenosyltransferase (*cobA*). Adenosyl-cobyrate synthase (*cbiP*). Adenosylcobinamide-phosphate synthase (*cbiB*). Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase (*cobU*). Adenosylcobalamin synthase (*cobS*). Adenosyl-cobalamin (5'-phosphate) synthase (*cobV*). α-ribazole-5'-phosphate phosphatase (*cobC*). Aspartate oxidase (*nadB*). Quinolinate synthase (*nadA*). Nicotinate phosphoribosyltransferase (*pncB*). Quinolinate phosphoribosyltransferase (*nadC*). NAD kinase (*nadK*). NH_3-dependent NAD^+ synthase (*nadE*). Nicotinamide-mononucleotide adenylyltransferase (*nadM*). Riboflavin kinase (*ribK*). 3,4-dihydroxy-2-butanone-4-phosphate synthase (*ribB*). FAD synthetase (*ribL*). Riboflavin synthase (*ribC*). 6,7-dimethyl-8-ribityllumazine synthase (*ribH*). Bifunctional diaminohydroxy phosphoribosylaminopyrimidine deaminase /5-amino-6-(5-phosphoribosylamino) uracil reductase (*ribD*). GTP cyclohydrolase II (*ribA*). Molybdopterin biosynthesis protein A (*moaA*).

Table A.4.351. Genes involved in secretion

	ISO4-H5	ISO4-G1	ISO4-G11[#]	BRNA1	RumEnM1	RumEnM2	1R26	Mx1201	Mx1	B10*	MpT1
Signal recognition riboprotein subunits											
SRP54	AR505_0758	ISO4G1_0902	ISO4G11_0348	TALC_00918	AOA80_11290	AOA81_05835 (pseudo)	AUQ37_01505	MMALV_07830	MMINT_10970	WP_019177011.1	Mpt1_c04960
SRP19	AR505_1473	ISO4G1_1274	ISO4G11_0677	TALC_00338	AOA80_02415	AOA81_00415	AUQ37_03240	MMALV_03090	MMINT_04930	WP_019177450.1	Mpt1_c10520
Signal recognition receptor											
ftsY	AR505_0006	ISO4G1_0051	ISO4G11_0807	TALC_01572	AOA80_08185	AOA81_02495	AUQ37_02005	MMALV_16910	MMINT_19350	WP_019178395.1	Mpt1_c14550
Signal peptide cleavage											
Signal sequence peptidase SPI	AR505_1828	ISO4G1_0002	ISO4G11_0813	TALC_00002	AOA80_00795 (pseudo)	AOA81_02525	AUQ37_01975	MMALV_00030	MMINT_00020	WP_019178258.1	Mpt1_c12360
TFPP-like signal peptidase	-	-	-	-	-	-	-	-	-	-	-
Signal peptide peptidease SPP	-	-	-	-	-	-	-	-	-	-	-
Sec-independent protein translocase											
tatC	-	-	-	-	-	-	-	-	-	-	-
Sec61 translocon											
secG	AR505_0386	ISO4G1_0481	ISO4G11_0029	TALC_01224	AOA80_01490	AOA81_02815	AUQ37_03815	MMALV_12690	MMINT_06015	WP_019177600.1	Mpt1_c07850
secY	AR505_0241	ISO4G1_0319	ISO4G11_1567	TALC_01311	AOA80_08750 (pseudo)	AOA81_04135	AUQ37_05770	MMALV_14030	MMINT_04560	WP_019177507.1	Mpt1_c09390
secE	AR505_1799	ISO4G1_0032	ISO4G11_0848	TALC_00039	-	AOA81_02630	AUQ37_01800	MMALV_00430	MMINT_17390	WP_026069057.1	Mpt1_c12020

*NCBI accession number is shown instead of locus_tag due to draft genome. [#] Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes. Signal recognition particle protein (SRP54, SRP19), signal recognition particle receptor (*ftsY*). Preprotein translocase (*secGYE*), twin arginine protein translocation system (*tatC*).

Table A.5.36. T-test of spectra bins in sample (paired) and log transformed with *P*-value threshold of 0.05

Data transformation - Log							
Data scaling - Auto scaling				**Data scaling - Pareto scaling**			
Chemical shift	***P*-value**	**-log10(*P*)**	**FDR**	**Chemical shift**	***P*-value**	**-log10(*P*)**	**FDR**
0.618576	0.027484	1.5609	0.26061	0.618576	0.027484	1.5609	0.26061
0.628577	0.017632	1.7537	0.21773	0.628577	0.017632	1.7537	0.21773
0.638578	0.011383	1.9437	0.17372	0.638578	0.011383	1.9437	0.17372
0.878595	0.040705	1.3904	0.3226	0.878595	0.040705	1.3904	0.3226
0.888596	0.020211	1.6944	0.22433	0.888596	0.020211	1.6944	0.22433
0.898597	0.019024	1.7207	0.22433	0.898597	0.019024	1.7207	0.22433
0.958601	0.029485	1.5304	0.27467	0.958601	0.029485	1.5304	0.27467
1.04861	0.02035	1.6914	0.22433	1.04861	0.02035	1.6914	0.22433
1.05861	0.044553	1.3511	0.33797	1.05861	0.044553	1.3511	0.33797
1.17862	0.037261	1.4288	0.30364	1.17862	0.037261	1.4288	0.30364
1.47864	0.031946	1.4956	0.27918	1.47864	0.031946	1.4956	0.27918
1.75866	0.044437	1.3523	0.33797	1.75866	0.044437	1.3523	0.33797
2.10868	0.005681	2.2456	0.13115	2.10868	0.005681	2.2456	0.13115
2.11868	7.21E-05	4.1419	0.019151	2.11868	7.21E-05	4.1419	0.019151
2.12869	0.008258	2.0831	0.15019	2.12869	0.008258	2.0831	0.15019
2.13869	0.030346	1.5179	0.27782	2.13869	0.030346	1.5179	0.27782
2.14869	0.008075	2.0929	0.15019	2.14869	0.008075	2.0929	0.15019
2.15869	0.009218	2.0354	0.1579	2.15869	0.009218	2.0354	0.1579
2.16869	0.012009	1.9205	0.17714	2.16869	0.012009	1.9205	0.17714
2.17869	0.01145	1.9412	0.17372	2.17869	0.01145	1.9412	0.17372
2.3287	0.020701	1.684	0.22433	2.3287	0.020701	1.684	0.22433
2.3387	0.015524	1.809	0.20106	2.3387	0.015524	1.809	0.20106
2.3487	0.001564	2.8058	0.074874	2.3487	0.001564	2.8058	0.074874
2.3587	0.037636	1.4244	0.30364	2.3587	0.037636	1.4244	0.30364
2.3787	0.017502	1.7569	0.21773	2.3787	0.017502	1.7569	0.21773
2.39871	0.005456	2.2631	0.13115	2.39871	0.005456	2.2631	0.13115
2.46871	0.021927	1.659	0.22829	2.46871	0.021927	1.659	0.22829
2.47871	0.012356	1.9081	0.17733	2.47871	0.012356	1.9081	0.17733
2.48871	0.001371	2.8631	0.074874	2.48871	0.001371	2.8631	0.074874
2.49871	0.006868	2.1632	0.15019	2.49871	0.006868	2.1632	0.15019
2.63872	0.007569	2.121	0.15019	2.63872	0.007569	2.121	0.15019
2.64872	0.002787	2.5548	0.09867	2.64872	0.002787	2.5548	0.09867
2.65872	0.003385	2.4704	0.099858	2.65872	0.003385	2.4704	0.099858
2.66872	0.018192	1.7401	0.21955	2.66872	0.018192	1.7401	0.21955
2.67873	0.001583	2.8005	0.074874	2.67873	0.001583	2.8005	0.074874
2.68873	0.001521	2.818	0.074874	2.68873	0.001521	2.818	0.074874
2.76873	0.037741	1.4232	0.30364	2.76873	0.037741	1.4232	0.30364
2.77873	0.027188	1.5656	0.26061	2.77873	0.027188	1.5656	0.26061
2.78873	0.000195	3.7106	0.032031	2.78873	0.000195	3.7106	0.032031
2.79873	2.42E-05	4.6161	0.012854	2.79873	2.42E-05	4.6161	0.012854
2.80873	0.00114	2.9433	0.074874	2.80873	0.00114	2.9433	0.074874
2.81874	0.000241	3.6175	0.032031	2.81874	0.000241	3.6175	0.032031
2.84874	0.003661	2.4364	0.10136	2.84874	0.003661	2.4364	0.10136
2.85874	0.036325	1.4398	0.30364	2.85874	0.036325	1.4398	0.30364
2.89874	0.010108	1.9953	0.16773	2.89874	0.010108	1.9953	0.16773
2.90874	0.015144	1.8198	0.20106	2.90874	0.015144	1.8198	0.20106
2.92874	0.008485	2.0713	0.15019	2.92874	0.008485	2.0713	0.15019
2.94875	0.001841	2.7349	0.074874	2.94875	0.001841	2.7349	0.074874
2.99875	0.025107	1.6002	0.25638	2.99875	0.025107	1.6002	0.25638
3.00875	0.00791	2.1018	0.15019	3.00875	0.00791	2.1018	0.15019
3.01875	0.008201	2.0862	0.15019	3.01875	0.008201	2.0862	0.15019
3.02875	0.021636	1.6648	0.22829	3.02875	0.021636	1.6648	0.22829
3.59879	0.032072	1.4939	0.27918	3.59879	0.032072	1.4939	0.27918
3.88881	0.015477	1.8103	0.20106	3.88881	0.015477	1.8103	0.20106
5.18891	0.026876	1.5706	0.26061	5.18891	0.026876	1.5706	0.26061
5.22891	0.003343	2.4759	0.099858	5.22891	0.003343	2.4759	0.099858
5.24891	0.010487	1.9793	0.16875	5.24891	0.010487	1.9793	0.16875
5.78895	0.032889	1.483	0.28167	5.78895	0.032889	1.483	0.28167
6.06897	0.043671	1.3598	0.33797	6.06897	0.043671	1.3598	0.33797
7.72909	0.001365	2.8648	0.074874	7.72909	0.001365	2.8648	0.074874
7.8591	0.031981	1.4951	0.27918	7.8591	0.031981	1.4951	0.27918
8.17912	0.025848	1.5876	0.25896	8.17912	0.025848	1.5876	0.25896
8.18912	0.003208	2.4938	0.099858	8.18912	0.003208	2.4938	0.099858
8.23913	0.001974	2.7046	0.074874	8.23913	0.001974	2.7046	0.074874
8.24913	0.003818	2.4182	0.10136	8.24913	0.003818	2.4182	0.10136
8.25913	0.001873	2.7275	0.074874	8.25913	0.001873	2.7275	0.074874
8.32913	0.020533	1.6875	0.22433	8.32913	0.020533	1.6875	0.22433
8.59915	0.001268	2.897	0.074874	8.59915	0.001268	2.897	0.074874
8.60915	0.004463	2.3504	0.11285	8.60915	0.004463	2.3504	0.11285
8.70916	0.013543	1.8683	0.18924	8.70916	0.013543	1.8683	0.18924

Table A.5.37. T-test of spectra bins in control (paired) and log transformed with *P*-value threshold of 0.05

Data transformation - Log							
Data scaling - Auto scaling				**Data scaling - Pareto scaling**			
Chemical shift	***P*-value**	**-log10(*P*)**	**FDR**	**Chemical shift**	***P*-value**	**-log10(*P*)**	**FDR**
2.68873	0.044431	1.3523	0.58204	2.68873	0.044431	1.3523	0.49948
2.78873	0.008903	2.0505	0.35067	2.78873	0.008903	2.0505	0.27808
2.83874	0.048355	1.3156	0.59317	2.83874	0.048355	1.3156	0.50522
2.84874	0.022264	1.6524	0.4726	2.84874	0.022264	1.6524	0.40575
2.85874	0.021226	1.6731	0.4726	2.85874	0.021226	1.6731	0.40575
2.89874	0.002473	2.6068	0.14064	2.89874	0.002473	2.6068	0.10533
2.91874	0.019869	1.7018	0.4726	2.91874	0.019869	1.7018	0.40575
2.92874	0.030446	1.5165	0.4726	2.92874	0.030446	1.5165	0.40575
2.94875	0.012311	1.9097	0.43641	2.94875	0.012311	1.9097	0.34405
2.99875	0.027886	1.5546	0.4726	2.99875	0.027886	1.5546	0.40575
3.00875	0.042708	1.3695	0.58204	3.00875	0.042708	1.3695	0.493
3.02875	0.028912	1.5389	0.4726	3.02875	0.028912	1.5389	0.40575
3.03875	0.033688	1.4725	0.48744	3.03875	0.033688	1.4725	0.41601
3.04875	0.016086	1.7936	0.4726	3.04875	0.016086	1.7936	0.40575
3.05875	0.025489	1.5936	0.4726	3.05875	0.025489	1.5936	0.40575
3.08876	0.011244	1.9491	0.41959	3.08876	0.011244	1.9491	0.3317
3.09876	0.021254	1.6726	0.4726	3.09876	0.021254	1.6726	0.40575
3.10876	0.032959	1.482	0.48684	3.10876	0.032959	1.482	0.41601
3.59879	0.002777	2.5564	0.14064	3.59879	0.002777	2.5564	0.10533
3.6788	0.017958	1.7457	0.4726	3.6788	0.017958	1.7457	0.40575
3.7288	0.027635	1.5585	0.4726	3.7288	0.027635	1.5585	0.40575
3.77881	0.029032	1.5371	0.4726	3.77881	0.029032	1.5371	0.40575
3.79881	0.001418	2.8485	0.10051	3.79881	0.001418	2.8485	0.075273
3.80881	0.000131	3.8835	0.030905	3.80881	0.000131	3.8835	0.023146
3.81881	0.035005	1.4559	0.49637	3.81881	0.035005	1.4559	0.41778
3.86881	0.045151	1.3453	0.58204	3.86881	0.045151	1.3453	0.49948
3.88881	0.000933	3.0299	0.092066	3.88881	0.000933	3.0299	0.068952
3.89881	8.32E-05	4.0799	0.030905	3.89881	8.32E-05	4.0799	0.023146
3.90881	0.018015	1.7444	0.4726	3.90881	0.018015	1.7444	0.40575
3.93882	0.022965	1.6389	0.4726	3.93882	0.022965	1.6389	0.40575
3.97882	0.004121	2.385	0.18262	3.97882	0.004121	2.385	0.14589
3.98882	0.001165	2.9336	0.092066	3.98882	0.001165	2.9336	0.068952
4.00882	0.047788	1.3207	0.59317	4.00882	0.047788	1.3207	0.50522
4.07883	0.001717	2.7653	0.11064	4.07883	0.001717	2.7653	0.082866
4.10883	0.023896	1.6217	0.4726	4.10883	0.023896	1.6217	0.40575
4.11883	0.029976	1.5232	0.4726	4.11883	0.029976	1.5232	0.40575
5.26891	0.048524	1.314	0.59317	5.26891	0.048524	1.314	0.50522
5.32892	0.003118	2.5061	0.1474	5.37892	0.000863	3.0642	0.068952
5.34892	0.025055	1.6011	0.4726	5.38892	0.002643	2.5779	0.10533
5.37892	0.000863	3.0642	0.092066	6.06897	0.035405	1.4509	0.41778
5.38892	0.002643	2.5779	0.14064	6.10897	0.029042	1.537	0.40575
5.98897	0.019564	1.7085	0.4726	7.38907	0.024604	1.609	0.40575
6.10897	0.029042	1.537	0.4726	7.73909	0.01876	1.7268	0.40575
6.14898	0.040083	1.397	0.55723	7.98911	0.004858	2.3136	0.16122
6.27899	0.026223	1.5813	0.4726	8.18912	0.000108	3.967	0.023146
6.419	0.013273	1.877	0.44812	8.19913	0.017089	1.7673	0.40575
7.38907	0.024604	1.609	0.4726	8.22913	0.031329	1.5041	0.40575
7.73909	0.01876	1.7268	0.4726	8.23913	0.000502	3.2991	0.066676
7.98911	0.004858	2.3136	0.2026	8.24913	0.030878	1.5103	0.40575
8.18912	0.000108	3.967	0.030905	8.25913	0.000836	3.0775	0.068952
8.19913	0.017089	1.7673	0.4726	8.32913	0.001169	2.9323	0.068952
8.22913	0.031329	1.5041	0.4726				
8.23913	0.000502	3.2991	0.089027				
8.24913	0.030878	1.5103	0.4726				
8.25913	0.000836	3.0775	0.092066				
8.32913	0.001169	2.9323	0.092066				
8.54915	0.04449	1.3517	0.58204				
8.67916	0.019664	1.7063	0.4726				

Table A.5.38. Gene expressions of ISO4-H5

Locus_tag	Product	Trt1	Trt2	Trt3	Trt4
AR505_0005*	adhesin-like protein	462*	152	153	76*
AR505_0024*	hypothetical transmembrane protein	158*	167	113	71*
AR505_0025*	ATP-dependent DNA helicase	39	17*	33	59*
AR505_0028*	endonuclease IV	257	251*	423	597*
AR505_0032*	thiamine biosynthesis protein ThiC1	107*	161	131	241*
AR505_0038*	ribosomal RNA large subunit methyltransferase J RrmJ	144	245*	113*	140
AR505_0039*	ATP-dependent DNA helicase	52*	43	25*	49
ISO4H5_0043*	tRNA 2'-*O*-methylase (pseudo)	170*	48*	65	71
AR505_0059*	radical SAM domain protein	77*	161	192*	177
AR505_0067*	hydrogenase accessory protein HypB	79*	134	163*	136
AR505_0072*	hypothetical protein	126*	195	234	305*
AR505_0080*	hydrogenase expression/formation protein HypE	34*	84*	46	53
AR505_0093*	molecular chaperone GrpE	309*	134	151	100*
AR505_0094*	chaperone protein DnaK	925*	302	420	275*
AR505_0095*	chaperone protein DnaJ	545*	176*	354	291
AR505_0098*	pyridoxamine 5'-phosphate oxidase-related protein	395*	236	144*	153
AR505_0114*	Methylase involved in ubiquinone/menaquinone biosynthesis	103*	37	27*	47
AR505_0126*	hypothetical transmembrane protein	118	172*	56*	92
AR505_0133*	conserved hypothetical protein	132*	101	75	57*
AR505_0167*	conserved hypothetical protein	92*	220*	162	158
AR505_0169*	dihydropteroate synthase FolP	272	344*	228	161*
AR505_0183*	hypothetical protein	43	115*	63	21*
AR505_0189*	DNA repair and recombination protein RadB	163*	95	74	60*
AR505_0199*	transcriptional regulator AsnC family	72	45*	101*	45
AR505_0201*	Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit B GatB	72*	21*	34	25
AR505_0206*	nitrogen regulatory protein P-II GlnK	104	86	131*	56*
AR505_0218*	ribosomal protein L3P Rpl3p	436*	192*	196	255
AR505_0220*	ribosomal protein L23P Rpl23p	229*	62*	124	93
AR505_0221*	ribosomal protein L2P Rpl2p	237*	56*	201	121
AR505_0222*	ribosomal protein S19P Rps19p	253*	44*	303	134
AR505_0223*	ribosomal protein L22P Rpl22p	259*	70*	227	164
AR505_0224*	ribosomal protein S3P Rps3p	156*	41*	138	94
AR505_0225*	ribosomal protein L29P Rpl29p	235	81*	301*	177
AR505_0226*	translation initiation factor aSUI1	298	65*	358*	163
AR505_0227*	ribonuclease P subunit P29	297*	61*	293	153
AR505_0228*	ribosomal protein S17P Rps17p	118*	20*	78	46
AR505_0229*	ribosomal protein L14e Rpl14e	214*	30*	173	103
AR505_0230*	ribosomal protein L24P Rpl24p	119*	19*	85	51
AR505_0231*	ribosomal protein S4e Rps4e	192*	58*	163	106
AR505_0232*	ribosomal protein L5P Rpl5p	133*	24*	85	66
AR505_0233*	ribosomal protein S8P Rps8p	161*	39*	112	43
AR505_0234*	ribosomal protein L6P Rpl6p	168	51*	178*	94
AR505_0235*	ribosomal protein L32e Rpl32e	251*	61*	186	96
AR505_0236*	ribosomal protein L19e Rpl19e	411*	117*	345	237
AR505_0237*	ribosomal protein L18P Rpl18p	284*	65*	232	157
AR505_0238*	ribosomal protein S5P Rps5p	340*	77*	331	183
AR505_0239*	ribosomal protein L30P Rpl30p	336*	68*	277	177
AR505_0240*	ribosomal protein L15P Rpl15p	389*	84*	318	128
AR505_0241*	preprotein translocase subunit SecY	180*	26*	114	56
AR505_0242*	hypothetical transmembrane protein	151*	18*	90	44
AR505_0260*	phosphate uptake regulator PhoU	123*	54	42*	70
AR505_0267*	hypothetical protein	31	48*	22	11*
AR505_0270*	peptidase M16 family	54*	96	63	130*
AR505_0271*	hydrogenase maturation protease HycI	154*	283	195	401*
AR505_0273*	CoB--CoM heterodisulfide reductase subunit C HdrC	93	79*	681*	166
AR505_0274*	CoB--CoM heterodisulfide reductase subunit B HdrB	182	110*	836*	279
AR505_0275*	TPR repeat-containing protein	13	8*	38*	16
AR505_0281*	dihydroorotase PyrC	134*	82	80	59*
AR505_0287*	sodium/proline symporter PutP	141*	56	39	32*
AR505_0288*	hypothetical transmembrane protein	229	296	301*	141*
AR505_0289*	NADPH-dependent FMN reductase	339*	233	125	78*
AR505_0299*	cobalamin biosynthesis protein CbiX	316*	278	127*	207
AR505_0359*	nitrogenase iron protein NifH	84*	34*	57	68
AR505_0376*	N-carbamoyl-D-amino acid amidohydrolase AguB	38	64*	46	26*
AR505_0377*	NAD synthetase NadE	36*	107	129*	95
AR505_0384*	hypothetical protein	4047	1276*	3195	4321*
AR505_0385*	transporter Na^+/H^+ antiporter family	280*	134	78*	219
AR505_0388*	DNA topoisomerase VI subunit A	43	21	21*	52*
AR505_0394*	histidyl-tRNA synthetase HisS	44*	64	81	103*
AR505_0396*	hypothetical protein	46	64*	59	16*
AR505_0418*	ABC transporter permease protein	70*	10	8*	11
AR505_0443*	universal stress protein UspA	57*	62	182	475*
AR505_0456*	transporter DMT family	402*	1463*	513	905

AR505_0460*	flap endonuclease Fen	72*	111	121	161*
AR505_0461*	ribulose-phosphate 3-epimerase Rpe	94*	170	224*	186
AR505_0464*	ribonuclease Z Rnz	109*	54	43*	57
AR505_0465*	HTH/CBS domain-containing protein	52*	20	12	4*
AR505_0470*	phosphopyruvate hydratase Eno	48	55*	28	21*
AR505_0474*	2,3-bisphosphoglycerate-independent phosphoglycerate mutase ApgM	28*	42	72*	38
AR505_0485*	PHP domain-containing protein	147*	67	54*	89
AR505_0498*	transposase IS605 OrfB family	53*	41	32	22*
AR505_0508*	phospho-2-dehydro-3-deoxyheptonate aldolase/fructose-bisphosphate aldolase Fba	115*	338*	288	184
AR505_0509*	3-dehydroquinate synthase AroB1	112*	250*	185	128
AR505_0518*	hypothetical protein	37	20*	45*	21
AR505_0532*	HTH domain-containing protein	167*	344*	216	239
AR505_0536*	TPR repeat-containing protein	122	195*	88*	125
AR505_0549*	glucose-1-phosphate cytidylyltransferase RfbF	141	322*	241	115*
AR505_0558*	LPS biosynthesis protein LICD family	120*	75	71	45*
AR505_0577*	hypothetical protein	85*	240*	121	137
AR505_0594*	adhesin-like protein	326*	57	41	26*
AR505_0601*	hydroxymethylglutaryl-CoA synthase	672*	197	271	188*
AR505_0602*	acetyl-CoA acetyltransferase	1081*	337	480	277*
AR505_0603*	DNA-binding protein	529*	141	257	91*
AR505_0604	phosphoglycolate phosphatase	73	10	12	17
AR505_0605*	ABC transporter permease protein	142*	26*	38	66
AR505_0609*	low molecular weight phosphotyrosine protein phosphatase	35*	53	72	157*
AR505_0619*	hypothetical transmembrane protein	11*	660	349	727*
AR505_0630*	desulfoferrodoxin Dfx	171*	228	414*	226
AR505_0631*	2-isopropylmalate synthase LeuA	69*	175*	105	149
AR505_0636*	pap2 family protein	83	37*	68	88*
AR505_0640*	hypothetical protein	1639*	3634	2549	4169*
AR505_0651*	hypothetical protein	306	466*	291	216*
AR505_0660*	adhesin-like protein	126*	108	101	51*
AR505_0664*	adhesin-like protein	111*	49*	83	92
AR505_0667*	signal peptidase I	243	434*	182	156*
AR505_0670*	adhesin-like protein	310*	206	222	139*
AR505_0675*	acetylglutamate kinase ArgB	29	15*	48*	34
AR505_0678*	citrate synthase	55*	56	111*	62
AR505_0689*	hypothetical transmembrane protein	150	235*	104*	148
AR505_0696*	transporter CDF family	97*	252	266*	173
AR505_0723*	hypothetical transmembrane protein	207*	465*	322	342
AR505_0740*	TATA-box-binding protein Tbp	68*	258*	140	207
AR505_0749*	trimethylamine permease	47	27*	119	1143*
AR505_0750*	hypothetical protein	33	27*	72	955*
AR505_0754*	nitrate/sulfonate/bicarbonate ABC transporter substrate-binding protein	295*	705	407	771*
AR505_0777*	DNA-binding protein	213*	89*	102	148
AR505_0793*	TPR repeat-containing protein	63	95*	58	39*
AR505_0799*	*O*-acetylserine sulfhydrylase	20	15	13*	43*
AR505_0803*	hypothetical protein	85*	47	54	27*
AR505_0809*	tRNA-dihydrouridine synthase DusA	519*	165	126*	280
AR505_0815*	Na⁾dependent transporter SBF family	38*	97*	49	63
AR505_0823*	hypothetical transmembrane protein	58*	38	32	22*
AR505_0826*	hypothetical transmembrane protein	79*	70	29*	58
AR505_0827*	wyosine biosynthesis protein TYW1	58*	22	32	16*
AR505_0830*	amino acid/peptide transporter	58*	23*	31	39
AR505_0833*	hypothetical protein	83*	142	187*	130
AR505_0850*	hypothetical protein	10	3*	18	140*
AR505_0856*	hypothetical transmembrane protein	137	549*	396	52*
AR505_0859*	transcriptional regulator TetR family	6152*	7818	14364	31275*
AR505_0897*	hypothetical protein	166*	72	60	33*
AR505_0902*	Fe-S oxidoreductase	43	87*	71	41*
AR505_0903*	transposase IS605 OrfB family	100*	139	223*	133
AR505_0905*	EamA-like transporter family	52	25*	33	68*
AR505_0909*	cell surface protein	128	263*	114*	164
AR505_0921*	hypothetical protein	78	143*	76	56*
AR505_0924*	pyridoxamine 5'-phosphate oxidase family protein	103*	140	211	267*
AR505_0931*	phage integrase	19*	164*	81	111
AR505_0932*	fructose-bisphosphate aldolase Fba	3016	2045*	4228	5566*
AR505_0935*	PHP domain-containing protein	30*	56	59	87*
AR505_0939*	FeS assembly protein SufBD	40*	88*	37	57
AR505_0942*	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase GpmA	50	89	40*	104*
AR505_0946*	hypothetical protein	129	193*	150	70*
AR505_0947*	cobalamin biosynthesis protein CobW	48*	76	115*	65
AR505_0948*	hypothetical protein	841*	1131	1639*	1214

AR505_0954*	hypothetical protein	147*	30*	46	45
AR505_0960*	Arylsulfotransferase AssT	66*	104	109	152*
AR505_0974*	hypothetical protein	968	509*	790	1631*
AR505_0975*	cell division protein FtsZ	664*	308*	423	531
AR505_0985*	adhesin-like protein	198*	327	546*	422
AR505_0988*	hypothetical transmembrane protein	289	211*	390	599*
AR505_1008*	SAM-dependent methyltransferases	16*	42	34	89*
AR505_1016*	hypothetical protein	134*	41	21*	30
AR505_1022*	hypothetical protein	38	22*	34	245*
AR505_1027*	cysteine desulfurase SufS subfamily SufS	33*	80*	44	44
AR505_1032*	adhesin-like protein	63*	157*	127	84
AR505_1039*	NADPH-dependent FMN reductase	122*	194	207	284*
AR505_1053*	hypothetical protein	115	143*	115	49*
AR505_1055*	site-specific recombinase	97*	13	10*	29
AR505_1062*	thermosome subunit	578*	844	1243*	1007
AR505_1066*	methylcobalamin:CoM methyltransferase MtaA	52*	63	105	881*
AR505_1068*	4Fe-4S Fdx iron-sulfur binding domain-containing protein	149	116*	222	1139*
AR505_1073*	histidinol dehydrogenase HisD	155*	412*	307	384
AR505_1083*	hypothetical protein	155*	77	104	53*
AR505_1092*	CRISPR-associated protein Cas7/Cse4/CasC	67*	12	21	10*
AR505_1097*	hypothetical protein	79	101*	71	38*
AR505_1107*	hypothetical protein	28*	41	84*	60
AR505_1108*	conserved hypothetical protein	198*	524	740*	498
AR505_1115*	DEAD/DEAH box helicase domain-containing protein	57*	26	22*	34
AR505_1118*	MTA/SAH nucleosidase MtnN	35	61*	26*	45
AR505_1127*	hypothetical protein	42*	119*	73	96
AR505_1138*	transporter SDF family	21	34	18*	52*
AR505_1146*	ribosomal-protein-alanine acetyltransferase RimI	28	59*	29	22*
AR505_1149*	fructose 1,6-bisphosphatase Fbp	194*	156	159	93*
AR505_1155*	quinolinate synthetase A protein NadA	113	196*	98*	186
AR505_1157*	hypothetical protein	4*	48*	27	23
AR505_1170*	signal transduction histidine kinase	20*	61	31	117*
AR505_1171*	hypothetical protein	33*	85	55	86*
AR505_1177*	MMPL domain-containing protein	69*	15	10*	19
AR505_1178*	radical SAM domain containing protein	364	459	250*	583*
AR505_1185*	hydrogenase nickel insertion protein HypA1	14*	15	34	1017*
AR505_1187*	conserved hypothetical	221	1091*	159*	474
AR505_1188*	hypothetical transmembrane protein	67	125*	45*	102
AR505_1216*	TPR repeat-containing protein	63	98*	42*	51
AR505_1218*	haloacid dehalogenase-like hydrolase	696	901*	490	311*
AR505_1219*	transposase IS605 OrfB family	99	124*	112	42*
AR505_1220*	hypothetical transmembrane protein	42	103*	69	26*
AR505_1222*	ferrous iron transport protein B FeoB	89*	43	16*	32
AR505_1224*	hypothetical protein	100*	81	32*	50
AR505_1227*	heavy metal translocating P-type ATPase	46	26*	46	63*
AR505_1233*	pseudouridylate synthase	48	87*	38*	63
AR505_1242*	dimethylamine permease	29	17*	58	223*
AR505_1257*	NYN domain-containing protein	87*	127	229*	174
AR505_1263*	molybdenum cofactor biosynthesis protein MoaA1	189*	24*	64	35
AR505_1266*	hypothetical protein	283*	53	35	27*
AR505_1274*	tRNA pseudouridine synthase D TruD	73*	47	32*	56
AR505_1281*	hypothetical protein	151*	51*	104	134
AR505_1297*	adenosylcobalamin synthase CobS	115*	45	39	26*
AR505_1300*	hypothetical protein	138*	30*	30	76
AR505_1301*	hypothetical protein	320*	119	130*	215
AR505_1310*	small multidrug resistance protein	30	11*	34	163*
AR505_1311*	small multidrug resistance protein	29	5*	30	99*
AR505_1313*	hypothetical transmembrane protein	218*	488*	310	313
AR505_1315*	hypothetical protein	311	575*	370	273*
AR505_1322*	pyrrolysine biosynthesis protein PylD	179*	99	105	55*
AR505_1323*	pyrrolysine biosynthesis protein PylC	198*	107	127	90*
AR505_1325*	pyrrolysine--tRNA ligase PylS	289*	769*	428	514
AR505_1328*	monomethylamine methyltransferase MtmB	132	72*	145	559*
AR505_1329*	methyltransferase cognate corrinoid proteins	75	49*	76	291*
AR505_1332*	dimethylamine:corrinoid methyltransferase MtbB	1183	1018*	2138	3030*
AR505_1333*	dimethylamine corrinoid protein	1800*	2213	3479	5257*
AR505_1335*	conserved hypothetical protein	79	101	185*	73*
AR505_1351*	TPR repeat-containing protein	51	92*	40*	59
AR505_1369*	ATP-dependent DNA helicase	73*	19*	31	32
AR505_1376*	isopropylmalate/isohomocitrate dehydrogenase	80*	32*	56	60
AR505_1382*	*N*-methylhydantoinase A/acetone carboxylase beta subunit	33	20*	34	60*
AR505_1383*	glycolate oxidase, subunit GlcD3	281*	491	489	691*
AR505_1385*	methanogenesis marker protein 7	277*	77*	249	162
AR505_1386*	methanogenesis marker protein 17	381	118*	411*	209
AR505_1387*	methanogenesis marker protein 15	256	95*	298*	136

AR505_1388*	methanogenesis marker protein 5	101	48*	131*	53
AR505_1395*	hypothetical protein	1683	999*	4926*	1831
AR505_1396*	methyl-CoM reductase alpha subunit McrA	1696	953*	2883*	1328
AR505_1400*	hypothetical protein	158*	260	399	400*
AR505_1401*	ornithine carbamoyltransferase ArgF	136*	251	236	274*
AR505_1402*	ssDNA exonuclease RecJ3	141*	220	280	327*
AR505_1426*	seryl-tRNA synthetase SerS	144	78*	108	247*
AR505_1435*	molybdenum cofactor biosynthesis protein A MoaA	197*	97	114	93*
AR505_1437*	GTPase	131*	221	303*	178
AR505_1442*	ribosomal protein S13P Rps13p	134*	81	86	52*
AR505_1459*	translation initiation factor aIF-2 alpha subunit	211*	444*	362	294
AR505_1470*	hypothetical transmembrane protein	53*	14	7*	28
AR505_1472*	ribosomal protein S8e Rps8e	260	381*	216	161*
AR505_1476*	methyl-viologen-reducing hydrogenase alpha subunit MvhA	436	320*	1771*	861
AR505_1477*	methyl-viologen-reducing hydrogenase gamma subunit MvhG	515	364*	1870*	1014
AR505_1478*	methyl-viologen-reducing hydrogenase delta subunit	400	361*	2055*	1072
AR505_1479*	CoB--CoM heterodisulfide reductase subunit A HdrA	381	327*	1194*	748
AR505_1481*	hypothetical protein	214	171*	406	580*
AR505_1489*	hypothetical protein	175*	74	114	53*
AR505_1490*	prefoldin beta subunit PfdB	97	36*	72	186*
AR505_1493*	phosphoesterase DHHA1	78*	98	118	165*
AR505_1543*	phage integrase	54	20*	37	57*
AR505_1551*	ABC transporter ATP-binding protein	69*	39	31*	40
AR505_1559*	adhesin-like protein	35	28*	48	418*
AR505_1566*	death-on-curing family protein	2163*	7422*	4528	3251
AR505_1578*	excinuclease ABC A subunit UvrA	64*	31*	45	33
AR505_1581*	cytosine deaminase	37*	45	39	75*
AR505_1583*	tRNA nucleotidyltransferase Cca	68	138*	54*	108
AR505_1587*	prenyltransferase UbiA	107	142*	73	59*
AR505_1598*	hypothetical protein	54*	110*	88	87
AR505_1601*	RNA 3'-phosphate cyclase RtcA	141*	198	291*	231
AR505_1603*	translation initiation factor aIF-2 beta subunit	61*	43	37	27*
AR505_1622*	$F_{420}H_2$ dehydrogenase subunit N FpoN	538	210	705*	86*
AR505_1623*	$F_{420}H_2$ dehydrogenase subunit M FpoM	359	150	411*	50*
AR505_1624*	$F_{420}H_2$ dehydrogenase subunit L FpoL	541	232	589*	86*
AR505_1625*	$F_{420}H_2$ dehydrogenase subunit K FpoK	404*	99	355	63*
AR505_1626*	hypothetical transmembrane protein	453*	181	439	66*
AR505_1628*	NADH:quinone oxidoreductase I FpoI	633*	214	546	82*
AR505_1629*	NADH dehydrogenase subunit H FpoH	376*	149	289	60*
AR505_1630*	NADH dehydrogenase subunit D FpoD	721*	317	468	134*
AR505_1631*	NADH dehydrogenase subunit C FpoC	765*	437	519	135*
AR505_1632*	NADH dehydrogenase subunit B FpoB	373*	230	261	69*
AR505_1635*	Fe-S oxidoreductase	107*	53	74	48*
AR505_1640*	phage integrase	194*	94*	120	110
AR505_1648*	transketolase subunit B	120	65*	170*	87
AR505_1653*	RNA-metabolising metallo-beta-lactamase	82	57*	99	125*
AR505_1658*	exosome complex RNA-binding protein Csl4	161*	94	124	69*
AR505_1661*	hypothetical protein	87	170*	56*	88
AR505_1667*	5-formaminoimidazole-4-carboxamide-1-(beta)-D- ribofuranosyl 5'-monophosphate-formate ligase PurP	116*	460*	365	302
AR505_1674*	proteasome-activating nucleotidase	39	18*	31	52*
AR505_1675*	radical SAM domain-containing proteinB	85	114*	50*	89
AR505_1692*	transposase IS605 OrfB family	50*	97	118*	80
AR505_1694*	hypothetical protein	60*	102	122*	81
AR505_1696*	CBS domain-containing protein	59	86*	50	29*
AR505_1718*	hypothetical protein	38	63	82*	17*
AR505_1719*	hypothetical protein	41*	71	116*	62
AR505_1723*	hypothetical protein	295*	429	517	594*
AR505_1741*	adhesin-like protein	171*	294	454*	230
AR505_1755*	translation initiation factor aIF-2	178*	327	374*	290
AR505_1758*	ribosomal protein S28e Rps28e	205*	92	87	49*
AR505_1759*	ribosomal protein L7Ae Rpl7ae	466*	268	221	50*
AR505_1760*	hypothetical transmembrane protein	56*	26	36	24*
AR505_1766*	hypothetical protein	194*	182	151	60*
AR505_1775*	transposase IS605 OrfB family	53	52	80*	32*
AR505_1780*	malate dehydrogenase Mdh	37	28*	35	60*
AR505_1782*	DNA primase large subunit PriB	128	187*	93*	102
AR505_1795*	acidic ribosomal protein P0 RplPO	970*	290*	712	595
AR505_1796*	ribosomal protein L1P Rpl1p	685*	252*	507	393
AR505_1810*	AAA family ATPase CDC48 subfamily	647*	243	235*	427
AR505_1811*	transcriptional regulator	115*	169	290*	198
AR505_1817*	hypothetical protein	163*	26*	86	54
AR505_1818*	A_1A_0 ATP synthase subunit D	402*	57	188	64*
AR505_1819*	A_1A_0ATP synthase subunit B	353*	40*	131	44
AR505_1820*	A_1A_0ATP synthase subunit A	449*	48*	118	55

AR505_1821*	A_1A_0ATP synthase subunit F	211*	29*	44	33
AR505_1822*	A_1A_0ATP synthase subunit C	522*	56	132	37*
AR505_1823*	A_1A_0ATP synthase subunit E	395*	71	150	55*
AR505_1824*	A_1A_0ATP synthase subunit K	600*	92	172	66*
AR505_1825*	A_1A_0ATP synthase subunit I	722*	189	193	107*
AR505_1826*	A_1A_0ATP synthase subunit H	660*	261	220	151*
AR505_1827*	CMP/dCMP deaminase	64*	216*	149	92

Trt1: ISO4-H5 enrichment culture with high H_2 on methanol. Trt2: ISO4-H5 enrichment culture +FD1 with high H_2 on methanol. Trt3: ISO4-H5 enrichment culture +FD1 with low H_2 on methanol. Trt4: ISO4-H5 enrichment culture +FD1 with low H_2 on monomethylamine. *: Significantly differentially expressed genes between two conditions by both Krusgal-Wallis test and Benjamini-Hochberg test of $q<0.05$ and above two fold difference in expression were marked with *, the pair of conditions are highlighted with black background and white letter.

Table A.5.4. Gene expressions of *Succinivibrio dextrinosolvens* H5

Locus_tag	Product	Trt1	Trt2	Trt3	Trt4
T508DRAFT_00038*	integral membrane protein, TerC family	93*	33	7*	63
T508DRAFT_00103*	arabinose efflux permease	73	47	9*	106*
T508DRAFT_00131*	aspartate ammonia-lyase	744*	102	99	28*
T508DRAFT_00132*	anaerobic C_4-dicarboxylate membrane transporter family protein	1179*	205	171	10*
T508DRAFT_00165*	transcriptional regulator	430*	171	20	9*
T508DRAFT_00172*	30*S* ribosomal protein S11	319*	31	34	0*
T508DRAFT_00173*	30*S* ribosomal protein S13	622*	204	86	0*
T508DRAFT_00179*	ribosomal protein L18, bacterial type	663*	228	172	0*
T508DRAFT_00230*	outer membrane protein and related peptidoglycan-associated (lipo)proteins	1328*	232	29*	51
T508DRAFT_00241*	threonine ammonia-lyase, biosynthetic, long form	199*	22*	80	134
T508DRAFT_00281*	co-chaperonin GroES (HSP10)	827*	114*	249	553
T508DRAFT_00303*	ATP-binding cassette protein, ChvD family	189*	52	4*	40
T508DRAFT_00354*	phosphoglucomutase, alpha-D-glucose phosphate-specific	212*	36	4*	16
T508DRAFT_00369*	ADP-heptose:LPS heptosyltransferase	259*	71	20*	86
T508DRAFT_00413*	RNA polymerase, sigma 32 subunit, RpoH	35	192*	11*	61
T508DRAFT_00450*	predicted Zn-dependent proteases and their inactivated homologs	53*	331*	106	108
T508DRAFT_00525*	glycogen/starch/alpha-glucan phosphorylases	348*	44	27	3*
T508DRAFT_00592*	sugar phosphate permease	195*	69	17*	45
T508DRAFT_00643*	transposase and inactivated derivatives	412*	191	49*	211
T508DRAFT_00647*	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain	246	566*	435	83*
T508DRAFT_00694*	ribosomal large subunit pseudouridine synthase A (EC 5.4.99.-)	318*	131	40	10*
T508DRAFT_00730*	acetyl-CoA carboxylase, biotin carboxylase subunit	103*	30	12	6*
T508DRAFT_00782*	hypothetical protein	994*	434	159	69*
T508DRAFT_00826*	uncharacterised protein conserved in bacteria	179*	34	35	0*
T508DRAFT_00857*	hypothetical protein	6644*	4720	1142*	1823
T508DRAFT_00876*	DNA polymerase III, epsilon subunit and related 3'-5' exonucleases	280*	53	12	0*
T508DRAFT_00891*	conjugative transfer signal peptidase TraF	291*	113	20*	36
T508DRAFT_00914*	malate dehydrogenase (NAD) (EC 1.1.1.37)	699*	174	16*	71
T508DRAFT_00918*	cell division protein FtsZ	308*	219	69	22*
T508DRAFT_00919*	ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components	45	28	107*	7*
T508DRAFT_00953*	ribosomal protein S9	584*	241	106	0*
T508DRAFT_00970*	*N*-acetylglucosamine-6-phosphate deacetylase	628*	3366*	789	2399
T508DRAFT_00983*	tRNA(Ile)-lysidine synthetase, *N*-terminal domain/tRNA(Ile)-lysidine synthetase, C-terminal domain	194*	556	496	1062*
T508DRAFT_01005*	NADH:ubiquinone oxidoreductase, Na^+-translocating, F subunit	299*	92	31	0*
T508DRAFT_01024*	hypothetical protein	78*	425*	531	156
T508DRAFT_01118*	hypothetical protein	722*	3180*	2249	3409
T508DRAFT_01186*	hypothetical protein	58	217*	9*	30
T508DRAFT_01252*	hypothetical protein	120	390*	228	53*
T508DRAFT_01281*	23*S* rRNA (uracil-5-)-methyltransferase RumA	43	74*	10*	22
T508DRAFT_01385*	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	530*	273	22	0*
T508DRAFT_01387*	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	726*	49	29	16*
T508DRAFT_01388*	pyruvate:Fdx oxidoreductase and related 2-oxoacid:Fdx oxidoreductases, beta subunit	957	60	61	7*
T508DRAFT_01404*	uncharacterised oxidoreductases, Fe-dependent alcohol dehydrogenase family	127*	61	11*	37
T508DRAFT_01410*	hypothetical protein	8709*	6323	1109*	1910
T508DRAFT_01412*	alpha-1,4-glucan:alpha-1,4-glucan 6-glycosyltransferase	122*	17	5*	24
T508DRAFT_01465*	predicted transcriptional regulator containing an HTH domain and an uncharacterized domain shared with the mammalian protein Schlafen	157*	55	4*	12
T508DRAFT_01506*	succinate dehydrogenase and fumarate reductase iron-sulfur protein	252*	31	29	0*
T508DRAFT_01507*	fumarate reductase subunit C	329*	97	59	0*
T508DRAFT_01509*	pyruvate kinase	318*	78	9*	187
T508DRAFT_01527*	glutamate dehydrogenase (NADP) (EC 1.4.1.4)	122*	82	6*	79
T508DRAFT_01537*	translation initiation factor IF-2	83*	22	3*	8
T508DRAFT_01623*	phosphoglycerate mutase (EC 5.4.2.1)	495*	184	27*	94
T508DRAFT_01775*	ribosomal S4P (gammaproteobacterial)	317*	950*	550	694
T508DRAFT_01778*	predicted AAA-ATPase	310*	169	39*	147
T508DRAFT_01805*	ribosomal protein S1	259*	28	8*	14
T508DRAFT_01832*	type I secretion outer membrane protein, TolC family	108	150*	76	0*
T508DRAFT_01917*	uncharacterised conserved protein	242*	85	4*	57
T508DRAFT_01956*	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	103	12	288*	0*
T508DRAFT_02013*	ATPase, P-type (transporting), HAD superfamily, subfamily IC/heavy metal translocating P-type ATPase	78*	15	3*	16
T508DRAFT_02014*	hypothetical protein	128*	844*	440	201
T508DRAFT_02093*	hypothetical protein	5652*	1840	599*	778
T508DRAFT_02094*	hypothetical protein	3544*	1260	190*	658

T508DRAFT_02129*	translation elongation factor TU	2034*	313	56	47*
T508DRAFT_02143*	tyrosine lyase ThiH	133*	52	21*	63
T508DRAFT_02152*	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	96	86	14*	167*
T508DRAFT_02173*	flagellin and related hook-associated proteins	151*	77	34	14*
T508DRAFT_02218*	hypothetical protein	260*	563	548	897*
T508DRAFT_02255*	transposase DDE domain	1881*	556	149*	357
T508DRAFT_02257*	transposase DDE domain	1695*	948	97*	604
T508DRAFT_02258*	hypothetical protein	1571*	697	151*	470
T508DRAFT_02264*	Permeases of the drug/metabolite transporter (DMT) superfamily	530*	181	113*	322
T508DRAFT_02280*	transcriptional regulator, IclR family	532*	379	31*	62
T508DRAFT_02283*	hypothetical protein	71*	21	15	0*
T508DRAFT_02301*	hypothetical protein	8490*	5171	891*	3785
T508DRAFT_02302*	translation elongation factor TU	2091*	379	57	47*
T508DRAFT_02320*	hypothetical protein	8440*	5161	891*	3761
T508DRAFT_02322*	outer membrane protein (porin)	688*	289	139	29*
T508DRAFT_02323*	outer membrane protein (porin)	3829*	744	591	55*
T508DRAFT_02325*	hypothetical protein	8763*	4823	1211*	2346
T508DRAFT_02331*	hypothetical protein	6670*	4755	1124*	1827
T508DRAFT_02366*	FOG: transposase and inactivated derivatives	1250*	355	24*	405
T508DRAFT_02404*	hypothetical protein	45*	10	14	0*
T508DRAFT_02410*	hypothetical protein	460*	313	18*	130
T508DRAFT_02418*	flagellin and related hook-associated proteins	138*	60	11*	116
T508DRAFT_02428*	flagellin and related hook-associated proteins	288	105*	153	463*

Trt1: ISO4-H5 enrichment culture with high H_2 on methanol. Trt2: ISO4-H5 enrichment culture +FD1 with high H_2 on methanol. Trt3: ISO4-H5 enrichment culture +FD1 with low H_2 on methanol. Trt4: ISO4-H5 enrichment culture +FD1 with low H_2 on monomethylamine. *: Significantly differentially expressed genes between two conditions by both Krusgal-Wallis test and Benjamini-Hochberg test of $q<0.05$ and above two fold difference in expression were marked with *, the pair of conditions are highlighted with black background and white letter.

Table A.5.5. Gene expressions of *Ruminococcus flavefaciens* FD1

Locus_tag	Predicted gene product	Trt2	Trt3	Trt4
FD1_0006*	pseudouridine synthase	2949*	890	47*
FD1_0040*	transglutaminase-like protein	172*	19	0*
FD1_0041*	hypothetical protein	761*	129	16*
FD1_0042*	50*S* ribosomal protein L11 methyltransferase	353*	108	0*
FD1_0045*	hypothetical protein	105*	39	0*
FD1_0048**	ribonuclease III	31**	0**	0
FD1_0052**	GNAT family acetyltransferase	35**	0**	0
FD1_0058*	NADH-quinone oxidoreductase, F subunit	124*	31	12*
FD1_0059*	NADH-quinone oxidoreductase, G subunit	113*	14	0*
FD1_0072*	hypothetical protein	196*	36	0*
FD1_0077*	hypothetical protein	2461*	91*	282
FD1_0108*	glucose-1-phosphate adenylyltransferase	116*	54	0*
FD1_0109*	glucose-1-phosphate adenylyltransferase	181*	86	0*
FD1_0110*	glycogen/starch synthase	213*	68	0*
FD1_0141*	hypothetical protein	298*	187	0*
FD1_0142*	hypothetical protein	325*	143	0*
FD1_0192*	glycoside hydrolase family protein	351	505*	58*
FD1_0223*	dockerin (pseudogene)	1064*	81	12*
FD1_0237**	hypothetical protein	42**	0**	0
FD1_0243**	hypothetical protein	22**	0**	0
FD1_0244*	type III pantothenate kinase	702*	228	0*
FD1_0273**	RNA polymerase sigma-70 factor	0**	48**	0
FD1_0292**	cobalt-precorrin-6A synthase (pseudogene)	22**	0**	0
FD1_0322*	UDP-*N*-acetylmuramoyl-L-alanine--D-glutamate ligase	51*	18	0*
FD1_0333*	RNA polymerase subunit sigma-24	481*	175	0*
FD1_0338*	triosephosphate isomerase	88*	15	0*
FD1_0341*	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	64*	31	0*
FD1_0366*	hypothetical protein	312*	111	17*
FD1_0408**	pyruvate:Fdx oxidoreductase and related 2-oxoacid:Fdx oxidoreductase beta subunit	82**	7**	31
FD1_0409**	pyruvate:Fdx oxidoreductase and related 2-oxoacid:Fdx oxidoreductase gamma subunit	162**	15**	0
FD1_0429*' **	hypothetical protein	447**	2083*,**	0*
FD1_0431**	carbohydrate binding module (family 6)/glycoside hydrolase 97	240**	21**	0
FD1_0436**	multidrug transporter	328**	5**	17
FD1_0451*	hypothetical protein	300*	92	0*
FD1_0453*	glycoside hydrolase family 9	8	26*	0*
FD1_0454*'**	hypothetical protein	536*,**	120**	0*
FD1_0455*'**	Fe-S oxidoreductase	241*,**	51**	0*
FD1_0465**	diguanylate cyclase	98*	5*	0
FD1_0467*'**	transposase, partial	642*,**	91**	0*
FD1_0472**	hypothetical protein	38**	4**	0
FD1_0480*	hypothetical protein	1305	2210*	0*
FD1_0481*	hypothetical protein	1346	2295*	21*
FD1_0498*	von Willebrand factor type A	189*	125	0*
FD1_0499*	hypothetical protein	280*	122	0*
FD1_0514*	sortase	187*	81	0*
FD1_0515*	hypothetical protein	449*	375	15*
FD1_0519**	hypothetical protein	112**	11**	0
FD1_0520*'**	*N*-acetylglucosamine-1-phosphate uridyltransferase	33**	93*,**	0*
FD1_0524**	hypothetical protein	71**	10**	0
FD1_0525**	ATP-dependent chaperone protein ClpB	45**	8**	0
FD1_0537**	phosphoglucosamine mutase	41**	4**	0
FD1_0555*	heat shock protein Hsp20	119	209*	0*
FD1_0562*	hypothetical protein	400*	64	0*
FD1_0563*	esterase/lipase-like protein (pseudogene)	183*	33	0*
FD1_0567*	transcriptional regulator, LysR family	225*	15	0*
FD1_0568*'**	transposase, partial	936*,**	143**	0*
FD1_0572**	D-alanyl-D-alanine carboxypeptidase	152**	3**	0
FD1_0576**	GCN5 family acetyltransferase	228**	15**	0
FD1_0577*'**	tRNA-splicing ligase RtcB	292*,**	18**	0*
FD1_0580**	hypothetical protein	238**	25**	0
FD1_0584*	hypothetical protein	83*	48	0*
FD1_0586**	flavodoxin	284**	22**	0
FD1_0587**	diguanylate cyclase	80**	0**	0
FD1_0588**	exodeoxyribonuclease III	135**	5**	0
FD1_0589**	hypothetical protein	180**	30**	0
FD1_0590*	hypothetical protein	199*	119	0*
FD1_0591*	hypothetical transmembrane protein	144*	81	0*
FD1_0592*	hypothetical transmembrane protein	455*	279	0*
FD1_0593*	XRE family transcriptional regulator	223*	129	0*
FD1_0598*'**	hypothetical protein	2785*,**	319**	11*

FD1_0599*'**	sugar ABC transporter permease	270*,**	46**	0*
FD1_0600*'**	sugar ABC transporter permease	171*,**	25**	0*
FD1_0612**	exodeoxyribonuclease VII large subunit	112**	10**	0
FD1_0615*'**	1-deoxy-D-xylulose-5-phosphate synthase	139*,**	9**	0*
FD1_0616**	TlyA family rRNA methyltransferase/putative hemolysin	100**	8**	0
FD1_0617**	NAD^+ kinase	102**	7**	0
FD1_0619*'**	DNA repair protein RecN	107*,**	13**	0*
FD1_0621*'**	beta-galactosidase	239*,**	24**	0*
FD1_0622*	D-tyrosyl-$tRNA^{Tyr}$ deacylase	182*	76	0*
FD1_0623**	penicillin-binding protein 2	2500**	221**	0
FD1_0626*	serine/threonine protein kinase	35*	17	0*
FD1_0627*	hypothetical protein	59*	27	0*
FD1_0628**	DNA mismatch repair protein MutS	44**	3**	0
FD1_0635**	1-phosphofructokinase	131**	7**	0
FD1_0636*	PTS fructose transporter subunit IIC	41*	11	0*
FD1_0639*	adhesin-like protein	22	44*	0*
FD1_0641**	sugar ABC transporter substrate-binding protein	181**	2**	0
FD1_0650**	sugar ABC transporter permease	66**	7**	9
FD1_0651*'**	carbohydrate ABC transporter substrate-binding protein	163*,**	35**	9*
FD1_0655**	hypothetical protein	115**	1**	0
FD1_0664*'**	hypothetical protein	483*,**	67**	0*
FD1_0679*	adhesin-like protein	117*	41	0*
FD1_0683*'**	elongation factor Tu	2692*,**	227**	17*
FD1_0684*'**	serine/threonine phosphatase	31**	223*,**	0*
FD1_0691**	hypothetical protein	194**	2497**	0
FD1_0692**	DNA-directed RNA polymerase subunit sigma	153**	2049**	0
FD1_0693*'**	acetate kinase	283*,**	42**	0*
FD1_0721**	relaxase/mobilization nuclease family protein (pseudogene)	422**	76**	84
FD1_0722*'**	hypothetical protein	711*,**	155**	31*
FD1_0733*'**	hypothetical protein	249*,**	55**	0*
FD1_0740**	response regulator receiver/ANTAR domain- containing protein	106**	16**	0
FD1_0741**	hypothetical transmembrane protein	383**	24**	0
FD1_0742*'**	RNA polymerase sigma factor, sigma-70 family protein	265*,**	12**	0*
FD1_0743*'**	glycoside hydrolase family 43	134*,**	38**	14*
FD1_0744*'**	citrate transporter	331*,**	19**	0*
FD1_0745**	hypothetical transmembrane protein	143**	10**	0
FD1_0746*'**	beta-galactosidase	50*,**	11**	0*
FD1_0748**	alpha-*N*-arabinofuranosidase	67**	4**	6
FD1_0753**	xylanase	186**	16**	13
FD1_0754**	glycerate kinase	243**	11**	0
FD1_0761**	oxidoreductase	42**	6**	0
FD1_0766**	hypothetical protein	249**	44**	0
FD1_0789**	cell division protein FtsH	56**	5**	0
FD1_0805**	aldo/keto reductase (pseudogene)	52**	7**	0
FD1_0844*'**	Fdx	120*,**	16**	0*
FD1_0862**	RNA polymerase subunit sigma	26**	346**	0
FD1_0867*'**	XRE family transcriptional regulator	99*,**	10**	0*
FD1_0895*	ABC transporter	41	771*	0*
FD1_0901**	hypothetical protein	22**	576**	24
FD1_0902**	restriction endonuclease-like protein	16**	372**	24
FD1_0908**	twitching motility protein pilT	175**	31**	14
FD1_0909*'**	type IV-A pilus assembly ATPase PilB	76*,**	9**	0*
FD1_0910*'**	type II secretory pathway prepilin signal peptidase PulO and related peptidase	1530*,**	260**	21*
FD1_0911*'**	prepilin-type cleavage/methylation protein	1380*,**	203**	50*
FD1_0913*'**	adhesin-like protein	541*,**	111**	22*
FD1_0914*	hypothetical protein	1591*	572	20*
FD1_0915*'**	type IV pilus assembly protein PilM	379*,**	71**	0*
FD1_0916*'**	type II secretion protein F	383*,**	77**	0*
FD1_0918*	spore coat protein cotH	52*	21	5*
FD1_0919**	hypothetical protein	59**	5**	22
FD1_0945*	toxic anion resistance protein TelA	41*	9	0*
FD1_0946*'**	tellurium resistance protein TerD	287*,**	7**	0*
FD1_0953*'**	calcium-translocating P-type ATPase	44*,**	3**	0*
FD1_0967*'**	Ig domain protein group 2 domain protein	286*,**	59**	0*
FD1_0970*	serine/threonine protein phosphatase	77*	16	0*
FD1_0971*	hypothetical protein	1082*	451	0*
FD1_0972*	thiazolinyl imide reductase (pseudogene)	436*	247	0*
FD1_0973*'**	thiazolinyl imide reductase (pseudogene)	890*,**	101**	35*
FD1_0975**	hypothetical protein	25**	0**	0
FD1_0986*	transposase	150*	46	0*
FD1_0997**	hypothetical protein	198**	4**	12
FD1_0998**	hypothetical protein	61**	10**	0
FD1_1015**	TrpR like protein, YerC/YecD	53**	0**	0
FD1_1016*'**	glycoside hydrolase family 9	129*,**	31**	2*

FD1_1021*'**	glycoside hydrolase (pseudogene)	123*,**	27**	0*
FD1_1023*'**	glycosyl hydrolase family 9	248*,**	39**	13*
FD1_1025**	translation initiation factor IF-3	284**	21**	0
FD1_1026**	50*S* ribosomal protein L359	319**	1**	0
FD1_1036**	signal recognition particle-docking protein FtsY	36**	2**	0
FD1_1050*'**	elongation factor P	305*,**	65**	0*
FD1_1051**	hypothetical protein	1884**	136**	0
FD1_1059**	biotin/acetyl-CoA-carboxylase ligase	61**	2**	0
FD1_1061**	cytochrome b/b6	144**	16**	56
FD1_1065*'**	GerA spore germination protein	80*,**	16**	0*
FD1_1066**	hypothetical transmembrane protein	73**	10**	0
FD1_1068*	recombinase	50*	14	0*
FD1_1069**	hypothetical protein	61**	7**	0
FD1_1083*	hypothetical protein	260*	101	0*
FD1_1084*	hypothetical protein	205*	50	7*
FD1_1085*'**	large mechanosensitive ion channel protein MscL	427*,**	98**	19*
FD1_1087*'**	gDSL-like Lipase/Acylhydrolase	142*,**	36**	18*
FD1_1093**	FAD-dependent oxidoreductase	80**	4**	0
FD1_1100*'**	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	136*,**	24**	0*
FD1_1101*'**	glycoside hydrolase family 9	268*,**	12**	0*
FD1_1102*	glycoside hydrolase family 9	3117*	909	58*
FD1_1103*	hypothetical protein	61*	12	0*
FD1_1175**	glutamate dehydrogenase	82**	3**	0
FD1_1181**	FolC bifunctional protein	144**	20**	0
FD1_1186*	adhesin-like protein	1379	3216*	19*
FD1_1188*	ATP-dependent DNA helicase PcrA	676	1323*	8*
FD1_1199*'**	hypothetical protein	169*,**	33**	0*
FD1_1201*	two component sensor kinase	46*	15	0*
FD1_1202**	hypothetical protein	172**	39**	24
FD1_1205*	hypothetical protein	37*	10	0*
FD1_1220*'**	pyruvate phosphate dikinase	438*,**	29**	5*
FD1_1226*'**	hypothetical protein	1646*,**	335**	40*
FD1_1230**	hypothetical protein	40**	0**	0
FD1_1231**	hypothetical protein	109**	0**	0
FD1_1255*	efflux ABC transporter, permease	36*	11	0*
FD1_1258**	galactokinase	118**	7**	0
FD1_1265**	cell division ATP-binding protein FtsE	51**	6**	0
FD1_1266**	protein insertion permease FtsX	63**	1**	0
FD1_1280*'**	polysaccharide lyase	441*,**	58**	12*
FD1_1288**	acetyl xylan esterase	34**	5**	0
FD1_1292*'**	1,4-beta-xylanase	124*,**	17**	0*
FD1_1294**	hypothetical transmembrane protein	79**	5**	0
FD1_1296**	hypothetical transmembrane protein	204**	25**	0
FD1_1306**	amidohydrolase	103**	4**	0
FD1_1340*'**	hypoxanthine phosphoribosyltransferase	127*,**	25**	0*
FD1_1341*	adenine permease	161*	46	11*
FD1_1342*'**	hypothetical protein	77*,**	9**	0*
FD1_1343*'**	tRNA-splicing ligase RtcB	152*,**	26**	0*
FD1_1347*,**	fibronectin (pseudogene)	306*,**	49**	0*
FD1_1351*	cellulase (glycosyl hydrolase family 5)	65*	38	0*
FD1_1356*'**	GDSL family lipase	261*,**	47**	0*
FD1_1357*'**	hypothetical transmembrane protein	447*,**	32**	0*
FD1_1365**	carbonic anhydrase	304**	23**	0
FD1_1388*'**	hypothetical secreted protein	86*,**	20**	0*
FD1_1393*	metallo-beta-lactamase	119*	88	0*
FD1_1400*	nuclease	150*	71	0*
FD1_1406**	glycogen synthase	71**	4**	0
FD1_1410*'**	DNA-binding protein	958*,**	208**	0*
FD1_1412*	hypothetical protein	130*	83	0*
FD1_1413**	hemolysin activation protein	69**	7**	0
FD1_1415*'**	cytidylate kinase	510*,**	98**	0*
FD1_1419*	plectin	299*	131	0*
FD1_1420*	(4Fe-4S)-binding protein	840*	319	28*
FD1_1468*	anaerobic ribonucleoside-triphosphate reductase	326*	95	0*
FD1_1469*'**	rubrerythrin	1562*,**	261**	58*
FD1_1472*'**	hypothetical protein	174*,**	44**	12*
FD1_1473*'**	sugar ABC transporter substrate-binding protein	146*,**	25**	0*
FD1_1478**	asparagine synthase	39**	2**	0
FD1_1481*	hypothetical protein	228*	52	0*
FD1_1482**	hypothetical protein	396**	21**	73
FD1_1483*'**	peptidase S1 and S6 chymotrypsin/Hap	182*,**	15**	0*
FD1_1486*	ATP synthase subunit A	36*	18	0*
FD1_1500*	alpha-L-arabinofuranosidase	162*	39	0*
FD1_1506*	orotate phosphoribosyltransferase	115*	35	0*
FD1_1519*	hypothetical transmembrane protein	26*	5	0*

FD1_1525*'**	glycoside hydrolase (pseudogene)	5**	27*,**	0*
FD1_1539*'**	transposase IS116/IS110/IS902 family protein	937*,**	143**	0*
FD1_1546**	aspartate--ammonia ligase	103**	8**	27
FD1_1547*'**	hypothetical protein	979*,**	49**	0*
FD1_1548**	DNA mistmatch repair protein MutT	169**	14**	0
FD1_1558**	DNA-binding protein	154**	21**	0
FD1_1559*	RNA-binding protein S4	463*	167	0*
FD1_1593*'**	alpha-amylase	108**	265*,**	0*
FD1_1611**	CRISPR-associated protein Csd1	23**	3**	17
FD1_1614*	DEAD/DEAH box helicase	36*	12	0*
FD1_1619*	dockerin type I repeat-containing domain protein	23	44*	0*
FD1_1623*	ABC transporter permease	42*	11	0*
FD1_1626*'**	histidine kinase	124*,**	21**	0*
FD1_1629**	dinitrogenase iron-molybdenum cofactor biosynthesis protein	149**	16**	0
FD1_1631**	metallo-beta-lactamase	173**	9**	0
FD1_1633**	long-chain-fatty-acid--CoA ligase (pseudogene)	56**	10**	0
FD1_1641*	transglutaminase/protease	77*	62	5*
FD1_1659**	GNAT family acetyltransferase	47**	6**	0
FD1_1681*	GTPase HflX	90	124*	0*
FD1_1689**	conjugal transfer protein TraX	91**	9**	0
FD1_1705**	transcriptional regulator	259**	34**	0
FD1_1719**	preprotein translocase subunit SecY	93**	0**	0
FD1_1731**	50S ribosomal protein L29	52**	0**	0
FD1_1780**	alpha-glucan phosphorylase (pseudogene)	19**	2**	0
FD1_1799**	glycoside hydrolase family 3	50**	7**	0
FD1_1801*	ricin-type beta-trefoil lectin domain protein	188*	60	0*
FD1_1803*'**	glycoside hydrolase family protein	78*,**	16**	0*
FD1_1814*'**	hypothetical secreted protein	181*,**	40**	0*
FD1_1816*	hypothetical protein	22	55*	0*
FD1_1828*	hypothetical protein	190*	137	0*
FD1_1831*'**	glycosyl hydrolase family 9	117**	388*,**	4*
FD1_1832*'**	adhesin-like protein	77**	413*,**	0*
FD1_1833*	sortase	35	168*	0*
FD1_1834*	hypothetical secreted protein	58	121*	0*
FD1_1835*'**	hypothetical transmembrane protein	4**	50*,**	0*
FD1_1838**	transposase	151**	22**	0
FD1_1842**	diguanylate cyclase	59**	6**	10
FD1_1843**	ABC transporter substrate-binding protein	135**	16**	21
FD1_1844**	carbohydrate-binding protein	84**	5**	0
FD1_1871**	integrase family protein	79**	0**	0
FD1_1883*	histidine kinase	38*	8	0*
FD1_1893*	hypothetical secreted protein	294*	111	0*
FD1_1900*'**	1-phosphofructokinase	5**	58*,**	0*
FD1_1903**	Rrf2 family transcriptional regulator	0**	63**	0
FD1_1910**	hypothetical secreted protein	108**	4**	0
FD1_1920*'**	hypothetical protein	1847*,**	330**	63*
FD1_1932*	hypothetical protein	196*	147	10*
FD1_1964**	hypothetical protein	104**	7**	0
FD1_1965*'**	type I site-specific deoxyribonuclease	333*,**	104**	16*
FD1_1966*	KAP P-loop domain protein	269*	122	0*
FD1_1967*	hypothetical protein	104	172*	0*
FD1_1976**	excisionase family DNA binding domain-containing protein	1396**	110**	85
FD1_1991**	resolvase	587**	135**	95
FD1_2075**	SPFH/Band 7/PHB domain protein	1937**	549**	0
FD1_2076*	PTS ascorbate transporter subunit IIC	646*	174	0*
FD1_2086*'**	histidine kinase-DNA gyrase B-and HSP90-like ATPase	8*,**	274*,**	81
FD1_2087**	lytTr DNA-binding domain protein	6**	88**	26
FD1_2088**	DNA-binding helix-turn-helix protein	350**	3483**	269
FD1_2091**	hypothetical protein	0**	15**	0
FD1_2103*	hypothetical transmembrane protein	26*	15	0*
FD1_2111**	protein-(glutamine-N^5) methyltransferase	27**	4**	2
FD1_2112**	histidine kinase	34**	6**	13
FD1_2131*	transposase	278*	206	0*
FD1_2147**	MIP family channel proteins	281**	22**	0
FD1_2153**	hypothetical protein	26**	3**	0
FD1_2167**	glmZ(sRNA)-inactivating NTPase	84**	12**	0
FD1_2177**	aminopeptidase	73**	3**	0
FD1_2178**	LysR family transcriptional regulator	1194**	373**	0
FD1_2179*'**	ATP-dependent helicase	654*,**	181**	20*
FD1_2180**	biotin carboxylase	575**	171**	0
FD1_2181**	tellurite-resistance protein	491**	108**	0
FD1_2186*'**	transglutaminase domain-containing protein	48*,**	7**	0*
FD1_2187*	hypothetical transmembrane protein	63*	30	0*
FD1_2191*'**	hypothetical protein	104**	1829*,**	0*
FD1_2192*'**	hypothetical protein	225**	942*,**	0*

FD1_2193*'**	hypothetical protein	76**	1219*,**	0*
FD1_2194*'**	hypothetical protein	73**	665*,**	0*
FD1_2195**	hypothetical protein	6**	509**	0
FD1_2196*'**	hypothetical protein	20**	321*,**	0*
FD1_2197*'**	hypothetical protein	65**	886*,**	0*
FD1_2200**	lantibiotic lacticin	0**	184**	0
FD1_2201**	lantibiotic lacticin	0**	162**	0
FD1_2202*'**	hypothetical protein	45**	378*,**	0*
FD1_2203**	NHLM bacteriocin system ABC transporter, peptidase/ATP-binding protein	15**	607**	17
FD1_2204*'**	type 2 lantibiotic biosynthesis protein LanM	19**	1113*,**	6*
FD1_2205*'**	type 2 lantibiotic biosynthesis protein LanM	26**	1029*,**	6*
FD1_2206**	hypothetical transmembrane protein	43**	1406**	53
FD1_2207**	ABC transporter substrate-binding protein	70**	1218**	0
FD1_2208**	ABC-2 type transport system permease	96**	4958**	102
FD1_2209*'**	lantibiotic protection ABC transporter, ATP-binding subunit	90**	4482*,**	54*
FD1_2211*	hypothetical protein	453*	203	0*
FD1_2212*	hypothetical protein	631*	241	35*
FD1_2222**	30*S* ribosomal protein S9	235**	8**	0
FD1_2223*	50*S* ribosomal protein L13	167*	62	0*
FD1_2229*	hypothetical protein	34*	10	0*
FD1_2230*	hypothetical protein	729*	453	72*
FD1_2231*	molybdopterin-guanine dinucleotide biosynthesis protein MobA	322*	215	6*
FD1_2236*	ABC transporter substrate-binding protein	88	158*	0*
FD1_2241*	hypothetical transmembrane protein	470	1136*	97*
FD1_2243**	GHH signature containing HNH/Endo VII supernuclease toxin family protein	349**	1236**	85
FD1_2244**	hypothetical protein	339**	833**	41
FD1_2260**	nitrogen-fixing protein NifU	121**	21**	0
FD1_2264*'**	adhesin-like protein	101*,**	17**	0*
FD1_2265**	adhesin-like protein	47**	2**	0
FD1_2274*'**	large mechanosensitive ion channel protein MscL	427*,**	98**	19*
FD1_2275*	hypothetical protein	300*	102	2*
FD1_2292**	urea ABC transporter, permease protein UrtC	66**	4**	0
FD1_2293**	D-xylose ABC transporter, ATP-binding protein	41**	3**	9
FD1_2296*'**	phosphotransacetylase	117*,**	12**	0*
FD1_2301*	sugar ABC transporter permease	233	491*	0*
FD1_2302*	ABC transporter permease	130	270*	0*
FD1_2308**	multidrug transporter MatE	156**	4**	0
FD1_2310*'**	hypothetical protein	323*,**	12**	0*
FD1_2313**	protein kinase	513**	5**	0
FD1_2315*'**	periplasmic binding protein	519*,**	112**	12*
FD1_2316*'**	diguanylate cyclase	272*,**	31**	0*
FD1_2319**	carboxypeptidase	71**	3**	0
FD1_2341*	Z1 domain-containing protein	35	43*	0*
FD1_2347*'**	hypothetical protein	2910**	8038*,**	919*
FD1_2365**	von Willebrand factor A	31**	2**	0
FD1_2400**	hypothetical protein	666**	24**	0
FD1_2409**	DNA-directed RNA polymerase subunit beta	48**	3**	0
FD1_2410*'**	DNA-directed RNA polymerase subunit beta'	69*,**	10**	0*
FD1_2411*	metallo-beta-lactamase	138*	100	0*
FD1_2420**	hypothetical transmembrane protein	98**	3**	0
FD1_2423**	phosphomethylpyrimidine synthase ThiC	55**	4**	0
FD1_2424*	beta-mannosidase	198*	75	0*
FD1_2425*	lysophospholipase L1 and related esterase	95*	40	0*
FD1_2429**	anthranilate synthase subunit II	113**	13**	0
FD1_2430*	anthranilate phosphoribosyltransferase	149*	29	0*
FD1_2434*'**	tryptophan synthase subunit beta	80*,**	15**	0*
FD1_2435**	tryptophan synthase subunit alpha	69**	4**	0
FD1_2436*	peptidase S1 and S6 chymotrypsin/Hap	86*	47	0*
FD1_2447*'**	pyruvate formate lyase-activating protein	924*,**	286**	0*
FD1_2479*	gCN5-related N-acetyltransferase	1103*	524	0*
FD1_2481**	hypothetical transmembrane protein	144**	1**	0
FD1_2497*'**	helicase domain-containing protein	22*,**	3**	0*
FD1_2505*'**	phage integrase family protein	326*,**	90**	0*
FD1_2506**	hypothetical protein	95**	8**	0
FD1_2507**	hypothetical protein	124**	6**	0
FD1_2508*	repressor LexA	188*	56	0*
FD1_2514*'**	hypothetical transmembrane protein	395*,**	67**	0*
FD1_2519*	ChW repeat-/cell adhesion domain-containing transglutaminase-like protease	134	327*	0*
FD1_2520**	aminotransferase	44**	4**	0
FD1_2522*'**	hypothetical transmembrane protein	31*,**	4**	0*
FD1_2526**	hAD-superfamily phosphatase subfamily IIIC/FkbH-like domain	27**	2*	0
FD1_2530*	peptidase (pseudogene)	493*	210	0*

FD1_2531*	peptidase (pseudogene)	398*	123	0*
FD1_2532*	hypothetical protein	509*	206	0*
FD1_2549*'**	hypothetical protein	1761*,**	76**	0*
FD1_ 2550**	formate *C*-acetyltransferase glycine radical	133**	16**	0
FD1_2552*	formate acetyltransferase	77*	17	0*
FD1_2554*'**	diguanylate cyclase (GGDEF) domain-containing protein	78*,**	11**	0*
FD1_2623**	sulfide dehydrogenase	75**	13**	0
FD1_2630**	hypothetical protein	187**	794**	0
FD1_2631**	ankyrin	92**	373**	0
FD1_2632**	TetR family transcriptional regulator	26**	301**	49
FD1_2644**	hypothetical protein	45**	1**	0
FD1_2645**	DNA gyrase subunit B	23**	2**	0
FD1_2651**	chromosomal replication initiator protein DnaA	113**	9**	0
FD1_2667**	flavodoxin	66**	8**	0
FD1_2680**	hypothetical protein	21**	0**	0
FD1_2683*	molecular chaperone	78*	16	0*
FD1_2685*'**	DNA-3-methyladenine glycosylase II	29**	147*,**	0*
FD1_2691*'**	hypothetical protein	424*,**	39**	0*
FD1_2692*'**	cell division protein FtsZ	231*,**	13**	0*
FD1_2697**	cell division protein FtsW	20**	3**	0
FD1_2715**	septation protein spoVG	357**	19**	0
FD1_2737**	hypothetical transmembrane protein	60**	11**	0
FD1_2743*	hypothetical secreted protein	106*	54	0*
FD1_2750*'**	endoglucanase family protein	258*,**	74**	0*
FD1_2757**	polyphosphate:nucleotide phosphotransferase	48**	5**	0
FD1_2759*'**	ricin B lectin (pseudogene)	635*,**	115**	12*
FD1_2760*'**	ricin B lectin (pseudogene)	1107*,**	205**	0*
FD1_2763**	collagen-binding protein A	148**	16**	17
FD1_2778*'**	hypothetical protein	464*,**	57**	0*
FD1_2779*'**	histidinol phosphate phosphatase HisJ family	527*,**	111**	0*
FD1_2780*	hypothetical transmembrane protein	160*	75	0*
FD1_2783*'**	ABC-type antimicrobial peptide transport system permease component	207*,**	33**	21*
FD1_2784*	transcriptional regulator	529*	269	0*
FD1_2785*	periplasmic sensor signal transduction histidine kinase	697*	299	67*
FD1_2786*	hypothetical protein	727*	235	0*
FD1_2803**	HNH endonuclease	147**	10**	0
FD1_2809**	virulence-related protein	146**	5**	35
FD1_2810**	hypothetical protein	342**	55**	0
FD1_2818*	phage Terminase (pseudogene)	40*	8	0*
FD1_2823**	phage major capsid protein, HK97 family	33**	3**	0
FD1_2826*	prophage pi2 protein 37	229*	60	0*
FD1_2832*	phage tail length tape measure protein	87	142*	0*
FD1_2848**	hypothetical secreted protein	279**	66**	41
FD1_2849*	hypothetical transmembrane protein	162*	96	9*
FD1_2851*	histidine kinase	202*	97	10*
FD1_2854*'**	hypothetical protein	915*,**	247**	36*
FD1_2859*'**	1,4-alpha-glucan branching protein	9**	48*,**	0*
FD1_2863*	cellulose binding domain./glycosyl hydrolase family 9	137*	47	0*
FD1_2866*'**	DNA repair protein RadC	149*,**	8**	0*
FD1_2867*'**	hypothetical transmembrane protein	120*,**	12**	0*
FD1_2869*'**	*S*-adenosylmethionine synthetase	179*,**	13**	0*
FD1_2874*'**	6-phosphofructokinase	331*,**	52**	0*
FD1_2875**	phosphoglucomutase	69**	3**	0
FD1_2880**	deoxyguanosinetriphosphate triphosphohydrolase	112**	6**	0
FD1_2881**	DNA primase	59**	2**	0
FD1_2882*	RNA polymerase sigma factor RpoD	126*	92	11*
FD1_2888*'**	hypothetical transmembrane protein	147**	535*,**	0*
FD1_2889*	DNA-directed RNA polymerase subunit sigma	106	290*	0*
FD1_2901*'**	sortase, SrtB family	66*,**	4**	0*
FD1_2902*	hypothetical transmembrane protein	23	66*	0*
FD1_2909*	hypothetical transmembrane protein	130*	52	0*
FD1_2914*	hypothetical transmembrane protein	64	121*	0*
FD1_2915*'**	alternate signal-mediated exported protein	116**	471*,**	19*
FD1_2916*	signal peptidase	135	192*	0*
FD1_2924*	dihydroxy-acid dehydratase	40*	8	0*
FD1_2925**	3-isopropylmalate dehydrogenase	33**	5**	0
FD1_2927**	L-threonine ammonia-lyase	94**	16**	22
FD1_2928**	3-isopropylmalate dehydratase large subunit	111**	6**	0
FD1_2930**	endoribonuclease L-PSP	142**	0**	0
FD1_2931**	50S rRNA methyltransferase	132**	19**	0
FD1_2933*	cysteinyl-tRNA synthetase	58*	15	0*
FD1_2938**	quinolinate synthetase	66**	9**	0
FD1_2941*'**	hypothetical transmembrane protein	142*,**	17**	0*
FD1_2949**	hypothetical secreted protein	0**	16**	0

FD1_2971*'**	collagen-binding protein A	511**	1759*,**	9*
FD1_2973*'**	RNA polymerase, sigma-24 subunit, ECF subfamily protein	288**	1278*,**	0*
FD1_2974*	hypothetical transmembrane protein	167	562*	0*
FD1_2975*	hypothetical transmembrane protein	126	298*	0*
FD1_2977*	molecular chaperone GroEL	17	48*	0*
FD1_2979*	phosphoenolpyruvate carboxykinase	193*	65	28*
FD1_2981*'**	XRE family transcriptional regulator	335*,**	57**	0*
FD1_2986*'**	scaffoldin C	2374*,**	288**	32*
FD1_2987*'**	hypothetical protein	1252*,**	325**	15*
FD1_2988*	cohesin	932*	289	59*
FD1_2989**	cohesin-dockerin -x Domain Complex, Chain B	996**	16**	0
FD1_2990**	cohesin-dockerin -x Domain Complex, Chain A	1599**	66**	44
FD1_2992*	hypothetical protein	4260*	2933	218*
FD1_2994**	5,10-methylenetetrahydrofolate reductase	55**	7**	0
FD1_2997*	UDP-glucose 4-epimerase	253*	112	0*
FD1_3030*	mobilization protein	536*	289	0*
FD1_3031*'**	hypothetical protein	982*,**	159**	0*
FD1_3033*	integrase	338*	175	0*
FD1_3034**	DNA-binding protein	276**	33**	0
FD1_3037*'**	ECF subfamily RNA polymerase sigma-24 factor	26**	201*,**	0*
FD1_3038*'**	hypothetical protein	19**	222*,**	0*
FD1_3046*'**	hypothetical protein	166*,**	20**	0*
FD1_3048**	hypothetical protein	839**	10**	0
FD1_3049*'**	hypothetical transmembrane protein	701*,**	64**	19*
FD1_3050*'**	transposase	432*,**	96**	0*
FD1_3057*'**	transposase, IS116/IS110/IS902 family	848*,**	137**	20*
FD1_3069**	argininosuccinate synthase	177**	11**	0
FD1_3085*	hydrolase GDSL	213*	98	5*
FD1_3086*'**	diguanylate cyclase	134*,**	5**	0*
FD1_3088**	DNA mismatch repair protein	158**	2**	0
FD1_3092*'**	LacI family transcriptional regulator	141*,**	17**	0*
FD1_3093*'**	sodium:proton antiporter	273*,**	36**	14*
FD1_3097**	hypothetical protein	31**	5**	0
FD1_3104**	helicase c2	55**	1**	0
FD1_3106*'**	hypothetical protein	57408*,**	993**	97*
FD1_3120**	hypothetical protein	82**	0**	0
FD1_3126*	ammonium transporter	370*	192	15*
FD1_3127*	glutamine synthetase	979*	312	43*
FD1_3128**	hypothetical protein	67**	12**	0
FD1_3129*'**	hypothetical protein	3298*,**	805**	17*
FD1_3130**	hypothetical protein	13546**	2452**	0
FD1_3131*'**	hypothetical protein	16922*,**	3179**	114*
FD1_3132*'**	DNA-binding helix-turn-helix protein	290*,**	53**	0*
FD1_3133*'**	hypothetical protein	548*,**	80**	0*
FD1_3134**	DNA repair protein (pseudogene)	880**	168**	0
FD1_3136**	amidotransferase	216**	11**	0
FD1_3137**	transcriptional regulator, LuxR family protein	256**	10**	0
FD1_3138*'**	urea carboxylase-associated protein 2	419*,**	48**	0*
FD1_3139*'**	urea carboxylase-associated protein 1	350*,**	75**	0*
FD1_3140**	urea carboxylase	370**	22**	19
FD1_3146**	hypothetical protein	81**	9**	0
FD1_3148**	hypothetical protein	146**	20**	0
FD1_3149*'**	restriction-modification system control element Bcll	481**	1737*,**	140*
FD1_3150*'**	type II restriction endonuclease ScaI	361**	1089*,**	135*
FD1_3151**	hypothetical protein	455**	1049**	35
FD1_3154**	hypothetical protein	140**	11**	0
FD1_3160**	hypothetical protein	70**	0**	0
FD1_3162**	urea carboxylase	537**	19**	0
FD1_3163**	allophanate hydrolase	456**	11**	0
FD1_3164**	nitrogen regulatory protein PII	308**	14**	0
FD1_3165**	amino acid transporter	271**	6**	0
FD1_3171*'**	RNA polymerase subunit sigma-24	637*,**	160**	0*
FD1_3172*'**	hypothetical transmembrane protein	584*,**	49**	0*
FD1_3173*'**	hypothetical transmembrane protein	376*,**	51**	11*
FD1_3191**	methyltransferase	94**	6**	6
FD1_3196*	FeS assembly ATPase SufC	120*	77	0*
FD1_3200*'**	hypothetical transmembrane protein	15**	295*,**	0*
FD1_3206*	*O*-glycosyl hydrolase	99*	36	10*
FD1_3208*	ABC transporter substrate-binding protein	72	90*	0*
FD1_3209*	diguanylate cyclase	34*	15	0*
FD1_3210*'**	phospho-2-dehydro-3-deoxyheptonate aldolase	66*,**	9**	0*
FD1_3212**	glutamine synthetase type III (pseudogene)	7**	65**	0
FD1_3221*	transposase	150*	46	0*
FD1_3232**	1-phosphofructokinase	114**	5**	0
FD1_3233*	PTS fructose transporter subunit IIC	41*	11	0*

FD1_3254*	pentapeptide repeat-containing protein	95*	40	0*
FD1_3255*'**	L,D-transpeptidase catalytic domain protein	86**	357*,**	0*
FD1_3256*	hypothetical transmembrane protein	43	102*	0*
FD1_3280**	(p)ppGpp synthetase	37**	3**	0
FD1_3307*	mannan endo-1,4-beta-mannosidase	29*	6	0*
FD1_3312*'**	endo-arabinase	80*,**	8**	0*
FD1_3322**	transposase	151**	22**	0
FD1_3328*	hypothetical protein	20	67*	0*
FD1_3330*'**	hypothetical protein	31**	242*,**	0*
FD1_3332**	glutamine-fructose-6-phosphate transaminase (isomerizing)	37**	1**	0
FD1_3341*	30*S* ribosomal protein S15	98	159*	0*
FD1_3343*'**	hypothetical protein	1949*,**	541**	0*
FD1_3348**	serine hydroxymethyltransferase	95**	5**	0
FD1_3361**	permease	97**	9**	0
FD1_3362**	fumarate reductase	329**	15**	0
FD1_3365**	sulfite reductase, subunit A	103**	8**	0
FD1_3368**	heterodisulfide reductase subunit A	68**	4**	0
FD1_3369**	heterodisulfide reductase subunit C-like protein	84**	10**	0
FD1_3373**	isocitrate dehydrogenase	155**	16**	0
FD1_3377**	aconitate hydratase	73**	3**	0
FD1_3378*	diaminopimelate decarboxylase	235*	75	0*
FD1_3379**	ribosome biogenesis GTP-binding protein YsxC	229**	13**	55
FD1_3381**	Clp protease ATP-binding protein	219**	11**	12
FD1_3394**	GNAT family acetyltransferase	42**	3**	0
FD1_3400**	phosphoglycerol transferase	75**	15**	0
FD1_3406*'**	adhesin-like protein	238*,**	60**	4*
FD1_3407*	sortase	319*	94	0*
FD1_3408*'**	signal peptidase I	577*,**	149**	0*
FD1_3415*'**	ATPase AAA	17**	72*,**	0*
FD1_3451*	hypothetical protein	588*	197	0*
FD1_3452**	hypothetical protein	17**	131**	21
FD1_3453*	type-IV secretion system protein TraC	787	1329*	41*
FD1_3485*'**	glutamine--fructose-6-phosphate aminotransferase (pseudogene)	43*,**	6**	0*
FD1_3489*'**	integrase	131*,**	25**	0*
FD1_3498*	hypothetical protein	146*	42	0*
FD1_3504*'**	radical SAM peptide maturase, CXXX-repeat target family	178*,**	17**	0*
FD1_3515*	LIM zinc-binding protein	463*	150	0*
FD1_3518*	conjugative coupling factor TraD, SXT/TOL subfamily	25	82*	0*
FD1_3519**	hypothetical protein	22**	2**	0
FD1_3537**	VanZ-like protein	28**	0**	0
FD1_3546**	LICD family protein	60**	5**	0
FD1_3552**	UDP-*N*-acetylglucosamine 2-epimerase	30**	0**	0
FD1_3558**	helicase	85**	19**	9
FD1_3602*	hypothetical protein	420*	100	0*
FD1_3614**	LysR family transcriptional regulator	112**	9**	24
FD1_3615**	hypothetical transmembrane protein	105**	7**	0
FD1_3620*	hypothetical protein	65*	43	0*
FD1_3622**	hypothetical protein	3**	58**	0
FD1_3623*'**	2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase	10**	116*,**	0*
FD1_3632**	hypothetical transmembrane protein	291**	5968**	289
FD1_3633**	hypothetical secreted protein	109**	4690**	322
FD1_3634**	serine/threonine protein kinase	200**	3846**	270
FD1_3635*'**	antitoxin HicB	287**	2317*,**	124*
FD1_3636*'**	toxin HicA	186**	1863*,**	0*
FD1_3645**	hypothetical protein	0**	92**	0
FD1_3647**	hypothetical protein	5**	1375**	97
FD1_3648**	hypothetical protein	3**	827**	63
FD1_3649**	hypothetical protein	8**	465**	7
FD1_3675*'**	adhesin-like protein	99*,**	21**	0*
FD1_3676**	branched-chain amino acid aminotransferase	436**	5**	0
FD1_3680*'**	hypothetical protein	454*,**	41**	0*
FD1_3692**	acetolactate synthase, small subunit	98**	5**	0
FD1_3696*'**	hypothetical transmembrane protein	520*,**	61**	0*
FD1_3701*	ubiquinone/menaquinone biosynthesis methyltransferase	35	83*	0*
FD1_3702*'**	glycosyltransferase family 36	77*,**	13**	0*
FD1_3704**	citrate synthase	100**	6**	0
FD1_3710**	beta-1 4-xylanase	2128**	416**	0
FD1_3717**	NADH-dependent butanol dehydrogenase	189**	8**	0
FD1_3729**	hypothetical protein	378**	18**	0
FD1_3735**	glutathione peroxidase	0**	66**	0
FD1_3740**	hypothetical transmembrane protein	39**	4**	0
FD1_3744**	cotH protein	37**	1**	0
FD1_3746*	pseudouridine synthase	251*	66	0*
FD1_3747**	NADPH-dependent FMN reductase	240**	12**	0
FD1_3748*'**	sodium transporter	290*,**	42**	0*

FD1_3771*	glycoside hydrolase 48 family protein0	4143*	2079	67*
FD1_3776*'**	primosomal protein N'	35*,**	6**	0*
FD1_3783*'**	serine/threonine protein kinase	65*,**	4**	0*
FD1_3800*	isoleucyl-tRNA synthetase	31*	13	0*
FD1_3809*	DNA methylase	37	49*	0*
FD1_3815**	dockerin	42**	3**	0
FD1_3826**	hypothetical secreted protein	48**	5**	0
FD1_3828**	quinolinate synthetase	66**	9**	0
FD1_3833*	cysteinyl-tRNA synthetase	58*	15	0*
FD1_3835**	rRNA large subunit m3Psi methyltransferase RlmH	132**	19**	0
FD1_3836**	endoribonuclease L-PSP	142**	0**	0
FD1_3838**	3-isopropylmalate dehydratase large subunit	111**	6**	0
FD1_3839**	threonine ammonia-lyase	94**	16**	22
FD1_3841**	3-isopropylmalate dehydrogenase	33**	5**	0
FD1_3842*	dihydroxy-acid dehydratase	40*	8	0*
FD1_3868*'**	hypothetical protein	358*'**	67**	0*
FD1_3890**	30*S* ribosomal protein S16	32**	0**	0
FD1_3901**	adhesin-like protein	78**	345**	0
FD1_3921**	adhesin-like protein	6**	112**	53
FD1_3922**	ISPsy14, transposase family protein	119**	880**	0
FD1_3923*'**	ATPase AAA	144**	794*'**	0*
FD1_3924*'**	ABC transporter	198**	1435*'**	0*
FD1_3937**	hypothetical protein	66**	5**	0
FD1_3938**	RNA polymerase sigma factor, sigma-70 family	354**	49**	62
FD1_3941*'**	XRE family transcriptional regulator	515**	1555*'**	13*
FD1_3942*'**	adhesin-like protein	302**	1571*'**	53*
FD1_3943*'**	hypothetical protein	180**	750*'**	73*
FD1_3945*	adhesin-like protein	86	118*	0*
FD1_3946*	adhesin-like protein	543*	301	0*
FD1_3947*	hypothetical protein	459*	96	0*
FD1_3948*'**	accessory gene regulator	70*'**	10**	0*
FD1_3975*'**	ATPase AAA	6**	57*'**	0*
FD1_4004*'**	electron transport complex, RnfABCDGE type, B subunit	153*'**	15**	0*
FD1_4007*'**	transcriptional regulator, AraC family	241*'**	51**	0*
FD1_4015**	sulfate permease	61**	12**	0
FD1_4045*'**	hypothetical transmembrane protein	199*'**	22**	0*
FD1_4079**	resolvase	72**	2**	0
FD1_4080**	helicase SNF2	98**	27**	0
FD1_4093**	hypothetical transmembrane protein	362*	83*	0
FD1_4094**	type-IV secretion system protein TraC	499*	89*	0
FD1_4095**	conjugal transfer protein	396*	29*	0
FD1_4096*'**	conjugal transfer protein TrbL	326*'**	33**	0*
FD1_4097**	tRNA (adenine-N^6)-methyltransferase	288*	13*	0
FD1_4109*'**	type IV conjugative transfer system coupling protein TraD	229*'**	15**	0*
FD1_4110*	hypothetical protein	160	41	0
FD1_4113*'**	transcriptional regulator	111*'**	10**	0*
FD1_4114**	hypothetical protein	151**	14**	0
FD1_4116*	transposase	132*	44	0*
FD1_4123**	glyceraldehyde-3-phosphate dehydrogenase	3968**	46**	0
FD1_4128*'**	GTPase Era	219*'**	14**	0*
FD1_4131*'**	diguanylate cyclase (GGDEF) domain (pseudogene)	129*'**	20**	0*
FD1_4166*'**	cellulose 1,4-beta-cellobiosidase	191*'**	21**	0*

Trt2: ISO4-H5 enrichment culture +FD1 with high H_2 on methanol. Trt3: ISO4-H5 enrichment culture +FD1 with low H_2 on methanol. Trt4: ISO4-H5 enrichment culture +FD1 with low H_2 on monomethylamine. *: Genes that are above two fold differentially expressed with q<0.05 for both Krusgal-Wallis test and Benjamini-Hochberg test. **: Genes that are above two fold differentially expressed between Trt2 and Trt3, with q<0.05 for both Krusgal-Wallis test and Benjamini-Hochberg test excluding Trt4.

Table A.5.6. FD1 genes with 2-fold expression difference in Trt2 and Trt3

Locus_tag	Predicted gene product		
Genes with higher expression in Trt2		**Expression**	
Carbohydrate Utilisation		**Trt2**	**Trt3**
FD1_0599	sugar ABC transporter permease	270	46
FD1_0600	sugar ABC transporter permease	171	25
FD1_0621	beta-galactosidase	239	24
FD1_0641	sugar ABC transporter substrate-binding protein	181	2
FD1_0650	sugar ABC transporter permease	66	7
FD1_0651	carbohydrate ABC transporter substrate-binding protein	163	35
FD1_0743	glycoside hydrolase family 43	134	38
FD1_0746	beta-galactosidase	50	11
FD1_0748	alpha-*N*-arabinofuranosidase	67	4
FD1_0753	xylanase	186	16
FD1_0754	glycerate kinase	243	11
FD1_1016	glycoside hydrolase family 9	129	31
FD1_1021	glycoside hydrolase (pseudogene)	123	27
FD1_1023	glycosyl hydrolase family 9	248	39
FD1_1101	glycoside hydrolase family 9	268	12
FD1_1258	galactokinase	118	7
FD1_1280	polysaccharide lyase	441	58
FD1_1288	acetyl xylan esterase	34	5
FD1_1292	1,4-beta-xylanase	124	17
FD1_1473	sugar ABC transporter substrate-binding protein	146	25
FD1_1799	glycoside hydrolase family 3	50	7
FD1_1803	glycoside hydrolase family protein	78	16
FD1_1844	carbohydrate-binding protein	84	5
FD1_2293	D-xylose ABC transporter, ATP-binding protein	41	3
FD1_3702	glycosyltransferase family 36	77	13
FD1_3710	beta-1 4-xylanase	2128	416
FD1_4166	cellulose 1,4-beta-cellobiosidase	191	21
Cellulosome			
FD1_2986	scaffoldin C	2374	288
FD1_2987	hypothetical protein	1252	325
FD1_2989	cohesin-dockerin -x Domain Complex, Chain B	996	16
FD1_2990	cohesin-dockerin -x Domain Complex, Chain A	1599	66
Carbon metabolism			
FD1_0408	pyruvate:Fdx oxidoreductase and related 2-oxoacid:Fdx oxidoreductase beta subunit	82	7
FD1_0409	pyruvate:Fdx oxidoreductase and related 2-oxoacid:Fdx oxidoreductase gamma subunit	162	15
FD1_0431	carbohydrate binding module (family 6)/glycoside hydrolase 97	240	21
FD1_0635	1-phosphofructokinase	131	7
FD1_0693	acetate kinase	283	42
FD1_1181	FolC bifunctional protein	144	20
FD1_1365	carbonic anhydrase	304	23
FD1_1406	glycogen synthase	71	4
FD1_2447	pyruvate formate lyase-activating protein	924	286
FD1_2874	6-phosphofructokinase	331	52
FD1_2875	phosphoglucomutase	69	3
FD1_3069	argininosuccinate synthase	177	11
FD1_3362	fumarate reductase	329	15
FD1_3373	isocitrate dehydrogenase	155	16
FD1_3377	aconitate hydratase	73	3
FD1_3704	citrate synthase	100	6
FD1_3717	NADH-dependent butanol dehydrogenase	189	8
FD1_4123	glyceraldehyde-3-phosphate dehydrogenase	3968	46
Nitrogen metabolism			
FD1_1175	glutamate dehydrogenase	82	3
FD1_2260	nitrogen-fixing protein NifU	121	21
FD1_2292	urea ABC transporter, permease protein UrtC	66	4
FD1_3138	urea carboxylase-associated protein 2	419	48
FD1_3139	urea carboxylase-associated protein 1	350	75
FD1_3140	urea carboxylase	370	22
FD1_3162	urea carboxylase	537	19
FD1_3163	allophanate hydrolase	456	11
FD1_3164	nitrogen regulatory protein PII	308	14
FD1_3165	amino acid transporter	271	6
Electron transfer			
FD1_0455	Fe-S oxidoreductase	241	51
FD1_0586	flavodoxin	284	22
FD1_0844	Fdx	120	16
FD1_1061	cytochrome b/b6	144	16
FD1_1469	rubrerythrin	1562	261
FD1_2667	flavodoxin	66	8
FD1_4004	electron transport complex, RnfABCDGE type, B subunit	153	15

Transporters			
FD1_0436	multidrug transporter	328	5
FD1_0744	citrate transporter	331	19
FD1_1085	large mechanosensitive ion channel protein MscL	427	98
FD1_1266	protein insertion permease FtsX	63	1
FD1_2147	MIP family channel proteins	281	22
FD1_2274	large mechanosensitive ion channel protein MscL	427	98
FD1_3093	sodium:proton antiporter	273	36
FD1_3361	permease	97	9
FD1_3748	sodium transporter	290	42
FD1_4015	sulfate permease	61	12
Cell surface protein			
FD1_0908	twitching motility protein pilT	175	31
FD1_0909	type IV-A pilus assembly ATPase PilB	76	9
FD1_0910	type II secretory pathway prepilin signal peptidase PulO and related peptidase	1530	260
FD1_0911	prepilin-type cleavage/methylation protein	1380	203
FD1_0913	adhesin-like protein	541	111
FD1_0915	type IV pilus assembly protein PilM	379	71
FD1_0916	type II secretion protein F	383	77
FD1_0946	tellurium resistance protein TerD	287	7
FD1_0953	calcium-translocating P-type ATPase	44	3
FD1_0967	Ig domain protein group 2 domain protein	286	59
FD1_1689	conjugal transfer protein TraX	91	9
FD1_2181	tellurite-resistance protein	491	108
FD1_2264	adhesin-like protein	101	17
FD1_2265	adhesin-like protein	47	2
FD1_2315	periplasmic binding protein	519	112
FD1_2316	diguanylate cyclase	272	31
FD1_2763	collagen-binding protein A	148	16
FD1_3406	adhesin-like protein	238	60
FD1_3675	adhesin-like protein	99	21
FD1_3815	dockerin	42	3
FD1_4094	type-IV secretion system protein TraC	499	89
FD1_4095	conjugal transfer protein	396	29
FD1_4096	conjugal transfer protein TrbL	326	33
FD1_4109	type IV conjugative transfer system coupling protein TraD	229	15
Secretion system			
FD1_1036	signal recognition particle-docking protein FtsY	36	2
FD1_1719	preprotein translocase subunit SecY	93	0
FD1_2901	sortase, SrtB family	66	4
FD1_3408	signal peptidase I	577	149
FD1_3901	signal recognition particle protein	37	0
Cell replication and house-keeping genes			
FD1_0619	DNA repair protein RecN	107	13
FD1_0628	DNA mismatch repair protein MutS	44	3
FD1_0683	elongation factor Tu	2692	227
FD1_0742	RNA polymerase sigma factor, sigma-70 family protein	265	12
FD1_0789	cell division protein FtsH	56	5
FD1_1025	translation initiation factor IF-3	284	21
FD1_1026	50*S* ribosomal protein L359	319	1
FD1_1050	elongation factor P	305	65
FD1_1265	cell division ATP-binding protein FtsE	51	6
FD1_1343	tRNA-splicing ligase RtcB	152	26
FD1_1548	DNA mistmatch repair protein MutT	169	14
FD1_1731	50*S* ribosomal protein L29	52	0
FD1_2179	ATP-dependent helicase	654	181
FD1_2222	30*S* ribosomal protein S9	235	8
FD1_2409	DNA-directed RNA polymerase subunit beta	48	3
FD1_2410	DNA-directed RNA polymerase subunit beta'	69	10
FD1_2645	DNA gyrase subunit B	23	2
FD1_2651	chromosomal replication initiator protein DnaA	113	9
FD1_2692	cell division protein FtsZ	231	13
FD1_2697	cell division protein FtsW	20	3
FD1_2715	septation protein spoVG	357	19
FD1_2881	DNA primase	59	2
FD1_2931	50*S* rRNA methyltransferase	132	19
FD1_3104	helicase c2	55	1
FD1_3171	RNA polymerase subunit sigma-24	637	160
FD1_3379	ribosome biogenesis GTP-binding protein YsxC	229	13
FD1_3558	helicase	85	19
FD1_3835	rRNA large subunit m3Psi methyltransferase RlmH	132	19
FD1_3890	30*S* ribosomal protein S16	32	0
FD1_3938	RNA polymerase sigma factor, sigma-70 family	354	49
FD1_4079	resolvase	72	2
FD1_4080	helicase SNF2	98	27

Cofactor biosynthesis			
FD1_0292	cobalt-precorrin-6A synthase (pseudogene)	22	0
FD1_0537	phosphoglucosamine mutase	41	4
FD1_0617	NAD^+ kinase	102	7
FD1_1629	dinitrogenase iron-molybdenum cofactor biosynthesis protein	149	16
FD1_2869	*S*-adenosylmethionine synthetase	179	13
FD1_2938	quinolinate synthetase	66	9
FD1_2994	5,10-methylenetetrahydrofolate reductase	55	7
FD1_3828	quinolinate synthetase	66	9
Amino acid biosynthesis			
FD1_1546	aspartate--ammonia ligase	103	8
FD1_2429	anthranilate synthase subunit II	113	13
FD1_2434	tryptophan synthase subunit beta	80	15
FD1_2435	tryptophan synthase subunit alpha	69	4
FD1_2779	histidinol phosphate phosphatase HisJ family	527	111
FD1_2925	3-isopropylmalate dehydrogenase	33	5
FD1_2928	3-isopropylmalate dehydratase large subunit	111	6
FD1_3348	serine hydroxymethyltransferase	95	5
FD1_3676	branched-chain amino acid aminotransferase	436	5
FD1_3692	acetolactate synthase, small subunit	98	5
FD1_3838	3-isopropylmalate dehydratase large subunit	111	6
FD1_3841	3-isopropylmalate dehydrogenase	33	5
Genes with higher expression in Trt3			
Carbohydrate Utilisation		**Trt2**	**Trt3**
FD1_1525	glycoside hydrolase (pseudogene)	5	27
FD1_1593	alpha-amylase	108	265
FD1_1831	glycosyl hydrolase family 9	117	388
Lantibiotic island			
FD1_2191	hypothetical protein	104	1829
FD1_2192	hypothetical protein	225	942
FD1_2193	hypothetical protein	76	1219
FD1_2194	hypothetical protein	73	665
FD1_2195	hypothetical protein	6	509
FD1_2196	hypothetical protein	20	321
FD1_2197	hypothetical protein	65	886
FD1_2200	lantibiotic lacticin	0	184
FD1_2201	lantibiotic lacticin	0	162
FD1_2202	hypothetical protein	45	378
FD1_2203	NHLM bacteriocin system ABC transporter, peptidase/ATP-binding protein	15	607
FD1_2204	type 2 lantibiotic biosynthesis protein LanM	19	1113
FD1_2205	type 2 lantibiotic biosynthesis protein LanM	26	1029
FD1_2206	hypothetical transmembrane protein	43	1406
FD1_2207	ABC transporter substrate-binding protein	70	1218
FD1_2208	ABC-2 type transport system permease	96	4958
FD1_2209	lantibiotic protection ABC transporter, ATP-binding subunit	90	4482
Cell surface proteins			
FD1_1832	adhesin-like protein	77	413
FD1_2971	collagen-binding protein A	511	1759
FD1_3632	hypothetical transmembrane protein	291	5968
FD1_3921	adhesin-like protein	6	112
FD1_3942	adhesin-like protein	302	1571
Toxin/antitoxin system			
FD1_3635	antitoxin HicB	287	2317
FD1_3636	toxin HicA	186	1863
Transporter			
FD1_3924	ABC transporter	198	1435
Secreted proteins			
FD1_2915	alternate signal-mediated exported protein	116	471
FD1_3633	hypothetical secreted protein	109	4690
DNA modification			
FD1_0902	restriction endonuclease-like protein	16	372
FD1_3149	restriction-modification system control element Bcll	481	1737
FD1_3150	type II restriction endonuclease ScaI	361	1089
RNA polymerase			
FD1_0273	RNA polymerase sigma-70 factor	0	48
FD1_0692	DNA-directed RNA polymerase subunit sigma	153	2049
FD1_0862	RNA polymerase subunit sigma	26	346
FD1_2973	RNA polymerase, sigma-24 subunit, ECF subfamily protein	288	1278
FD1_3037	ECF subfamily RNA polymerase sigma-24 factor	26	201

Trt2: ISO4-H5 enrichment culture +FD1 with high H_2 on methanol. Trt3: ISO4-H5 enrichment culture +FD1 with low H_2 on methanol. Genes that are above two fold differentially expressed between Trt2 and Trt3, with $q<0.05$ for both Krusgal-Wallis test and Benjamini-Hochberg test excluding Trt4.

Table A.6.1. Manual functional annotation of the *Methanobrevibacter* sp. D5 predicted ORFs. Table excludes hypothetical proteins.

****

AMINO ACID METABOLISM

Alanine
D5_0658 alanine aminotransferase
Arginine
D5_1847 acetylglutamate kinase ArgB
D5_2420 acetylornithine aminotransferase ArgD
D5_1643 argininosuccinate lyase ArgH
D5_0217 argininosuccinate synthase ArgG
D5_1862 bifunctional ornithine acetyltransferase/N-acetylglutamate synthase protein ArgJ
D5_0549 *N*-acetyl-gamma-glutamyl-phosphate reductase ArgC
D5_0054 ornithine carbamoyltransferase ArgF
Aspartate/asparagine
D5_1577 asparagine synthase (glutamine-hydrolyzing) AsnB
Chorismate
D5_1672 3-dehydroquinate dehydratase type I AroD
D5_1468 3-dehydroquinate synthase AroB
D5_1705 3-phosphoshikimate 1-carboxyvinyltransferase AroA
D5_0787 chorismate synthase AroC
D5_0101 shikimate 5-dehydrogenase AroE
D5_0604 shikimate kinase AroK
Cysteine
D5_1336 cysteine synthase CysKM1
D5_1703 cysteine synthase CysKM2
D5_1702 serine *O*-acetyltransferase CysE
General
D5_0852 aspartate aminotransferase
Glutamate/glutamine
D5_0429 glutamate dehydrogenase GdhA
D5_1876 glutamate synthase domain containing protein
D5_0780 glutamine amidotransferase
D5_2271 glutamine amidotransferase
D5_2378 glutamine synthetase GlnA
D5_2379 transcriptional repressor of nif and glnA operons NrpR
Glycine
D5_2464 serine hydroxymethyltransferase GlyA
Histidine
D5_0004 ATP phosphoribosyltransferase HisG1
D5_0992 ATP phosphoribosyltransferase HisG2
D5_0078 bifunctional imidazoleglycerol-phosphate dehydratase HisB
D5_1014 hisA/hisF family protein HisAF
D5_0040 histidinol dehydrogenase HisD
D5_2254 histidinol-phosphate aminotransferase HisC
D5_0143 imidazole glycerol phosphate synthase glutamine amidotransferase subunit HisH
D5_2425 imidazoleglycerol-phosphate synthase cyclase subunit HisF
D5_0099 phosphoribosyl-AMP cyclohydrolase HisI
D5_0201 phosphoribosyl-ATP pyrophosphohydrolase HisE
D5_0551 phosphoribosylformimino-5- aminoimidazole carboxamide ribotide isomerase HisA
Homoserine
D5_1574 allosteric regulator of homoserine dehydrogenase
D5_1573 homoserine dehydrogenase Hom
Lysine
D5_0440 aspartate kinase Ask
D5_0608 aspartate kinase Ask
D5_0609 dihydrodipicolinate synthase DapA
D5_0610 dihydrodipicolinate reductase DapB
D5_0611 aspartate-semialdehyde dehydrogenase Asd
D5_0757 diaminopimelate aminotransferase DapL
D5_2417 diaminopimelate decarboxylase LysA
D5_2416 diaminopimelate epimerase DapF
Methionine
D5_1070 cystathionine beta-lyase
D5_1588 homoserine *O*-acetyltransferase MetX1
D5_1798 homoserine *O*-acetyltransferase MetX2
D5_2016 homoserine *O*-acetyltransferase MetX3
D5_2017 homoserine *O*-acetyltransferase MetX4
D5_2019 homoserine *O*-acetyltransferase MetX5
D5_1522 methionine synthase MetE
D5_1635 *O*-acetylhomoserine/*O*-acetylserine sulfhydrylase MetZ/CysK1
D5_1699 *O*-acetylhomoserine/*O*-acetylserine sulfhydrylase MetZ/CysK2
Phenylalanine/tyrosine
D5_0606 chorismate mutase AroH
D5_0250 prephenate dehydratase PheA
D5_0938 prephenate dehydrogenase TyrA1
D5_2241 prephenate dehydrogenase TyrA2
Polyamines
D5_0524 arginase/agmatinase family protein
D5_2023 *N*-carbamoyl-D-amino acid amidohydrolase AguB
D5_0522 pyruvoyl-dependent arginine decarboxylase PdaD
Serine
D5_2285 aminotransferase class V family
D5_1964 phosphoglycerate dehydrogenase SerA
D5_2363 phosphoserine phosphatase SerB
Threonine
D5_1657 threonine synthase ThrC
Tryptophan
D5_0159 anthranilate phosphoribosyltransferase TrpD
D5_0157 anthranilate synthase component I TrpE
D5_0158 anthranilate synthase component II TrpG
D5_0160 indole-3-glycerol phosphate synthase TrpC
D5_0161 phosphoribosylanthranilate isomerase TrpF
D5_0163 tryptophan synthase alpha subunit TrpA
D5_0162 tryptophan synthase beta subunit TrpB
D5_0030 tryptophan synthase beta subunit TrpB1
D5_2234 tryptophan-binding repressor TrpY
Valine/leucine/isoleucine
D5_0026 2-isopropylmalate synthase LeuA
D5_2494 3-isopropylmalate dehydratase large subunit LeuC
D5_2493 3-isopropylmalate dehydratase small subunit LeuD
D5_2492 3-isopropylmalate dehydrogenase LeuB
D5_0057 acetolactate synthase large subunit IlvB1
D5_2290 acetolactate synthase large subunit IlvB2
D5_0058 acetolactate synthase small subunit IlvN
D5_0080 branched-chain-amino-acid aminotransferase IlvE
D5_1814 citramalate synthase CimA
D5_0039 dihydroxy-acid dehydratase IlvD
D5_0061 ketol-acid reductoisomerase IlvC
Salvage- general
D5_1890 indolepyruvate Fdx oxidoreductase alpha subunit IorA
D5_1889 indolepyruvate Fdx oxidoreductase beta subunit IorB
D5_1795 ketoisovalerate Fdx oxidoreductase alpha subunit VorA
D5_1794 ketoisovalerate Fdx oxidoreductase beta subunit VorB
D5_1793 ketoisovalerate Fdx oxidoreductase gamma subunit VorC
Salvage- methionine
D5_2274 methylthioadenosine phosphorylase MtnP
D5_2375 *S*-adenosyl-L-homocysteine hydrolase AhcY
D5_1338 *S*-adenosylmethionine synthetase MetK

CELL CYCLE

Cell division
D5_0031 cell division ATPase MinD
D5_2242 cell division control protein Cdc48
D5_2227 cell division protein FtsZ
D5_2239 cell division protein pelota PelA
Chromosome replication
D5_2246 ATP-dependent DNA ligase DnlI
D5_0001 cdc6 family replication initiation protein Cdc6-1
D5_2279 cdc6 family replication initiation protein Cdc6-2
D5_0261 DNA polymerase family B PolB1
D5_0803 DNA polymerase family B PolB2
D5_2404 DNA polymerase large subunit DP2 PolD2
D5_0164 DNA polymerase sliding clamp subunit PCNA family Pcn
D5_2507 DNA polymerase small subunit DP1 PolD1
D5_1934 DNA primase DnaG
D5_1490 DNA primase large subunit PriB
D5_1495 DNA primase small subunit PriA
D5_1959 DNA-binding protein
D5_2373 flap endonuclease Fen
D5_0515 HIRAN domain-containing protein
D5_2250 OB fold nucleic acid binding domain-containing protein
D5_2470 replication factor A

D5_0103 replication factor C large subunit RfcL
D5_0104 replication factor C small subunit RfcS
D5_2033 replicative DNA helicase Mcm
D5_0320 ribonuclease HII RnhB
D5_1611 ribonuclease HI

Genome segregation

D5_1415 chromosome partitioning ParA/MinD-like ATPase
D5_2361 DNA topoisomerase I TopA
D5_0347 DNA topoisomerase VI subunit A
D5_0348 DNA topoisomerase VI subunit B
D5_1538 DNA double-strand break repair protein Rad50

CELL ENVELOPE

Cell surface proteins

D5_0014 adhesin-like protein
D5_0015 adhesin-like protein
D5_0016 adhesin-like protein
D5_0036 adhesin-like protein
D5_0038 adhesin-like protein
D5_0051 adhesin-like protein
D5_0111 adhesin-like protein
D5_0112 adhesin-like protein
D5_0127 adhesin-like protein
D5_0173 adhesin-like protein
D5_0188 adhesin-like protein
D5_0189 adhesin-like protein
D5_0218 adhesin-like protein
D5_0219 adhesin-like protein
D5_0252 adhesin-like protein
D5_0262 adhesin-like protein
D5_0452 adhesin-like protein
D5_0453 adhesin-like protein
D5_0458 adhesin-like protein
D5_0459 adhesin-like protein
D5_0478 adhesin-like protein
D5_0640 adhesin-like protein
D5_0660 adhesin-like protein
D5_0715 adhesin-like protein
D5_0821 adhesin-like protein
D5_0825 adhesin-like protein
D5_0867 adhesin-like protein
D5_0965 adhesin-like protein
D5_0970 adhesin-like protein
D5_0971 adhesin-like protein
D5_0973 adhesin-like protein
D5_0974 adhesin-like protein
D5_0976 adhesin-like protein
D5_0978 adhesin-like protein
D5_0979 adhesin-like protein
D5_1071 adhesin-like protein
D5_1092 adhesin-like protein
D5_1152 adhesin-like protein
D5_1322 adhesin-like protein
D5_1393 adhesin-like protein
D5_1400 adhesin-like protein
D5_1406 adhesin-like protein
D5_1411 adhesin-like protein
D5_1412 adhesin-like protein
D5_1413 adhesin-like protein
D5_1423 adhesin-like protein
D5_1424 adhesin-like protein
D5_1473 adhesin-like protein
D5_1476 adhesin-like protein
D5_1494 adhesin-like protein
D5_1610 adhesin-like protein
D5_1618 adhesin-like protein
D5_1619 adhesin-like protein
D5_1715 adhesin-like protein
D5_1716 adhesin-like protein
D5_1746 adhesin-like protein
D5_1758 adhesin-like protein
D5_1764 adhesin-like protein
D5_1896 adhesin-like protein
D5_1912 adhesin-like protein
D5_1921 adhesin-like protein
D5_2120 adhesin-like protein
D5_2160 adhesin-like protein
D5_2173 adhesin-like protein
D5_2175 adhesin-like protein
D5_2176 adhesin-like protein
D5_2261 adhesin-like protein
D5_2264 adhesin-like protein
D5_2265 adhesin-like protein
D5_2266 adhesin-like protein
D5_2335 adhesin-like protein
D5_2337 adhesin-like protein
D5_2354 adhesin-like protein
D5_2356 adhesin-like protein
D5_2357 adhesin-like protein
D5_2407 adhesin-like protein

Expolysaccharide synthesis

D5_0041 acetyltransferase
D5_0199 polysaccharide/polyol phosphate ABC transporter ATP-binding protein
D5_0200 polysaccharide/polyol phosphate ABC transporter permease protein
D5_0223 glycosyl transferase GT2 family
D5_0447 glycosyl transferase GT2 family
D5_0448 polysaccharide deacetylase domain-containing protein
D5_0462 glycosyl transferase GT2 family
D5_0463 glycosyl transferase
D5_0465 glycosyl transferase
D5_0467 glycosyl transferase GT2 family
D5_0474 polysaccharide/polyol phosphate ABC transporter permease protein
D5_0475 polysaccharide/polyol phosphate ABC transporter ATP-binding protein
D5_0484 glycosyl transferase GT2 family
D5_0485 glycosyl transferase GT2 family
D5_0488 glycosyl transferase GT2 family
D5_0493 glycosyl transferase GT2 family
D5_0498 UDP-glucose pyrophosphorylase GalU
D5_0501 glycosyl transferase GT2 family
D5_0504 glycosyl transferase
D5_0542 UDP-glucose/GDP-mannose dehydrogenase
D5_0583 UDP-*N*-acetylglucosamine 2-epimerase
D5_0603 glycosyl transferase GT4 family
D5_0612 glycosyl transferase
D5_0762 glycosyl transferase GT2 family with response regulator receiver domain
D5_0817 UDP-galactopyranose mutase Glf
D5_0818 glycosyl transferase GT2 family
D5_0836 UDP-galactopyranose mutase Glf
D5_0840 glycosyl transferase
D5_0842 UDP-*N*-acetylglucosamine 2-epimerase
D5_0848 polysaccharide biosynthesis protein
D5_0849 glycosyl transferase
D5_0884 polysaccharide biosynthesis protein
D5_0986 UDP-glucose pyrophosphorylase GalU
D5_1003 glycosyl transferase GT2 family
D5_1286 glycosyl transferase GT2 family
D5_1287 glycosyl transferase GT2 family
D5_1288 glycosyl transferase GT2 family
D5_1289 glycosyl transferase GT2 family
D5_1290 glycosyl transferase GT2 family
D5_1291 glycosyl transferase GT2 family
D5_1292 UDP-glucose 6-dehydrogenase
D5_1293 glycosyl transferase GT2 family
D5_1305 glycosyl transferase GT2 family
D5_1307 glycosyl transferase GT2 family
D5_1609 response regulator/glycosyl transferase GT2 family protein
D5_1756 dolichol kinase
D5_1929 glycosyl transferase GT2 family
D5_2471 exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase
D5_2472 glycosyl transferase GT2 family
D5_2473 glycosyl transferase GT2 family
D5_2474 glycosyl transferase GT2 family
D5_2475 CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase
D5_2478 dTDP-glucose 4,6-dehydratase RfbB
D5_2479 dTDP-4-dehydrorhamnose 3,5-epimerase RfbC

D5_2480 glucose-1-phosphate thymidylyltransferase RfbA
D5_2483 dTDP-4-dehydrorhamnose reductase RfbD
D5_2484 UDP-*N*-acetyl-D-mannosaminuronate dehydrogenase

Pseudomurein biosynthesis

D5_0075 polysaccharide biosynthesis protein
D5_0081 undecaprenyl-diphosphatase UppP
D5_0091 cell wall biosynthesis protein Mur ligase family
D5_0092 cell wall biosynthesis protein Mur ligase family
D5_0520 cell wall biosynthesis protein Mur ligase family
D5_0534 phosphosugar isomerase
D5_0838 polysaccharide biosynthesis protein
D5_0854 lipopolysaccharide cholinephosphotransferase
D5_0866 glucosamine-fructose-6-phosphate aminotransferase GlmS
D5_0908 glucosamine-fructose-6-phosphate aminotransferase GlmS
D5_0948 cell wall biosynthesis protein Mur ligase family
D5_1033 UDP-glucose 4-epimerase GalE
D5_1120 glucosamine-fructose-6-phosphate aminotransferase GlmS
D5_1126 diacylglycerol kinase DagK
D5_1483 cell wall biosynthesis protein phospho-N-acetylmuramoyl-pentapeptide-transferase family
D5_1512 undecaprenyl pyrophosphate synthetase UppS
D5_1536 cell wall biosynthesis protein Mur ligase family
D5_1788 NAD-dependent epimerase/dehydratase
D5_1822 cell wall biosynthesis protein Mur ligase family
D5_1823 cell wall biosynthesis protein phospho-N-acetylmuramoyl-pentapeptide-transferase family
D5_2172 PeiW-related protein
D5_2237 cell wall biosynthesis protein UDP-glycosyltransferase family
D5_2238 cell wall biosynthesis protein
D5_2251 phosphoglucosamine mutase GlmM1
D5_2256 UDP-*N*-acetylglucosamine diphosphorylase/glucosamine-1-phosphate N-acetyltransferase GlmU
D5_2258 phosphoglucosamine mutase GlmM2
D5_2488 cell wall biosynthesis glycosyl transferase GT2 family

Sialic acid biosynthesis

D5_0365 pseudaminic acid CMP-transferase PseF
D5_0366 *N*-acetyl neuramic acid synthetase NeuB
D5_0841 sialic acid synthase

Teichoic acid biosynthesis

D5_0197 glycosyl transferase GT2 family/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase
D5_0446 cell wall teichoic acid glycosylation protein
D5_0550 glycerol-3-phosphate cytidylyltransferase
D5_0853 glycerol-3-phosphate cytidylyltransferase
D5_1302 alcohol dehydrogenase
D5_1303 LPS biosynthesis protein
D5_1857 alcohol dehydrogenase
D5_1858 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase IspD

CELLULAR PROCESSES

Oxidative stress response

D5_0115 NADH oxidase Nox
D5_0599 thioredoxin
D5_1807 thioredoxin-disulfide reductase TrxB
D5_1416 NADH oxidase Nox
D5_1478 $F_{420}H_2$ oxidase FprA
D5_1600 rubredoxin Rub1
D5_1601 rubredoxin Rub2
D5_1622 rubrerythrin Rbr1
D5_1623 rubrerythrin Rbr2
D5_1624 rubrerythrin Rbr3
D5_1696 desulfoferrodoxin Dfx
D5_2257 rubredoxin Rub
D5_2452 $F_{420}H_2$ oxidase FprA
D5_2453 rubrerythrin Rbr4

Stress response

D5_1620 bile salt hydrolase
D5_1030 universal stress protein UspA
D5_1283 universal stress protein UspA
D5_2005 UspA domain-containing protein

CENTRAL CARBON METABOLISM

Acetate

D5_0784 ADP-dependent acetyl-CoA synthetase Acs
D5_1792 acetyl-CoA synthetase Acs
D5_1883 cation/acetate symporter, actP
D5_2122 pyruvate Fdx oxidoreductase gamma subunit PorC
D5_2123 pyruvate Fdx oxidoreductase delta subunit PorD
D5_2124 pyruvate Fdx oxidoreductase alpha subunit PorA
D5_2125 pyruvate Fdx oxidoreductase beta subunit PorB
D5_2126 pyruvate Fdx oxidoreductase-associated PorE
D5_2127 pyruvate Fdx oxidoreductase-associated PorF

Aromatic compounds

D5_1516 carboxymuconolactone decarboxylase family protein PcaC1
D5_1732 carboxymuconolactone decarboxylase family protein PcaC2
D5_1877 carboxymuconolactone decarboxylase family protein PcaC3
D5_2292 4-oxalocrotonate tautomerase family enzyme DmpI1
D5_2321 4-oxalocrotonate tautomerase family enzyme DmpI2

Bicarbonate

D5_0059 carbonic anhydrase Cab
D5_1726 bicarbonate ABC transporter ATP-binding protein BtcA
D5_1727 bicarbonate ABC transporter permease protein BtcB
D5_0782 bicarbonate ABC transporter substrate-binding protein BtcC

Formate

D5_0837 pyruvate formate-lyase-activating enzyme PflA
D5_0839 pyruvate-formate lyase
D5_2253 pyruvate formate-lyase-activating enzyme PflA

Gluconeogenesis

D5_0312 phosphoenolpyruvate synthase PpsA
D5_0339 glyceraldehyde-3-phosphate dehydrogenase Gap
D5_0405 triose-phosphate isomerase TpiA
D5_0406 phosphoglycerate kinase Pgk
D5_0655 2-phosphoglycerate kinase Pgk
D5_0728 phosphopyruvate hydratase Eno
D5_1509 cyclic 2,3-diphosphoglycerate-synthetase CpgS
D5_1572 2,3-bisphosphoglycerate-independent phosphoglycerate mutase ApgM
D5_1909 2-phosphoglycerate kinase Pgk2A
D5_2214 pyruvate kinase PykA
D5_2218 fructose 1,6-bisphosphatase Fbp
D5_2259 2,3-bisphosphoglycerate-independent phosphoglycerate mutase ApgM
D5_1469 phospho-2-dehydro-3-deoxyheptonate aldolase/fructose-bisphosphate aldolase

Glycolate salvage pathway

D5_0363 phosphoglycolate phosphatase Gph
D5_1053 phosphoglycolate phosphatase Gph

Inositol biosynthesis

D5_0521 bifunctional inositol-1 monophosphatase/fructose-1,6-bisphosphatase/ATP-NAD kinase
D5_0878 inositol-phosphate phosphatase domain containing protein
D5_0370 myo-inositol-1-phosphate synthase

Propanoate

D5_1954 2-methylcitrate dehydratase PrpD
D5_1951 2-methylcitrate synthase/citrate synthase II PrpC/CitZ

PRPP synthesis

D5_0954 ribose-phosphate diphosphokinase Prs
D5_1719 ribose 5-phosphate isomerase A RpiA

Ribulose monophosphate pathway

D5_0070 bifunctional formaldehyde-activating enzyme/3- hexulose-6-phosphate synthase Fae/Hps
D5_1128 3-hexulose-6-phosphate isomerase Phi1
D5_2395 3-hexulose-6-phosphate isomerase Phi1

Tricarboxylic cycle

D5_0007 succinate dehydrogenase/fumarate reductase flavoprotein subunit SdhA
D5_0372 pyruvate carboxylase subunit B PycB
D5_0396 fumarate hydratase FumA
D5_0397 2-oxoglutarate Fdx oxidoreductase subunit delta KorD
D5_0398 2-oxoglutarate Fdx oxidoreductase subunit alpha KorA
D5_0399 2-oxoglutarate Fdx oxidoreductase subunit beta KorB
D5_0400 2-oxoglutarate Fdx oxidoreductase subunit gamma KorC
D5_0401 succinyl-CoA synthetase beta subunit SucC
D5_0674 fumarate hydratase FumA
D5_0678 pyruvate carboxylase subunit A PycA
D5_1041 succinate dehydrogenase/fumarate reductase flavoprotein subunit SdhA
D5_1042 succinate dehydrogenase/fumarate reductase iron-sulfur protein SdhB

D5_1372 malate dehydrogenase Mdh
D5_1670 succinate-CoA ligase alpha subunit SucD
D5_1899 succinate dehydrogenase/fumarate reductase iron-sulfur protein SdhB
D5_1952 fumarate hydratase FumA
D5_2128 fumarate hydratase FumA

Other

D5_0594 deoxyribose-phosphate aldolase DeoC

ENERGY METABOLISM

Electron transfer

D5_0436 4Fe-4S Fdx binding domain-containing protein
D5_0525 4Fe-4S Fdx binding domain-containing protein
D5_0576 4Fe-4S Fdx binding domain-containing protein
D5_0663 4Fe-4S Fdx iron-sulfur binding domain-containing protein
D5_1060 4Fe-4S binding domain-containing protein
D5_1006 iron-sulfur cluster-binding protein
D5_1382 4Fe-4S Fdx binding domain-containing protein
D5_1556 4Fe-4S binding domain-containing protein
D5_1639 4Fe-4S binding domain-containing protein
D5_1767 4Fe-4S Fdx iron-sulfur binding domain-containing protein
D5_1848 4Fe-4S Fdx binding domain-containing protein
D5_1901 4Fe-4S binding domain-containing protein
D5_2139 4Fe-4S Fdx iron-sulfur binding domain-containing protein
D5_2206 4Fe-4S Fdx iron-sulfur binding domain protein
D5_0003 flavodoxin domain-containing protein
D5_0077 Fdx
D5_0183 flavodoxin
D5_0548 flavodoxin
D5_0662 Fdx
D5_0729 Fdx
D5_0823 flavodoxin
D5_1031 Fdx
D5_1137 flavodoxin
D5_1477 flavodoxin
D5_2010 flavodoxin
D5_0641 thioredixon TrxA
D5_0642 cytochrome C-type biogenesis protein DsbD
D5_0074 NADPH-dependent FMN reductase
D5_0824 NADPH-dependent FMN reductase
D5_1064 NADPH-dependent FMN reductase
D5_1065 NADPH-dependent FMN reductase
D5_1326 NADPH-dependent FMN reductase
D5_1359 NADPH-dependent FMN reductase
D5_1385 NADPH-dependent FMN reductase
D5_1472 NADPH-dependent FMN reductase
D5_1571 NADPH-dependent FMN reductase
D5_1834 NADPH-dependent FMN reductase
D5_1835 NADPH-dependent FMN reductase
D5_2273 NADPH-dependent FMN reductase
D5_1938 A_1A_0 archaeal ATP synthase subunit D AhaD
D5_1939 A_1A_0 archaeal ATP synthase subunit B AhaB
D5_1940 A_1A_0 archaeal ATP synthase subunit A AhaA
D5_1941 A_1A_0 archaeal ATP synthase subunit F AhaF
D5_1942 A_1A_0 archaeal ATP synthase subunit C AhaC
D5_1943 A_1A_0 archaeal ATP synthase subunit E AhaE
D5_1944 A_1A_0 archaeal ATP synthase subunit K AhaK
D5_1945 A_1A_0 archaeal ATP synthase subunit I AhaI
D5_1946 A_1A_0 archaeal ATP synthase subunit H AhaH

Alcohol metabolism

D5_1461 NADPH-dependent F_{420} reductase NpdG

Formate metabolism

D5_0738 formate dehydrogenase alpha subunit FdhA1
D5_0739 formate dehydrogenase beta subunit FdhB1
D5_0772 formate dehydrogenase accessory protein FdhD
D5_2391 formate dehydrogenase beta subunit FdhB2
D5_2392 formate dehydrogenase alpha subunit FdhA2
D5_2393 formate/nitrite transporter FdhC
D5_2394 formate dehydrogenase accessory protein FdhD

H_2 metabolism

D5_0174 coenzyme F_{420} hydrogenase alpha subunit FrhA
D5_0175 coenzyme F_{420} hydrogenase delta subunit FrhD
D5_0176 coenzyme F_{420} hydrogenase gamma subunit FrhG
D5_0177 coenzyme F_{420} hydrogenase beta subunit FrhB
D5_0225 energy-converting hydrogenase B subunit A EhbA
D5_0226 energy-converting hydrogenase B subunit B EhbB
D5_0227 energy-converting hydrogenase B subunit C EhbC
D5_0228 energy-converting hydrogenase B subunit D EhbD
D5_0229 energy-converting hydrogenase B subunit E EhbE
D5_0230 energy-converting hydrogenase B subunit F EhbF
D5_0231 energy-converting hydrogenase B subunit G EhbG
D5_0232 energy-converting hydrogenase B subunit H EhbH
D5_0233 energy-converting hydrogenase B subunit I EhbI
D5_0234 energy-converting hydrogenase B subunit J EhbJ
D5_0235 energy-converting hydrogenase B subunit K EhbK
D5_0236 energy-converting hydrogenase B subunit L EhbL
D5_0237 energy-converting hydrogenase B subunit M EhbM
D5_0238 energy-converting hydrogenase B subunit N EhbN
D5_0239 energy-converting hydrogenase B, subunit O, EhbO
D5_0240 energy-converting hydrogenase B subunit P EhbP
D5_0241 energy-converting hydrogenase B subunit Q EhbQ
D5_0302 methyl viologen-reducing hydrogenase delta subunit MvhD
D5_0303 methyl viologen-reducing hydrogenase gamma subunit MvhG
D5_0304 methyl viologen-reducing hydrogenase alpha subunit MvhA
D5_0305 methyl viologen-reducing hydrogenase beta subunit MvhB
D5_0740 methyl viologen-reducing hydrogenase delta subunit MvhD
D5_1770 energy-converting hydrogenase A subunit R EhaR
D5_1771 energy-converting hydrogenase A subunit Q EhaQ
D5_1772 energy-converting hydrogenase A subunit P EhaP
D5_1773 energy-converting hydrogenase A subunit O EhaO
D5_1774 energy-converting hydrogenase A subunit N EhaN
D5_1775 energy-converting hydrogenase A subunit M EhaM
D5_1776 energy-converting hydrogenase A subunit L EhaL
D5_1777 energy-converting hydrogenase A subunit K EhaK
D5_1778 energy-converting hydrogenase A subunit J EhaJ
D5_1779 energy-converting hydrogenase A subunit I EhaI
D5_1780 energy-converting hydrogenase A subunit H EhaH
D5_1781 energy-converting hydrogenase A subunit G EhaG
D5_1782 energy-converting hydrogenase A subunit F EhaF
D5_1783 energy-converting hydrogenase A subunit E EhaE
D5_1784 energy-converting hydrogenase A subunit D EhaD
D5_1785 energy-converting hydrogenase A subunit C EhaC
D5_1786 energy-converting hydrogenase A subunit B EhaB
D5_1787 energy-converting hydrogenase A subunit A EhaA
D5_0144 hydrogenase expression/formation protein, HypE
D5_0194 hydrogenase maturation factor HypF
D5_0364 hydrogenase expression/formation protein HypD
D5_0807 hydrogenase expression/formation protein HypE
D5_1527 hydrogenase accessory protein HypB
D5_1528 hydrogenase nickel insertion protein HypA
D5_1825 hydrogenase maturation protease HycI
D5_2235 hydrogenase assembly chaperone HypC

Methanogenesis pathway

D5_2110 5,10-methylenetetrahydromethanopterin reductase Mer
D5_0649 CoB--CoM heterodisulfide reductase subunit C HdrC
D5_0650 CoB--CoM heterodisulfide reductase subunit B HdrB
D5_1005 CoB--CoM heterodisulfide reductase subunit B HdrB2
D5_1502 CoB--CoM heterodisulfide reductase subunit A HdrA1
D5_1503 CoB--CoM heterodisulfide reductase subunit B HdrB3
D5_1504 CoB--CoM heterodisulfide reductase subunit C HdrC2
D5_2467 CoB--CoM heterodisulfide reductase subunit A HdrA2
D5_1879 F_{420}-dependent N^5,N^{10}-methenyltetrahydromethanopterin reductase Hmd (pseudo)
D5_2132 F_{420}-dependent N^5,N^{10}--methenyltetrahydromethanopterin reductase Hmd
D5_0079 F_{420}-dependent methylenetetrahydromethanopterin dehydrogenase Mtd
D5_0741 F_{420}-dependent methylenetetrahydromethanopterin dehydrogenase Mtd
D5_0210 formylmethanofuran-H_4MPT formyltransferase Ftr
D5_1769 formylmethanofuran-H_4MPT formyltransferase Ftr2
D5_0284 methyl-CoM reductase beta subunit McrB
D5_0285 methyl-CoM reductase D subunit McrD
D5_0286 methyl-CoM reductase C subunit McrC
D5_0287 methyl-CoM reductase gamma subunit McrG
D5_0288 methyl-CoM reductase alpha subunit McrA
D5_0328 methyl-CoM reductase component A2 AtwA
D5_0742 methyl-CoM reductase II beta subunit MrtB
D5_0743 methyl-CoM reductase II D subunit MrtD
D5_0744 methyl-CoM reductase II gamma subunit MrtG

D5_0745 methyl-CoM reductase II alpha subunit MrtA
D5_0289 H_4MPT *S*-methyltransferase subunit E MtrE
D5_0290 H_4MPT *S*-methyltransferase subunit D MtrD
D5_0291 H_4MPT S-methyltransferase subunit C MtrC
D5_0292 H_4MPT S-methyltransferase subunit B MtrB
D5_0293 H_4MPT S-methyltransferase subunit A MtrA
D5_0294 H_4MPT S-methyltransferase subunit F MtrF
D5_0295 H_4MPT S-methyltransferase subunit G MtrG
D5_0296 H_4MPT S-methyltransferase subunit H MtrH
D5_0350 H_4MPT S-methyltransferase subunit H MtrH2
D5_0510 H_4MPT S-methyltransferase subunit A MtrA
D5_1751 tungsten formylmethanofuran dehydrogenase subunit E FwdE
D5_2382 tungsten formylmethanofuran dehydrogenase subunit C FwdC
D5_2383 tungsten formylmethanofuran dehydrogenase subunit A FwdA
D5_2384 tungsten formylmethanofuran dehydrogenase subunit B FwdB
D5_2385 tungsten formylmethanofuran dehydrogenase subunit D FwdD
D5_2386 tungsten formylmethanofuran dehydrogenase subunit G FwdG
D5_2387 tungsten formylmethanofuran dehydrogenase subunit F FwdF
D5_2388 tungsten formylmethanofuran dehydrogenase subunit H FwdH
D5_1626 methenyltetrahydromethanopterin cyclohydrolase Mch
D5_0737 FlpE-related protein

LIPID METABOLISM

Biosynthesis bacterial
D5_1077 3-oxoacyl-(acyl-carrier-protein) reductase FabG
D5_1126 diacylglycerol kinase DagK
Biosynthesis general
D5_1032 geranylgeranyl reductase family protein
Lipid backbone
D5_1721 NAD(P)-dependent glycerol-1-phosphate dehydrogenase EgsA
Phospholipid biosynthesis
D5_0317 phosphatidylserine synthase PssA
D5_0369 geranylgeranylglyceryl phosphate synthase
D5_0799 membrane-associated phospholipid phosphatase
D5_1038 phosphatidylglycerophosphate synthase PgsA
D5_1542 digeranylgeranylglyceryl phosphate synthase
D5_2212 phosphatidylglycerophosphate synthase PgsA
Mevalonate pathway
D5_0732 mevalonate kinase Mvk
D5_0733 isopentenyl diphosphate kinase
D5_0734 isopentenyl diphosphate delta-isomerase Fni
D5_0872 hydroxymethylglutaryl-CoA synthase
D5_0873 acetyl-CoA acetyltransferase
D5_1669 hydroxymethylglutaryl-CoA reductase (NADPH) HmgA
D5_0731 phosphomevalonate decarboxylase
Elongation of isoprenoid side chains
D5_0736 bifunctional short chain isoprenyl diphosphate synthase IdsA

MOBILE ELEMENTS

CRISPR-associated genes
D5_0883 CRISPR-associated protein Cas1
D5_0882 CRISPR-associated protein Cas2
D5_0811 CRISPR-associated protein Cas4
D5_1211 CRISPR-associated protein Cas6
D5_1924 CRISPR-associated protein TIGR02710 family
Transposase
D5_0019 transposase
D5_0441 transposase
D5_0443 transposase
D5_0444 transposase
D5_0497 transposase
D5_0508 transposase
D5_0892 transposase
D5_0935 transposase
D5_1076 transposase
D5_1096 transposase
D5_1170 transposase
D5_1214 transposase
D5_1234 transposase
D5_1395 transposase
D5_1396 transposase
D5_1474 transposase
D5_1475 transposase
D5_1499 transposase
D5_1615 transposase
D5_1717 transposase
D5_1718 transposase
D5_1760 transposase
D5_1922 transposase
D5_1989 transposase
D5_2336 transposase
D5_2465 transposase
D5_0469 transposase IS4 family
D5_1997 transposase IS4 family protein
D5_0509 transposase IS605 OrfB family
D5_1021 transposase IS605 OrfB family
D5_1323 transposase IS605 OrfB family
D5_1324 transposase IS605 OrfB family
D5_2267 transposase IS605 OrfB family
D5_0830 transposase IS605 OrfB family, partial
D5_0832 transposase IS605 OrfB family, partial
D5_0712 transposase, partial
D5_0713 transposase, partial
D5_0714 transposase, partial

NITROGEN METABOLISM

Fixation
D5_1039 4Fe-4S iron sulfur cluster binding protein NifH/frxC family
D5_1379 NifU-like FeS cluster assembly scaffold protein
D5_1697 NifU-like FeS cluster assembly scaffold protein
D5_1698 cysteine desulfurase NifS
General
D5_0048 nitroreductase family protein
D5_0505 nitroreductase family protein
D5_0797 nitroreductase family protein
D5_0798 nitroreductase family protein
D5_1069 nitroreductase family protein
D5_1117 nitroreductase family protein
D5_1744 nitroreductase family protein
D5_1915 nitroreductase family protein
D5_1950 nitroreductase family protein
Other
D5_1162 ADP-ribosylglycohydrolase family protein
D5_1163 ADP-ribosylglycohydrolase family protein
D5_0749 hydroxylamine reductase Hcp
Regulation
D5_0958 nitrogen regulatory protein P-II GlnK
D5_1674 nitrogen regulatory protein P-II GlnK
Transport
D5_0957 ammonium transporter Amt
D5_1675 ammonium transporter Amt

NUCLEIC ACID METABOLISM

DNA-binding proteins
D5_0005 archaeal histone
D5_0593 archaeal histone
D5_0027 DNA-binding protein AlbA
D5_2345 DNA-binding protein
D5_0595 histone acetyltransferase ELP3 family
D5_0215 NAD-dependent deacetylase Sir2 family
Helicase
D5_0216 DNA helicase
D5_0220 DNA helicase
D5_0592 DEAD/DEAH box helicase domain-containing protein
D5_0813 DEAD/DEAH box helicase domain-containing protein
D5_0966 DEAD/DEAH box helicase domain-containing protein
D5_1025 helicase
D5_1225 ATP-dependent DNA helicase PcrA
D5_1239 ATP-dependent DNA helicase UvrD/REP family
D5_1246 ATP-dependent DNA helicase UvrD/REP family
D5_1255 ATP-dependent DNA helicase UvrD/REP family
D5_1261 ATP-dependent DNA helicase
D5_1530 DEAD/DEAH box helicase domain-containing protein
D5_1531 ATP-dependent DNA helicase UvrD/REP family
D5_2011 helicase SNF2 family
D5_2026 DEAD/DEAH box helicase domain-containing protein
D5_2298 helicase
D5_2438 ATP-dependent DNA helicase

Recombination and repair
D5_0087 RdgB/HAM1 family non-canonical purine NTP pyrophosphatase
D5_0089 ssDNA exonuclease RecJ
D5_0096 Hef nuclease
D5_0205 archaeal Holliday junction resolvase Hjc
D5_0338 endonuclease IV
D5_0616 exodeoxyribonuclease VII small subunit XseB
D5_0617 exodeoxyribonuclease VII large subunit XseA
D5_0618 exodeoxyribonuclease VII large subunit XseA
D5_0630 staphylococcal nuclease domain-containing protein
D5_0769 excinuclease ABC A subunit UvrA
D5_0801 exodeoxyribonuclease III Xth
D5_0810 nuclease
D5_0815 Single strand DNA binding protein
D5_0956 excinuclease ABC B subunit UvrB
D5_0959 excinuclease ABC A subunit UvrA
D5_1026 excinuclease ABC C subunit UvrC
D5_1043 6-*O*-methylguanine DNA methyltransferase Ogt
D5_1072 NADH pyrophosphatase NudC
D5_1198 exodeoxyribonuclease III Xth
D5_1199 endonuclease III-related protein
D5_1221 DNA mismatch endonuclease Vsr
D5_1233 exodeoxyribonuclease VII large subunit XseA
D5_1236 exodeoxyribonuclease VII small subunit XseB
D5_1349 DNA mismatch repair ATPase MutS family
D5_1364 uracil-DNA glycosylase Ung
D5_1538 DNA double-strand break repair protein Rad50
D5_1539 DNA double-strand break repair protein Mre11
D5_1558 exonuclease
D5_1694 archaea-specific RecJ-like exonuclease
D5_1704 endonuclease III Nth
D5_1870 DNA-3-methyladenine glycosylase I Tag
D5_2041 DNA mismatch repair ATPase MutS family
D5_2083 DNA repair photolyase
D5_2093 GIY-YIG catalytic domain-containing endonuclease
D5_2094 GIY-YIG catalytic domain-containing endonuclease
D5_2208 DNA repair and recombination protein RadB
D5_2243 exonuclease
D5_2299 DNA mismatch repair protein MutT
D5_2424 8-oxoguanine DNA-glycosylase Ogg
D5_2469 DNA repair and recombination protein RadA
Restriction and modification
D5_0066 type II restriction endonuclease
D5_0404 DNA methylase
D5_0512 type I restriction-modification enzyme S subunit HsdS2
D5_0513 type I restriction-modification system M subunit HsdM2
D5_1219 DNA-cytosine methyltransferase Dcm
D5_1220 DNA-cytosine methyltransferase Dcm
D5_2082 restriction endonuclease
D5_2084 type I site-specific deoxyribonuclease HsdR family
D5_2085 type I restriction endonuclease subunit S
D5_2086 type I restriction endonuclease subunit M
D5_2165 type III restriction endonuclease res subunit
D5_2166 type III restriction endonuclease methylation subunit
D5_2307 DNA phosphorothioation system restriction enzyme

PROTEIN FATE
Protein degradation
D5_0086 glycoprotease M22 family
D5_0105 peptidase M48 family
D5_0178 methionine aminopeptidase Map
D5_0221 succinylglutamate desuccinylase/aspartoacylase
D5_0362 peptidase U62 family
D5_0379 peptidase M48 family
D5_0545 peptidase U62 family
D5_0947 ATP-dependent protease S16 family
D5_1037 peptidase U32
D5_1555 glutamyl aminopeptidase PepA
D5_1998 Xaa-Pro aminopeptidase
D5_1999 peptidase M50 family
D5_2039 peptidase U32 family
D5_2040 peptidase U32 family
D5_2049 transglutaminase domain-containing protein
D5_2454 peptidase M50 family
D5_0266 proteasome beta subunit
D5_1686 proteasome alpha subunit PsmA
D5_1817 proteasome-activating nucleotidase
D5_2399 proteasome-activating nucleotidase
Protein folding
D5_0190 molecular chaperone DnaJ
D5_0191 chaperone protein DnaK
D5_0192 molecular chaperone GrpE
D5_0395 peptidyl-prolyl cis-trans isomerase FKBP-type
D5_0535 heat shock protein Hsp20/alpha crystallin family
D5_0614 thermosome subunit
D5_1012 prefoldin beta subunit PfdB
D5_1428 chaperone protein DnaJ
D5_1663 thermosome subunit
D5_1867 nascent polypeptide-associated complex protein
D5_2339 prefoldin alpha subunit PfdA
D5_2421 peptidyl-prolyl cis-trans isomerase
Protein secretion
D5_0046 signal peptidase I
D5_0315 sortase family protein
D5_0704 preprotein translocase subunit SecY
D5_0816 signal recognition particle SRP19 protein
D5_1565 type IV leader peptidase family protein
D5_1673 signal peptidase I
D5_1979 type II secretion system protein F GspF
D5_2197 type II secretion system protein E GspE
D5_2198 type II secretion system protein F GspF
D5_2226 preprotein translocase subunit SecE
D5_2338 signal recognition particle receptor FtsY
D5_2360 oligosaccharyl transferase
D5_2426 preprotein translocase subunit SecG
D5_2429 signal recognition particle SRP54 protein

PROTEIN SYNTHESIS
Other
D5_0134 translation-associated GTPase
D5_0265 RNA-metabolising metallo-beta-lactamase
D5_0349 RNA-binding protein
D5_0735 RNA-metabolising metallo-beta-lactamase
D5_1596 RNA-binding protein
D5_1815 RNA methylase
D5_1826 RNA methylase
D5_2201 RNA-binding protein
D5_2204 ATPase RIL
D5_2211 Sua5/YciO/YrdC/YwlC family translation factor
D5_2232 RNA-binding protein
D5_2344 tRNA methyltransferase subunit
D5_2347 RNA-binding protein
D5_2412 RNA-binding protein
D5_2428 pseudouridylate synthase
D5_2481 ribonuclease inhibitor
D5_2482 ribonuclease
Ribosomal proteins
D5_0088 ribosomal protein S15P Rps15p
D5_0166 ribosomal protein L44e Rpl44e
D5_0167 ribosomal protein S27e Rps27e
D5_0311 ribosomal protein L10e Rpl10e
D5_0414 ribosomal protein L30e Rpl30e
D5_0416 ribosomal protein S12P Rps12p
D5_0417 ribosomal protein S7P Rps7p
D5_0420 ribosomal protein S10P Rps10p
D5_0539 ribosomal protein L40e Rpl40e
D5_0607 ribosomal protein S17e Rps17e
D5_0681 ribosomal protein L3P Rpl3p
D5_0682 ribosomal protein L4p Rpl4p
D5_0683 ribosomal protein L23P Rpl23p
D5_0684 ribosomal protein L2P Rpl2p
D5_0685 ribosomal protein S19P Rps19p
D5_0686 ribosomal protein L22P Rpl22p
D5_0687 ribosomal protein S3P Rps3p
D5_0688 ribosomal protein L29P Rpl29p
D5_0691 ribosomal protein S17P Rps17p
D5_0692 ribosomal protein L14P Rpl14p
D5_0693 ribosomal protein L24P Rpl24p
D5_0694 ribosomal protein S4e Rps4e

D5_0695 ribosomal protein L5P Rpl5p
D5_0696 ribosomal protein S8P Rps8p
D5_0697 ribosomal protein L6P Rpl6p
D5_0698 ribosomal protein L32e Rpl32e
D5_0699 ribosomal protein L19e Rpl19e
D5_0700 ribosomal protein L18P Rpl18p
D5_0701 ribosomal protein S5P Rps5p
D5_0702 ribosomal protein L30P Rpl30p
D5_0703 ribosomal protein L15P Rpl15p
D5_0707 ribosomal protein L34e Rpl34e
D5_0709 ribosomal protein L14e Rpl14e
D5_0719 ribosomal protein S13P Rps13p
D5_0720 ribosomal protein S4P Rps4p
D5_0721 ribosomal protein S11P Rps11p
D5_0723 ribosomal protein L18e Rpl18e
D5_0724 ribosomal protein L13P Rpl13p
D5_0725 ribosomal protein S9P Rps9p
D5_0730 ribosomal protein S2P Rps2p
D5_0806 ribosomal protein S8e Rps8e
D5_1009 ribosomal protein L37Ae Rpl37ae
D5_1543 ribosomal protein L40e Rpl40e
D5_1593 ribosomal protein L37e Rpl37e
D5_1644 ribosomal protein S27ae Rps27ae
D5_1645 ribosomal protein S24e Rps24e
D5_1651 ribosomal protein S6e Rps6e
D5_1654 ribosomal protein L24e Rpl24e
D5_1655 ribosomal protein S28e Rps28e
D5_1656 ribosomal protein L7Ae Rpl7ae
D5_1755 ribosomal protein L15e Rpl15e
D5_2221 ribosomal protein L12P Rpl12p
D5_2222 acidic ribosomal protein P0 RplPO
D5_2223 ribosomal protein L1P Rpl1p
D5_2224 ribosomal protein L11P Rpl11p
D5_2225 ribosomal protein L24 family
D5_2269 ribosomal protein S3Ae Rps3ae
D5_2340 ribosomal protein LX RplX
D5_2342 ribosomal protein L31e Rpl31e
D5_2343 ribosomal protein L39e Rpl39e
D5_2346 ribosomal protein S19e Rps19e
D5_2410 ribosomal protein L21e Rpl21e

RNA processing

D5_0106 tRNA(1-methyladenosine) methyltransferase
D5_0146 tRNA pseudouridine synthase D TruD
D5_0169 H/ACA RNA-protein complex component Nop10p
D5_0185 tRNA-dihydrouridine synthase DusA1
D5_0246 queuosine biosynthesis protein QueD
D5_0247 7-cyano-7-deazaguanosine biosynthesis protein QueE
D5_0255 fibrillarin FlpA
D5_0256 pre-mRNA splicing ribonucleoprotein PRP31
D5_0272 N^2,N^2-dimethylguanosine tRNA methyltransferase Trm1
D5_0327 tRNA-dihydrouridine synthase
D5_0382 queuosine biosynthesis protein QueC
D5_0425 ribosomal-protein-alanine acetyltransferase RimI
D5_0581 tRNA pseudouridine synthase A TruA
D5_0591 MiaB-like tRNA modifying enzyme
D5_0648 archaeosine tRNA-ribosyltransferase TgtA
D5_0679 ribosomal protein L11 methyltransferase PrmA
D5_0710 H/ACA RNA-protein complex component Cbf5p
D5_0870 archaeosine tRNA-ribosyltransferase TgtA
D5_0920 RNA-splicingl igase RtcB
D5_1466 tRNA nucleotidyltransferase Cca
D5_1467 2'-5' RNA ligase LigT
D5_1660 tRNA intron endonuclease EndA
D5_1682 exosome complex RNA-binding protein Rrp42
D5_1683 exosome complex exonuclease Rrp41
D5_1684 exosome complex RNA-binding protein Rrp4
D5_1685 exosome subunit
D5_0690 ribonuclease P subunit P29
D5_1687 ribonuclease P subunit P14
D5_1688 ribonuclease P subunit P30
D5_2348 ribonuclease P subunit RPR2
D5_1748 exosome subunit
D5_1900 RNA methyltransferase TrmH family
D5_1932 H/ACA RNA-protein complex component Gar1
D5_1966 tRNA(His) guanylyltransferase ThgL
D5_2013 ribonuclease Z Rnz
D5_2031 ribosomal RNA large subunit methyltransferase J RrmJ
D5_2035 NMD3 family protein
D5_2413 dimethyladenosine transferase KsgA
D5_2432 exosome complex RNA-binding protein Csl4

Translation factors

D5_0168 translation initiation factor aIF-2 alpha subunit
D5_0358 translation initiation factor aIF-1A
D5_0418 translation elongation factor aEF-2
D5_0419 translation elongation factor aEF-1 alpha
D5_0428 peptide chain release factor aRF1
D5_0523 translation initiation factor aIF-5A
D5_0639 translation initiation factor aIF-2B alpha subunit
D5_0644 diphthine synthase DphB
D5_0689 translation initiation factor aSUI1
D5_0911 translation elongation factor aEF-2
D5_0994 deoxyhypusine synthase Dys
D5_1650 translation initiation factor aIF-2 gamma subunit
D5_1652 translation initiation factor IF-2
D5_2034 translation initiation factor aIF-2 beta subunit
D5_2200 translation elongation factor aEF-1 beta
D5_2203 peptidyl-tRNA hydrolase
D5_2341 translation initiation factor aIF-6
D5_2431 diphthamide biosynthesis protein

tRNA aminoacylation

D5_0013 Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit A GatA
D5_0042 aspartyl-tRNA synthetase AspS
D5_0047 arginyl-tRNA synthetase ArgS
D5_0069 threonyl-tRNA synthetase ThrS
D5_0100 histidyl-tRNA synthetase HisS
D5_0107 leucyl-tRNA synthetase LeuS
D5_0203 Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit B GatB
D5_0754 glutamyl-tRNA synthetase GltX
D5_0800 phenylalanyl-tRNA synthetase alpha subunit PheS
D5_1062 seryl-tRNA synthetase SerS
D5_1370 D-tyrosyl-tRNA^{Tyr} deacylase Dtd
D5_1489 methionyl-tRNA synthetase MetG
D5_1575 Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit C GatC
D5_1659 tryptophanyl-tRNA synthetase TrpS
D5_1700 cysteinyl-tRNA synthetase CysS
D5_1708 valyl-tRNA synthetase ValS
D5_1713 phenylalanyl-tRNA synthetase subunit beta PheT
D5_1722 prolyl-tRNA synthetase ProS
D5_1796 glutamyl-tRNA^{Gln} amidotransferase subunit D GatD
D5_1797 glutamyl-tRNA^{Gln} amidotransferase subunit E GatE
D5_1903 glycyl-tRNA synthetase GlyS
D5_1969 tRNA binding domain-containing protein
D5_2036 tyrosyl-tRNA synthetase TyrS
D5_2220 alanyl-tRNA synthetase AlaS
D5_2400 lysyl-tRNA synthetase LysS
D5_2457 isoleucyl-tRNA synthetase IleS

PURINES AND PYRIMIDINES

Purine biosynthesis

D5_0052 phosphoribosylamine--glycine ligase, purD
D5_0153 adenylosuccinate lyase PurB
D5_0195 phosphoribosylaminoimidazole carboxylase catalytic subunit PurE
D5_0264 phosphoribosylformylglycinamidine cyclo-ligase PurM
D5_0323 IMP cyclohydrolase PurO
D5_0862 phosphoribosylaminoimidazole-succinocarboxamide synthase PurC
D5_0863 phosphoribosylformylglycinamidine (FGAM) synthase PurS
D5_0864 phosphoribosylformylglycinamidine (FGAM) synthase PurQ
D5_1036 amidophosphoribosyltransferase PurF
D5_2029 5-formaminoimidazole-4-carboxamide-1-(beta)-D- ribofuranosyl 5'-monophosphate-formate ligase PurP
D5_2430 adenine phosphoribosyltransferase Apt
D5_2456 phosphoribosylformylglycinamidine (FGAM) synthase II PurL
D5_2486 N^5-carboxyaminoimidazole ribonucleotide synthetase purK

Purine interconversion

D5_0532 adenine deaminase Ade
D5_0705 adenylate kinase Adk
D5_2367 adenylate cyclase CyaA
D5_0766 adenylosuccinate synthetase PurA

Pyrimidine biosynthesis

D5_0002 aspartate carbamoyltransferase

D5_0258 dihydroorotate dehydrogenase PyrD
D5_0259 dihydroorotate dehydrogenase electron transfer subunit PyrK
D5_0307 dihydroorotase PyrC
D5_0517 orotate phosphoribosyltransferase PyrE
D5_0547 aspartate carbamoyltransferase regulatory subunit PyrI
D5_0620 orotate phosphoribosyltransferase PyrE
D5_0997 orotidine 5'-phosphate decarboxylase PyrF
D5_2008 carbamoyl-phosphate synthase large subunit CarB
D5_2009 carbamoyl-phosphate synthase small subunit CarA

Pyrimidine interconversion

D5_0708 cytidylate kinase Cmk
D5_1176 CMP/dCMP deaminase
D5_1497 thymidylate kinase Tmk
D5_1566 CTP synthase PyrG
D5_1617 thymidylate synthase ThyA
D5_1902 deoxycytidine triphosphate deaminase Dcd
D5_1916 uridylate kinase PyrH
D5_2038 thymidylate kinase Tmk
D5_2291 dUTP diphosphatase Dut
D5_2402 CMP/dCMP deaminase

Interconversion

D5_0065 5'-nucleotidase SurE
D5_1008 inosine-5'-monophosphate dehydrogenase GuaB
D5_1333 5'-nucleotidase SurE
D5_1653 nucleoside diphosphate kinase Ndk
D5_1808 GMP synthase subunit A GuaA
D5_1810 GMP synthase subunit B GuaAb
D5_2405 anaerobic ribonucleoside-triphosphate reductase NrdD

Salvage

D5_1762 uracil phosphoribosyltransferase Upp

Transport

D5_1315 xanthine/uracil permease
D5_1763 xanthine/uracil permease

REGULATION

Protein interaction

D5_0025 serine/threonine protein kinase related protein
D5_0357 serine/threonine protein kinase RIO1 family
D5_0753 TPR repeat-containing protein
D5_0904 WD40 repeat-containing protein
D5_1022 TPR repeat-containing protein
D5_1123 TPR repeat-containing protein
D5_1130 Hsp70 family protein with protein kinase domain
D5_1153 TPR repeat-containing protein
D5_1154 WD40 repeat-containing protein
D5_1240 TPR domain-containing protein
D5_1268 TPR repeat-containing protein
D5_1554 WD40 repeat-containing protein
D5_1664 TPR repeat-containing protein
D5_1884 TPR repeat-containing protein
D5_1992 anti-sigma factor antagonist
D5_1993 anti-sigma regulatory factor serine/threonine protein kinase
D5_1994 serine phosphatase
D5_1995 4'-phosphopantetheinyl transferase family protein
D5_2079 TPR repeat-containing protein
D5_2129 phosphate uptake regulator PhoU
D5_2252 TPR repeat-containing protein
D5_2444 TPR repeat-containing protein

Transcriptional regulator

D5_0022 transcriptional regulator PadR-like family
D5_0049 transcriptional regulator MarR family
D5_0113 transcriptional regulator
D5_0114 transcriptional regulator
D5_0129 transcriptional regulator HxlR family
D5_0154 HTH domain-containing protein
D5_0193 transcriptional regulator ArsR family
D5_0271 transcriptional regulator AsnC family
D5_0308 nucleotidyl transferase
D5_0352 transcriptional regulator
D5_0360 transcriptional regulator
D5_0380 transcriptional regulator
D5_0386 transcriptional regulator HxlR family
D5_0434 transcriptional regulator
D5_0552 transcriptional regulator AraC family
D5_0563 transcriptional regulator ArsR family
D5_0622 transcriptional regulator ArsR family
D5_0623 transcriptional regulator
D5_0624 transcriptional regulator ArsR family
D5_0748 transcriptional regulator TrmB family
D5_0796 transcriptional regulator HxlR family
D5_0814 transcriptional regulator MarR family
D5_0828 rOK family protein
D5_0942 transcriptional regulator HxlR family
D5_0993 transcriptional regulator
D5_1051 transcriptional regulator MarR family
D5_1055 transcriptional regulator AraC family
D5_1056 transcriptional regulator HxlR family
D5_1061 transcriptional regulator HxlR family
D5_1140 transcriptional regulator HxlR family
D5_1141 transcriptional regulator TetR family
D5_1155 transcriptional regulator
D5_1178 transcription regulator
D5_1184 transcriptional regulator MarR family
D5_1267 transcriptional regulator MarR family
D5_1269 transcription regulator
D5_1310 transcriptional regulator MarR family
D5_1342 transcriptional regulator MarR family
D5_1353 transcriptional regulator MarR family
D5_1361 transcriptional regulator
D5_1392 transcriptional regulator MarR family
D5_1510 transcriptional regulator
D5_1585 transcriptional regulator
D5_1661 iron-dependent repressor
D5_1690 transcription regulator
D5_1753 transcriptional regulator LytR family
D5_1791 HTH and cupin domain-containing protein
D5_1803 transcriptional regulator
D5_1804 transcriptional regulator
D5_1805 transcriptional regulator MarR family
D5_1818 HTH domain-containing protein
D5_1827 nickel responsive transcriptional regulator NikR
D5_1840 transcriptional regulator HxlR family
D5_1842 transcriptional regulator MarR family
D5_1904 transcriptional regulator
D5_1960 HTH domain-containing protein
D5_2191 transcriptional regulator MarR family
D5_2320 transcriptional regulator HxlR family
D5_2396 transcriptional regulator LysR family
D5_2440 transcriptional regulator ArsR family

Other

D5_0214 sugar fermentation stimulation protein SfsA
D5_1142 sugar fermentation stimulation protein SfsA

SECONDARY METABOLITES

NRPS

D5_0482 non-ribosomal surfactin synthase
D5_1740 non-ribosomal peptide synthetase

Other

D5_1741 anti-sigma factor antagonist
D5_1739 MatE efflux family protein

TRANSCRIPTION

RNA polymerase

D5_0409 DNA-directed RNA polymerase subunit H RpoH
D5_0410 DNA-directed RNA polymerase subunit B" RpoB2
D5_0411 DNA-directed RNA polymerase subunit B' RpoB1
D5_0412 DNA-directed RNA polymerase subunit A' RpoA1
D5_0413 DNA-directed RNA polymerase subunit A" RpoA2
D5_0722 DNA-directed RNA polymerase subunit D RpoD
D5_0726 DNA-directed RNA polymerase subunit N RpoN
D5_0727 DNA-directed RNA polymerase subunit K RpoK
D5_1647 DNA-directed RNA polymerase subunit E" RpoE2
D5_1648 DNA-directed RNA polymerase subunit E' RpoE1
D5_2411 DNA-directed RNA polymerase subunit F RpoF
D5_2433 DNA-directed RNA polymerase subunit L RpoL

Transcription factors

D5_0415 transcription elongation factor NusA-like protein
D5_1088 transcription factor S Tfs

D5_1931 transcription initiation factor TFIIB Tfb
D5_2109 transcription initiation factor TFIIB Tfb
D5_2230 transcription initiation factor TFIIE alpha subunit Tfe
D5_2364 TATA-box binding protein Tbp
D5_2436 transcription factor S Tfs

Other

D5_0063 LSM domain-containing protein
D5_0668 RNA-binding protein
D5_1589 ribonuclease III Rnc
D5_1594 LSM domain-containing protein
D5_2487 RNA-binding S1 domain-containing protein

TRANSPORTERS

Amino acids

D5_1090 amino acid carrier protein AGCS family
D5_1586 branched-chain amino acid transport protein AzlD
D5_1587 branched-chain amino acid transport protein AzlC

Anions

D5_0126 anion-transporting ATPase
D5_1116 voltage-gated chloride channel protein

Cations

D5_0151 heavy metal translocating P-type ATPase
D5_0207 potassium uptake protein TrkA family
D5_0208 potassium uptake protein TrkH family
D5_0212 potassium channel protein
D5_0276 K^+-dependent Na^+/Ca^+ exchanger
D5_0341 copper ion binding protein
D5_0342 heavy metal translocating P-type ATPase
D5_0422 cation-transporting P-type ATPase
D5_0516 transporter Na^+/H^+ antiporter family
D5_0564 heavy metal-translocating P-type ATPase
D5_0567 nickel ABC transporter substrate-binding protein NikA
D5_0568 nickel ABC transporter permease protein NikB
D5_0569 nickel ABC transporter permease protein NikC
D5_0570 nickel ABC transporter ATP-binding protein NikD
D5_0571 nickel ABC transporter ATP-binding protein NikE
D5_0573 nickel ABC transporter ATP-binding protein NikE
D5_0574 nickel ABC transporter ATP-binding protein NikD
D5_0577 nickel ABC transporter substrate-binding protein NikA
D5_0578 nickel ABC transporter permease protein NikB
D5_0579 nickel ABC transporter permease protein NikC
D5_0657 transporter CDF family
D5_0990 heavy metal translocating P-type ATPase
D5_1018 heavy metal efflux pump CzcD family
D5_1195 ion transport protein
D5_1201 ion transport protein (pseudo)
D5_1279 transporter CDF family
D5_1328 SPFH domain, Band 7 family protein
D5_1397 transporter Na^+/H^+ antiporter family (pseudo)
D5_1728 heavy metal translocating P-type ATPase
D5_1743 transporter Na^+/H^+ antiporter family
D5_1799 ferrous iron transport protein B FeoB
D5_1980 potassium uptake protein TrkH family
D5_2147 ferrous iron transport protein A FeoA
D5_2148 ferrous iron transport protein B FeoB
D5_2313 potassium channel protein (pseudo)
D5_2503 divalent cation transporter mgtE family
D5_2504 TrkA domain-containing protein

Other

D5_0023 ABC transporter ATP-binding protein
D5_0024 ABC transporter permease protein
D5_0050 MATE efflux family protein, matE
D5_0137 transporter
D5_0139 MatE efflux family protein
D5_0224 permease
D5_0243 Na^+dependent transporter SNF family
D5_0319 cell shape determining protein, MreB/Mrl family
D5_0321 transporter MotA/TolQ/ExbB proton channel family
D5_0322 transporter ExbD/TolR family
D5_0351 Na^+dependent transporter SBF family
D5_0385 MFS transporter
D5_0537 MFS transporter
D5_0553 ABC transporter ATP-binding/permease protein
D5_0554 ABC transporter ATP-binding/permease protein
D5_0555 ABC transporter ATP-binding protein
D5_0556 ABC transporter permease protein
D5_0557 ABC transporter permease protein
D5_0559 ABC transporter permease protein
D5_0572 MatE efflux family protein
D5_0779 ABC transporter substrate-binding protein
D5_0793 MatE efflux family protein
D5_0856 Na^+-dependent transporter SNF family
D5_0935 MFS transporter
D5_0937 ABC transporter ATP-binding/permease protein
D5_0934 MFS transporter
D5_0941 transporter SDF family
D5_0961 transporter MIP family
D5_0980 ABC transporter permease protein
D5_1029 ABC transporter ATP-binding protein
D5_1045 MFS transporter
D5_1050 MFS transporter
D5_1085 MotA/TolQ/ExbB proton channel family protein
D5_1097 MatE efflux family protein
D5_1100 MatE efflux family protein
D5_1145 MatE efflux family protein
D5_1149 MotA/TolQ/ExbB proton channel family protein
D5_1294 amino acid ABC transporter substrate-binding protein
D5_1295 amino acid ABC transporter permease protein
D5_1296 amino acid ABC transporter ATP-binding protein
D5_1311 MatE efflux family protein
D5_1354 MFS transporter
D5_1355 MFS transporter
D5_1371 transporter SDF family
D5_1570 MatE efflux family protein
D5_1576 MatE efflux family protein
D5_1580 ABC transporter permease protein
D5_1581 ABC transporter ATP-binding protein
D5_1613 ABC transporter ATP-binding/permease protein
D5_1614 ABC transporter ATP-binding/permease protein
D5_1759 MFS transporter
D5_1914 transporter MIP family
D5_2014 mechanosensitive ion channel protein
D5_2319 MFS transporter
D5_2333 Na^+-dependent transporter SNF family
D5_2409 MatE efflux family protein
D5_2422 MatE efflux family protein
D5_2423 MFS transporter
D5_2466 MFS transporter
D5_2468 transporter

UNKNOWN FUNCTION

Methanogenesis marker proteins

D5_0062 methanogenesis marker protein 12
D5_0142 methanogenesis marker protein 13
D5_0150 methanogenesis marker protein 16
D5_0171 methanogenesis marker protein 5
D5_0282 methanogenesis marker protein 7
D5_0283 methanogenesis marker protein 10
D5_0297 methanogenesis marker protein 14
D5_0332 methanogenesis marker protein 9
D5_0627 methanogenesis marker protein 2
D5_0631 methanogenesis marker protein 3
D5_0632 methanogenesis marker protein 6
D5_0633 methanogenesis marker protein 5
D5_0634 methanogenesis marker protein 15
D5_0635 methanogenesis marker protein 17
D5_1505 methanogenesis marker protein 16
D5_1511 methanogenesis marker protein 4
D5_2000 methanogenesis marker protein 1
D5_2244 methanogenesis marker protein 8
D5_2489 methanogenesis marker protein 11

Others

D5_0006 amidohydrolase
D5_0017 hydrolase HAD superfamily
D5_0021 xylose isomerase-like TIM barrel domain-containing protein
D5_0028 PHP domain-containing protein
D5_0032 ATPase AAA
D5_0033 radical SAM domain-containing protein
D5_0041 acetyltransferase
D5_0090 aconitase

D5_0093 pyridoxamine-phosphate oxidase (pseudo)
D5_0097 xylose isomerase-like TIM barrel domain-containing protein
D5_0098 PIN domain-containing protein
D5_0102 ATPase
D5_0119 phage antirepressor
D5_0130 macro domain-containing protein
D5_0133 calcineurin-like phosphoesterase
D5_0140 radical SAM domain-containing protein
D5_0141 C_GCAxxG_C_C family protein
D5_0149 amidohydrolase
D5_0155 CAAX amino terminal protease family protein
D5_0184 ainc dependent oxidoreductase
D5_0196 acetyltransferase GNAT family
D5_0202 CBS domain-containing protein
D5_0204 radical SAM domain-containing protein
D5_0206 metallo-beta-lactamase superfamily protein
D5_0213 carbohydrate kinase
D5_0248 CBS domain-containing protein
D5_0249 CBS domain-containing protein
D5_0267 Met-10 like-protein
D5_0273 aminotransferase DegT/DnrJ/EryC1/StrS family
D5_0275 PP-loop family protein
D5_0277 CBS domain-containing protein
D5_0279 TraB family protein
D5_0299 PRC-barrel domain-containing protein
D5_0335 ATPase
D5_0336 calcineurin-like phosphoesterase
D5_0340 hydrolase alpha/beta fold family
D5_0344 aldo/keto reductase family protein
D5_0345 aldo/keto reductase family protein
D5_0346 aldo/keto reductase family protein
D5_0373 NADH-dependent flavin oxidoreductase
D5_0377 acetyltransferase GNAT family
D5_0381 von Willebrand factor type A domain-containing protein
D5_0402 radical SAM domain-containing protein
D5_0424 maltose *O*-acetyltransferase
D5_0435 thioesterase family protein
D5_0439 cupin domain-containing protein
D5_0461 SAM-dependent methyltransferase
D5_0464 methyltransferase FkbM family
D5_0466 GtrA-like protein
D5_0470 acetyltransferase
D5_0471 SAM-dependent methyltransferase
D5_0479 SAM-dependent methyltransferase
D5_0499 acetyltransferase
D5_0507 phosphatase PAP2 family
D5_0518 oxidoreductase GFO/IDH/MOCA family
D5_0543 GTP-binding protein
D5_0544 AMMECR1 domain-containing protein
D5_0546 GMC oxidoreductase family protein
D5_0575 acetyltransferase GNAT family
D5_0580 acetyltransferase GNAT family
D5_0619 PRC-barrel domain-containing protein
D5_0637 phosphodiesterase MJ0936 family
D5_0638 manganese-dependent inorganic pyrophosphatase PpaC
D5_0643 Met-10 like-protein
D5_0653 calcineurin-like phosphoesterase
D5_0656 CBS domain-containing protein
D5_0659 radical SAM domain-containing protein
D5_0667 metallo-beta-lactamase superfamily protein
D5_0676 pyridoxal phosphate enzyme
D5_0718 LemA family protein
D5_0746 SAM-dependent methyltransferase
D5_0750 cupin 2 domain-containing protein
D5_0751 SAM-dependent methyltransferase
D5_0752 oxidoreductase aldo/keto reductase family
D5_0756 ATPase
D5_0767 acetyltransferase GNAT family
D5_0769 NADH-dependent flavin oxidoreductase
D5_0772 PP-loop family protein
D5_0776 GMC oxidoreductase family protein
D5_0778 SAM-dependent methyltransferase
D5_0786 metallo-beta-lactamase superfamily protein
D5_0792 DGC domain-containing protein
D5_0794 radical SAM domain-containing protein
D5_0795 radical SAM domain-containing protein
D5_0802 hydrolase HAD superfamily
D5_0812 radical SAM domain-containing protein
D5_0829 Fic family protein
D5_0833 ATPase
D5_0843 oxidoreductase domain-containing protein
D5_0844 oxidoreductase GFO/IDH/MOCA family
D5_0845 cytidyltransferase-related domain-containing protein
D5_0847 aminotransferase DegT/DnrJ/EryC1/StrS family
D5_0851 SAM-dependent methyltransferase
D5_0868 SAM-dependent methyltransferase
D5_0871 SAM-dependent methyltransferase
D5_0875 ATPase
D5_0877 von Willebrand factor type A domain-containing protein
D5_0929 ATPase
D5_0930 ATPase
D5_0931 acetyltransferase GNAT family (pseudo)
D5_0943 oxidoreductase aldo/keto reductase family
D5_0944 SAM-dependent methyltransferase
D5_0945 oxidoreductase aldo/keto reductase family
D5_0951 acetyltransferase GNAT family
D5_0987 oxidoreductase aldo/keto reductase family
D5_0989 radical SAM domain-containing protein
D5_1015 CAAX amino terminal protease family protein
D5_1019 carbohydrate kinase PfkB family
D5_1047 oxidoreductase aldo/keto reductase family
D5_1048 short-chain dehydrogenase family protein
D5_1049 short-chain dehydrogenase family protein
D5_1058 radical SAM domain-containing protein
D5_1078 ACT domain-containing protein
D5_1115 ATPase
D5_1129 FHA domain-containing protein
D5_1131 metallophosphoesterase
D5_1138 metallo-beta-lactamase superfamily protein
D5_1139 DSBA oxidoreductase
D5_1144 NADH-dependent flavin oxidoreductase
D5_1156 acyl-CoA synthetase
D5_1164 MazG domain-containing protein
D5_1172 thioesterase family protein
D5_1208 ATPase AAA family
D5_1209 MazG domain-containing protein
D5_1217 ATPase
D5_1265 ATPase AAA family
D5_1275 acetyltransferase GNAT family
D5_1299 nucleotidyltransferase
D5_1300 phosphoenolpyruvate phosphomutase AepX
D5_1301 phosphoenolpyruvate decarboxylase AepY
D5_1313 NADH-dependent flavin oxidoreductase
D5_1318 NADH-dependent flavin oxidoreductase
D5_1337 CAAX amino terminal protease family protein
D5_1339 flavin reductase domain-containing protein
D5_1340 metallo-beta-lactamase superfamily protein
D5_1343 ACT domain-containing protein
D5_1345 radical SAM domain-containing protein
D5_1348 C_GCAxxG_C_C family protein
D5_1357 pyridoxamine 5'-phosphate oxidase family protein
D5_1358 acetyltransferase GNAT family
D5_1360 acetyltransferase
D5_1367 toxic anion resistance protein
D5_1374 amidohydrolase
D5_1375 prophage antirepressor, partial
D5_1376 prophage antirepressor, partial
D5_1377 NUDIX domain-containing protein
D5_1383 hydrolase alpha/beta fold family
D5_1387 *O*-methyltransferase domain containing protein
D5_1389 SAM-dependent methyltransferase
D5_1409 peptide methionine sulfoxide reductase Msr
D5_1417 chloramphenicol *O*-acetyltransferase
D5_1470 SAM-dependent methyltransferase
D5_1479 PHP domain-containing protein
D5_1484 CAAX amino terminal protease family protein
D5_1485 CAAX amino terminal protease family protein
D5_1486 CAAX amino terminal protease family protein
D5_1513 hydrolase TatD family
D5_1529 oxidoreductase aldo/keto reductase family

D5_1534 MobA-related protein
D5_1540 ATPase
D5_1541 NurA domain-containing protein
D5_1557 HEAT repeat-containing protein
D5_1562 ATP-binding protein
D5_1568 acetylesterase
D5_1569 acetyltransferase GNAT family
D5_1569 acetyltransferase
D5_1602 acetyltransferase
D5_1616 acetyltransferase
D5_1641 ATPase
D5_1677 HD domain-containing protein
D5_1706 acetyltransferase
D5_1710 CAAX amino terminal protease family protein
D5_1725 DGC domain-containing protein
D5_1731 CAAX amino terminal protease family protein
D5_1766 CBS domain-containing protein
D5_1768 carbohydrate kinase PfkB family
D5_1789 CAAX amino terminal protease family protein
D5_1824 D-alanine--D-alanine ligase
D5_1836 radical SAM domain-containing protein
D5_1843 HEAT repeat-containing protein
D5_1845 AMP-binding enzyme
D5_1846 acetyltransferase GNAT family
D5_1859 SAM-dependent methyltransferase
D5_1860 SAM-dependent methyltransferase
D5_1866 calcineurin-like phosphoesterase
D5_1868 oxidoreductase aldo/keto reductase family
D5_1875 pyridoxamine 5'-phosphate oxidase family protein
D5_1878 short chain dehydrogenase
D5_1886 ACT domain-containing protein
D5_1891 YhgE/Pip-like protein
D5_1905 hydrolase TatD family
D5_1910 ZPR1 zinc-finger domain-containing protein
D5_1917 hydrolase TatD family
D5_1933 demethylmenaquinone methyltransferase
D5_1937 ATP-grasp domain-containing protein
D5_1986 acyltransferase
D5_1987 hydrolase alpha/beta fold family
D5_2001 TfuA-like protein
D5_2002 PP-loop family protein
D5_2004 CBS domain-containing protein
D5_2006 amidohydrolase
D5_2024 HEAT repeat-containing protein
D5_2025 short chain dehydrogenase
D5_2030 phosphodiesterase MJ0936 family
D5_2045 PAP2 family protein
D5_2048 metallophosphoesterase
D5_2051 ATPase AAA family
D5_2073 collagen triple helix repeat domain-containing protein
D5_2076 collagen triple helix repeat domain-containing protein
D5_2111 radical SAM domain-containing protein
D5_2113 cytidyltransferase-related domain-containing protein
D5_2117 PP-loop family protein
D5_2118 von Willebrand factor type A domain-containing protein
D5_2119 ATPase
D5_2158 collagen triple helix repeat domain-containing protein
D5_2195 SAM-dependent methyltransferase
D5_2205 DGC domain-containing protein
D5_2231 HD domain-containing protein
D5_2234 tryptophan-binding regulator TrpY
D5_2240 NADH-dependent flavin oxidoreductase
D5_2242 cell division control protein Cdc48
D5_2248 metallo-beta-lactamase superfamily protein
D5_2255 acetyltransferase
D5_2349 PIN domain-containing protein
D5_2352 GTP-binding protein
D5_2376 oxidoreductase aldo/keto reductase family
D5_2397 carbohydrate kinase PfkB family
D5_2398 thiamine biosynthesis protein ThiC
D5_2401 acetyltransferase GNAT family
D5_2413 acyltransferase
D5_2414 SAM-dependent methyltransferase HemK-related
D5_2418 NUDIX domain-containing protein
D5_2434 NUDIX domain-containing protein
D5_2441 hydrolase TatD family
D5_2443 HEAT repeat-containing protein
D5_2447 ATPase
D5_2449 ATPase
D5_2458 DSBA oxidoreductase
D5_2459 DSBA oxidoreductase

VITAMINS AND COFACTORS

Biotin

D5_2496 glycyl-radical enzyme activating protein
D5_0677 biotin-acetyl-CoA-carboxylase ligase BirA
D5_1332 biotin synthase BioB
D5_2133 biotin synthase BioB

Cobalamin

D5_0012 cobyrinic acid a,c-diamide synthase CbiA
D5_0044 precorrin-8X methylmutase CbiC
D5_0045 glutamate-1-semialdehyde-2,1-aminomutase HemL
D5_0068 cobyrinic acid a,c-diamide synthase CbiA
D5_0084 nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase CobT
D5_0128 precorrin-6Y C5,15-methyltransferase (decarboxylating) CbiE
D5_0187 cobaltochelatase CobN
D5_0298 GTP:adenosylcobinamide-phosphate guanylyltransferase CobY
D5_0333 siroheme synthase CysG
D5_0334 glutamyl-tRNA reductase HemA
D5_0384 alpha-ribazole phosphatase CobZ
D5_0389 cobalamin 5'-phosphate synthase CobS
D5_0421 precorrin-6X reductase CbiJ
D5_0519 porphobilinogen deaminase HemC
D5_0600 cobalamin biosynthesis protein CbiD
D5_0615 adenosylcobinamide amidohydrolase CbiZ
D5_0789 delta-aminolevulinic acid dehydratase HemB
D5_0819 uroporphyrinogen III synthase HemD
D5_0865 uroporphyrin-III *C*-methyltransferase CorA
D5_0946 cobyric acid synthase CbiP
D5_0998 cobalamin biosynthesis protein CbiM
D5_0999 cobalt transport protein CbiN
D5_1000 cobalt ABC transporter permease protein CbiQ
D5_1001 cobalt ABC transporter ATP-binding protein CbiO
D5_1082 cobalamin biosynthesis protein CbiM
D5_1146 cobaltochelatase CobN
D5_1519 precorrin-4 C11-methyltransferase CbiF
D5_1679 precorrin-6Y C5,15- methyltransferase (decarboxylating) CbiT
D5_2142 cobalamin biosynthesis protein CbiM
D5_2143 cobalt transport protein CbiN
D5_2144 cobalt ABC transporter permease protein CbiQ
D5_2145 cobalt ABC transporter ATP-binding protein CbiO
D5_2439 precorrin-2 C20-methyltransferase CbiL
D5_2498 cobalamin biosynthesis protein CbiX
D5_2499 cobalamin biosynthesis protein CbiX
D5_2505 precorrin-3B C17-methyltransferase CbiH
D5_2511 cobalamin biosynthesis protein CbiG
D5_2512 cobalamin biosynthesis protein CbiB

CoB

D5_0588 homoaconitase small subunit AksE
D5_1844 isohomocitrate dehydrogenase AksF
D5_2368 homocitrate synthase AksA
D5_2369 homoaconitase large subunit AksD

F_{420}

D5_0268 FO synthase subunit 1 CofG
D5_0324 F_{420}-*0*:gamma-glutamyl ligase CofE
D5_0325 LPPG:FO 2-phospho-L-lactate transferase CofD
D5_0503 F_{420}-*0*:gamma-glutamyl ligase CofE
D5_1079 F_{390} synthetase FtsA
D5_1498 FO synthase subunit 2 CofH
D5_1723 2-phospho-L-lactate guanylyltransferase CofC
D5_1885 F_{390} synthetase FtsA
D5_1974 lactaldehyde dehydrogenase CofA
D5_2229 F_{420}-*0*:gamma-glutamyl ligase CofE
D5_2508 fuculose 1-phosphate aldolase FucA

CoM

D5_0263 L-sulfolactate dehydrogenase ComC
D5_0280 2-phosphosulfolactate phosphatase ComB
D5_1500 sulfopyruvate decarboxylase subunit alpha ComD
D5_1501 sulfopyruvate decarboxylase subunit beta ComE

D5_1506 phosphosulfolactate synthase ComA

Glutathione

D5_0781 gamma-glutamylcysteine synthetase GshA
D5_1281 glutathione peroxidase GpxA
D5_1347 glutathione-disulfide reductase Gor
D5_1806 glutathione peroxidase GpxA
D5_2236 glutathione-disulfide reductase Gor
D5_2270 bifunctional glutamate-cysteine ligase/glutathione synthetase GshF

Metal-binding pterin

D5_0076 molybdate transport system regulatory protein ModE
D5_0359 molybdopterin biosynthesis protein MoeA
D5_0621 molybdenum cofactor biosynthesis protein B MoaB
D5_0774 molybdopterin biosynthesis protein MoeB
D5_1550 molybdenum cofactor biosynthesis protein MoaE
D5_1551 molybdenum-pterin binding protein Mop1
D5_1552 molybdenum-pterin binding protein Mop2
D5_1681 molybdopterin-guanine dinucleotide biosynthesis protein A MobA
D5_1871 molybdate ABC transporter ATP-binding protein ModC
D5_1872 molybdate ABC transporter ATP-binding protein ModC
D5_1873 molybdate ABC transporter permease protein ModB
D5_1874 molybdate ABC transporter substrate-binding protein ModA
D5_2377 molybdopterin biosynthesis protein MoeB
D5_2389 molybdopterin-guanine dinucleotide biosynthesis protein B MobB
D5_2390 molybdopterin cofactor biosynthesis protein A MobA
D5_2427 molybdenum cofactor biosynthesis protein C MoaC
D5_2455 molybdopterin biosynthesis protein MoeA

MF

D5_0313 L-tyrosine decarboxylase MfnA
D5_1007 2-furaldehyde phosphate synthase MfnB
D5_0584 tyramine—L-glutamate ligase MfnD
D5_2202 [5-(aminomethyl)furan-3-yl]methyl phosphate kinase MfnE
D5_0585 (4-{4-[2-(γ-L-glutamylamino)ethyl]phenoxymethyl}furan-2-yl)methanamine synthase MfnF

Methanopterin

D5_0270 GTP cyclohydrolase ArfA
D5_0314 dihydronepterin aldolase MtnD
D5_0587 7,8-dihydro-6-hydroxymethylpterin dimethyltransferase
D5_0790 ATP:dephospho-CoA triphosphoribosyl transferase CitG
D5_1075 beta-ribofuranosylaminobenzene 5'-phosphate synthase MptG
D5_1597 creatinine amidohydrolase ArfB
D5_2280 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase MptE
D5_2463 dihydromethanopterin reductase (acceptor) DmrX

Nicotinate

D5_0108 NAD synthetase NadE
D5_1381 ATP-NAD kinase NadK
D5_1549 nicotinamide-nucleotide adenylyltransferase NadD
D5_1621 nicotinate phosphoribosyltransferase
D5_1967 L-aspartate dehydrogenase AspDH
D5_2012 nicotinate-nucleotide pyrophosphorylase NadC
D5_2025 quinolinate synthetase A protein NadA

Others

D5_0300 FeS assembly ATPase SufC
D5_0301 FeS assembly protein SufBD
D5_0636 nitrogenase cofactor biosynthesis protein NifB
D5_2121 dihydropteroate synthase-related protein
D5_2272 dinitrogenase iron-molybdenum cofactor biosynthesis protein

Pantothenate and CoA

D5_0253 phosphopantothenoylcysteine decarboxylase CoaC
D5_0254 phosphopantothenate-cysteine ligase CoaB
D5_0661 phosphopantetheine adenylyltransferase CoaD
D5_1402 2-dehydropantoate 2-reductase PanE
D5_1471 pantothenate kinase CoaA
D5_1560 pantothenate synthase PanC
D5_1561 dephospho-CoA kinase CoaE

Riboflavin

D5_0008 riboflavin kinase RibK
D5_0010 3,4-dihydroxy-2-butanone 4-phosphate synthase RibB
D5_0326 GTP cyclohydrolase III ArfA
D5_1002 riboflavin synthase RibC
D5_1482 diaminohydroxyphosphoribosylaminopyrimidine reductase RibD
D5_2490 6,7-dimethyl-8-ribityllumazine synthase RibH

Thiamine

D5_0407 thiamine monophosphate synthase ThiE
D5_0408 hydroxyethylthiazole kinase ThiM
D5_0777 thiamine biosynthesis protein ThiS
D5_1334 hydroxymethylpyrimidine transporter CytX
D5_1335 phosphomethylpyrimidine kinase ThiD
D5_1724 phosphomethylpyrimidine kinase ThiD
D5_2116 thiamine biosynthesis protein ThiS
D5_2219 thiamine biosynthesis ATP pyrophosphatase ThiI
D5_2245 thiamine biosynthesis protein ThiC
D5_2398 thiamine biosynthesis protein ThiC
D5_2497 thiamine-monophosphate kinase ThiL
D5_1327 thiazole biosynthesis enzyme
D5_2137 ThiF family protein
D5_2442 ThiF family protein

Ubiquinone

D5_1678 3-polyprenyl-4-hydroxybenzoate decarboxylase UbiX
D5_1893 2-polyprenylphenol 6- hydroxylase UbiB
D5_1897 2-polyprenylphenol 6- hydroxylase UbiB2
D5_2042 2-polyprenylphenol 6- hydroxylase UbiB
D5_2485 UbiD family decarboxylase

Table A.6.2. Percentage codon usage in *Methanobrevibacter* genomes

Amino acid	Codon	D5	YE315	ZA-10^{T}	SM9	HO^{T}	PS^{T}	JMR01	M1^{T}	KM1H5-1P^{T}	YLM1	SH^{T}	AbM4	JH1^{T}	RFM-1^{T}	RFM-2^{T}	RFM-3^{T}	ANOR1	ATM^{T}
Ser (S)	UCA	3.6 [34.1]	2.0 [32.4]	3.7 [35.4]	3.4 [34.2]	2.1 [31.8]	2 [28.5]	2.2 [28.9]	2.7 [25.3]	1.7 [23.1]	1.7 [22.7]	1.9 [28.2]	1.9 [29.6]	2.0 [29.3]	1.7 [23.7]	1.9 [25.5]	1.8 [24.8]	1.9 [24.7]	1.5 [23.5]
	UCC	1.4 [13]	1.0 [15.4]	1.9 [17.7]	1.2 [11.8]	0.6 [8.6]	0.7 [10.2]	0.4 [4.8]	1.7 [16.1]	0.4 [5.2]	0.4 [5.1]	0.2 [2.7]	0.8 [13.9]	0.9 [13.8]	0.4 [5.4]	0.4 [5]	0.3 [4.7]	0.3 [3.6]	0.4 [5.5]
	UCG	0.2 [1.4]	0.1 [1.6]	0.3 [2.4]	0.1 [1.5]	0.1 [1.8]	0 [0.7]	0.1 [1.9]	0 [0.6]	0 [0.6]	0 [0.6]	0 [0.2]	0.04 [0.6]	0 [0.5]	0.1 [1.4]	0.1 [1.1]	0.1 [1.2]	0.1 [1.4]	0 [0.3]
	UCU	2.3 [21.7]	1.3 [20.6]	1.5 [14.4]	2.4 [24]	1.9 [28.9]	2.0 [28.7]	2.6 [33.9]	2.5 [24]	2.8 [38.9]	2.9 [38.6]	2.5 [37.9]	1.7 [25.6]	1.8 [25.9]	2.5 [34.9]	2.5 [34.7]	2.3 [30.8]	3.0 [39.7]	2 [31.6]
	AGC	1.1 [10.2]	0.8 [12.6]	1.8 [17.3]	0.9 [9.5]	0.4 [6.0]	0.6 [9]	0.3 [4.6]	2.2 [20.7]	0.5 [6.6]	0.5 [6.6]	0.1 [2.1]	0.4 [5.5]	0.4 [5.5]	0.5 [6.4]	0.5 [6.8]	0.4 [5.8]	0.4 [4.9]	0.2 [3.6]
	AGU	2.1 [19.6]	1.1 [17.5]	1.4 [12.8]	1.9 [18.9]	1.5 [22.9]	1.6 [22.9]	2.0 [25.9]	1.4 [13.3]	1.9 [25.6]	2.0 [26.5]	1.9 [28.9]	1.6 [24.9]	1.7 [25.1]	2 [28.1]	2 [26.9]	2.4 [32.6]	1.9 [25.8]	2.3 [35.5]
Phe (F)	UUC	2.7 [37.4]	1.8 [40.6]	3.3 [46]	2.2 [32.2]	1.2 [26.8]	1.2 [25.2]	0.8 [16.4]	2.7 [40.3]	1 [21.6]	1.0 [21.9]	0.8 [18.2]	1.2 [28.3]	1.2 [27.7]	1 [24.6]	0.7 [14.8]	1 [22.8]	0.8 [17]	1.1 [26.7]
	UUU	4.5 [62.6]	2.6 [59.4]	3.9 [54]	4.6 [67.8]	3.2 [73.2]	3.5 [74.8]	4.2 [83.6]	3.9 [59.7]	3.6 [78.4]	3.7 [78.1]	3.5 [81.8]	2.96 [71.7]	3.1 [72.3]	3.2 [75.4]	3.8 [85.2]	3.2 [77.2]	3.9 [83]	3 [73.3]
Thr (T)	ACA	3.8 [40.9]	2.1 [38.9]	2.7 [30.4]	3.3 [39.5]	2.0 [38.7]	2.2 [38.1]	2.3 [40.8]	3.0 [37.0]	1.6 [31.5]	1.7 [31.5]	2.0 [38.8]	1.9 [38.1]	2.1 [37.9]	2.1 [35.7]	2.1 [38.1]	2.4 [38.7]	1.9 [35.8]	2.2 [40.6]
	ACC	1.6 [17.1]	1.1 [20.3]	2.4 [27]	1.1 [13.9]	0.5 [10.2]	0.7 [12.6]	0.4 [7.5]	1.8 [22.9]	0.5 [9.3]	0.5 [9.4]	0.2 [4.6]	1.1 [21.1]	1.1 [20.6]	0.5 [9.3]	0.4 [8.1]	0.5 [7.5]	0.3 [6]	0.4 [8.1]
	ACG	0.2 [1.7]	0.1 [1.8]	1.1 [11.9]	0.3 [3.7]	0.1 [1.8]	0.1 [1.3]	0.1 [2.4]	0 [0.8]	0 [0.9]	0.1 [1]	0 [0.6]	0.1 [2.6]	0.1 [2.6]	0.1 [2.1]	0.1 [1.3]	0.1 [1.9]	0.1 [1.7]	0 [0.8]
	ACU	3.8 [40.3]	2.1 [39.0]	2.7 [30.7]	3.5 [42.9]	2.5 [49.3]	2.8 [48]	2.8 [49.3]	3.0 [38.4]	3 [58.3]	3.1 [58.1]	2.9 [56]	2.0 [38.2]	2.1 [38.9]	3.1 [52.9]	2.9 [52.4]	3.2 [51.9]	3 [56.6]	2.7 [50.4]
Asn (N)	AAC	3.6 [32.6]	2.0 [32.6]	4.7 [41.6]	2.8 [26.7]	1.3 [20.1]	1.8 [23.9]	1.3 [15.9]	3.3 [33.4]	1.1 [16.5]	1.2 [16.6]	0.9 [11.4]	1.6 [24.5]	1.7 [23.8]	1.7 [20.4]	1.7 [18.9]	1.7 [18.2]	1.2 [14.2]	1.2 [16.8]
	AAU	7.6 [67.4]	4.2 [67.4]	6.5 [58.4]	7.7 [73.3]	5.2 [79.9]	5.7 [76.1]	7 [84.1]	6.7 [66.6]	5.8 [83.5]	6.0 [83.4]	6.7 [88.6]	5.1 [75.5]	5.5 [76.2]	6.8 [79.6]	7.1 [81.1]	7.7 [81.8]	7 [85.8]	5.9 [83.2]
Lys (K)	AAA	9.8 [79.6]	5.4 [77.4]	7.4 [66.7]	9 [80.8]	7.0 [88.4]	7.7 [88.1]	8.5 [88.8]	6.0 [54.7]	7.4 [84.6]	7.6 [84]	8.2 [92.7]	6.1 [77]	6.4 [77]	7.1 [84.8]	7.8 [86.6]	6.9 [84.7]	7.8 [84.6]	7.5 [89.4]
	AAG	2.5 [20.4]	1.6 [22.6]	3.7 [33.3]	2.1 [19.2]	0.9 [11.6]	1.0 [11.9]	1.1 [11.2]	5.0 [45.3]	1.4 [15.4]	1.5 [16]	0.6 [7.3]	1.8 [23]	1.9 [23]	1.3 [15.2]	1.2 [13.4]	1.3 [15.3]	1.4 [15.4]	0.9 [10.6]
Glu (E)	GAA	9.6 [82.8]	6.1 [83.2]	9.3 [79.9]	8.7 [85.4]	6.2 [86.9]	6.8 [88.3]	6.6 [87]	8.1 [72]	6.9 [85]	7.1 [84.8]	6.3 [90.8]	5.1 [71.9]	5.3 [71.9]	5.6 [85.2]	5.5 [84.5]	5.3 [84.3]	6.4 [85.2]	6.4 [90.8]
	GAG	2.0 [17.2]	1.2 [16.8]	2.3 [20.1]	1.5 [14.6]	0.9 [13.1]	0.9 [11.7]	1 [13]	3.2 [28]	1.2 [15]	1.3 [15.6]	0.6 [9.2]	2.0 [28.1]	2.1 [28.1]	1 [14.8]	1 [15.5]	1 [15.7]	1.1 [14.8]	0.6 [9.2]
Tyr (Y)	UAC	2.0 [30.1]	1.2 [29.2]	2.6 [38]	1.5 [23.4]	0.8 [21.6]	1.1 [24.9]	0.8 [17.1]	1.7 [26.6]	0.8 [18.5]	0.8 [18.4]	0.5 [12.1]	0.8 [18.3]	0.8 [18]	0.7 [18.5]	0.9 [22.7]	0.9 [22.7]	0.4 [12]	0.8 [20.3]
	UAU	4.7 [69.9]	2.8 [70.8]	4.3 [62]	4.9 [76.6]	3.1 [78.4]	3.3 [75.1]	3.8 [82.9]	4.6 [73.4]	3.4 [81.5]	3.7 [81.6]	3.8 [87.9]	3.4 [81.7]	3.6 [82]	3.1 [81.5]	3.1 [77.3]	3 [77.3]	3.5 [88]	3.1 [79.7]
Val (V)	GUA	3.4 [30.1]	2.0 [28.8]	2.9 [27.3]	3 [30.3]	2.0 [31.3]	2.6 [35]	2.3 [32.4]	3.1 [32.9]	2.6 [40.4]	2.7 [40]	2.6 [41.8]	2.5 [38.1]	2.6 [37.8]	1.8 [29.6]	1.6 [25.8]	1.8 [27.5]	1.9 [28.9]	2.3 [37.6]
	GUC	1.3 [12.1]	0.9 [12.9]	1.9 [18]	1.1 [10.7]	0.4 [6.5]	0.5 [6.1]	0.3 [4.4]	1.5 [16.2]	0.3 [4.1]	0.3 [4.2]	0.2 [2.8]	0.5 [7]	0.5 [6.9]	0.4 [6.7]	0.4 [6.4]	0.3 [4.7]	0.3 [4.5]	0.2 [3.7]

	GUG	1.3 [11.2]	0.8 [10.9]	1.2 [10.8]	0.9 [9.5]	0.6 [9.5]	0.5 [7]	0.5 [6.8]	1.1 [11.9]	0.3 [5]	0.3 [5.1]	0.1 [2.3]	0.5 [7.5]	0.5 [7.7]	0.6 [9.7]	0.5 [8.7]	0.8 [12.5]	0.4 [6.4]	0.5 [7.7]
	GUU	5.2 [46.7]	3.3 [47.3]	4.7 [43.8]	4.9 [49.5]	3.5 [53.6]	3.9 [51.9]	4.0 [56.4]	3.7 [39.1]	3.3 [50.5]	3.4 [50.7]	3.4 [53.1]	3.1 [47.4]	3.2 [47.7]	3.3 [54]	3.6 [59.1]	3.6 [55.2]	3.9 [60.2]	3.2 [51.1]
Gln (Q)	CAA	2.5 [67.4]	1.5 [66.9]	1.6 [45.2]	2.2 [63.9]	1.8 [77.1]	1.6 [61.3]	2.2 [88.4]	2.2 [67.2]	1.9 [86.5]	2.0 [85.6]	1.9 [90.9]	1.5 [67.5]	1.6 [67.4]	1.6 [82.1]	1.7 [84.2]	1.6 [81.4]	1.8 [82.4]	2.3 [91.9]
	CAG	1.2 [32.6]	0.7 [33.1]	1.9 [54.8]	1.2 [36.1]	0.5 [22.9]	1.0 [38.7]	0.3 [11.6]	1.1 [32.8]	0.3 [13.5]	0.3 [14.4]	0.2 [9.1]	0.7 [32.5]	0.8 [32.6]	0.4 [17.9]	0.3 [15.8]	0.4 [18.6]	0.3 [17.6]	0.2 [8.1]
Met (M)	AUG	3.6 [100.0]	2.3 [100.0]	3.6 [100.0]	3.3 [100.0]	2.4 [100.0]	2.5 [100.0]	2.5 [100.0]	3.2 [100.0]	2.4 [100.0]	2.4 [100.0]	2.3 [100.0]	2.3 [100.0]	2.4 [100.0]	2.1 [100]	1.8 [100]	1.8 [100]	2.5 [100.0]	2.3 [100]
Cys (C)	UGC	0.6 [31.8]	0.5 [36.3]	0.9 [46.7]	0.5 [28.2]	0.3 [22.6]	0.3 [24.3]	0.2 [14]	3.7 [35.2]	0.2 [13.5]	0.2 [13.5]	0.1 [5.1]	0.2 [13.1]	0.2 [12.1]	0.1 [14.3]	0.1 [15.6]	0.1 [16.5]	0.1 [10.4]	0.1 [11.7]
	UGU	1.3 [68.2]	0.8 [63.7]	1.1 [53.3]	1.3 [71.8]	1.0 [77.4]	1.1 [75.7]	1.2 [86]	6.8 [64.8]	1.1 [86.5]	1.1 [86.5]	1.2 [94.9]	1.1 [86.9]	1.1 [87.9]	0.8 [85.7]	0.7 [84.4]	0.7 [83.5]	1 [89.6]	1.1 [88.3]
Leu (L)	CUA	0.9 [7]	0.6 [6.8]	0.6 [4.8]	0.8 [6.1]	0.5 [5.7]	0.4 [4.7]	0.7 [7.1]	1.3 [10.1]	0.7 [7.3]	0.7 [7.3]	0.3 [3.9]	0.8 [9.3]	0.8 [9]	0.9 [10.8]	0.7 [8.2]	0.8 [9.5]	0.7 [7.6]	0.6 [6.3]
	CUC	0.6 [4.8]	0.4 [4.4]	0.9 [6.8]	0.4 [3.4]	0.3 [3.0]	0.2 [2]	0.2 [1.8]	0.8 [6]	0.2 [2.1]	0.2 [2.1]	0.1 [1.4]	0.2 [2.4]	0.2 [2.4]	0.3 [3.6]	0.2 [2]	0.2 [2.9]	0.2 [2.4]	0.2 [2.1]
	CUG	1.1 [7.8]	0.7 [8.2]	2.1 [15.4]	0.9 [6.8]	0.3 [3.0]	0.6 [6.7]	0.1 [1.1]	0.6 [4.9]	0.1 [1.1]	0.1 [1.1]	0 [0.3]	0.2 [2.5]	0.2 [2.6]	0.2 [2.2]	0.1 [1.3]	0.2 [2]	0.1 [1.5]	0 [0.5]
	CUU	2.7 [20.1]	1.7 [19.7]	2.9 [21.2]	2.3 [18.4]	1.7 [20.1]	1.8 [19.5]	1.9 [18.8]	3.4 [26.6]	2.2 [22.7]	2.2 [22.7]	2.4 [27.5]	2.1 [24.3]	2.2 [24.6]	1.7 [20.6]	1.7 [19.7]	1.9 [23]	2.3 [24.7]	1.4 [15.6]
	UUA	5.3 [38.9]	3.2 [37.7]	4.4 [32.6]	5.4 [42.9]	4.6 [53.9]	4.8 [51.2]	6.1 [61.9]	4.2 [32.3]	5.8 [60]	5.9 [59.6]	5.6 [63.3]	4.3 [50.0]	4.5 [49.7]	4.2 [50.7]	5.1 [57.7]	4.3 [52.1]	5 [53.5]	6.3 [69.9]
	UUG	2.9 [21.5]	2.0 [23.2]	2.6 [19.3]	2.8 [22.4]	1.2 [14.1]	1.5 [16]	0.9 [9.3]	2.6 [20.2]	0.7 [6.9]	0.7 [7.2]	0.3 [3.6]	1.0 [11.4]	1.1 [11.7]	1 [12]	1 [11.2]	0.9 [10.6]	1 [10.3]	0.5 [5.5]
Ala (A)	GCA	4.8 [49.7]	2.8 [45.2]	4.2 [43]	4.2 [48.6]	2.9 [49.4]	2.9 [43.6]	2.7 [42.6]	4.4 [48.2]	3 [46.5]	3.0 [46.2]	3 [55.1]	2.7 [50.2]	2.9 [50.2]	1.9 [35.1]	1.9 [38.8]	2.1 [40.7]	2.1 [35.9]	2.7 [50.2]
	GCC	0.8 [9]	0.8 [12.4]	2 [20.1]	0.7 [8.3]	0.3 [5.6]	0.4 [6.8]	0.3 [4.2]	1.1 [12.2]	0.2 [3.1]	0.2 [3.1]	0 [0.9]	0.4 [8.1]	0.5 [8]	0.3 [5.5]	0.4 [7.7]	0.2 [4]	0.2 [3.4]	0.2 [3.3]
	GCG	0.5 [5.5]	0.4 [6.6]	0.8 [7.8]	0.4 [4.6]	0.2 [4.2]	0.1 [2.2]	0.2 [2.6]	0.3 [3]	0.1 [1.5]	0.1 [1.6]	0 [0.9]	0.2 [4.2]	0.2 [4.2]	0.1 [2.3]	0.1 [2.4]	0.2 [3.6]	0.1 [2]	0.1 [2.2]
	GCU	3.5 [35.9]	2.2 [35.7]	2.8 [29.1]	3.3 [38.5]	2.4 [40.8]	3.1 [47.4]	3.2 [50.7]	3.3 [36.5]	3.2 [48.9]	3.2 [49.2]	2.4 [43.1]	2.1 [37.6]	2.1 [37.5]	3.1 [57.1]	2.5 [51]	2.6 [51.8]	3.4 [58.7]	2.4 [44.2]
Trp (W)	UGG	1.1 [100.0]	0.7 [100.0]	1.1 [100.0]	1.0 [100.0]	0.7 [100.0]	0.7 [100.0]	0.7 [100.0]	1.0 [100.0]	0.7 [100.0]	0.7 [100.0]	0.7 [100.0]	0.6 [100.0]	0.7 [100.0]	0.6 [100]	0.7 [100]	0.7 [100]	0.7 [100.0]	0.6 [100]
Pro (P)	CCA	1.8 [37.8]	1.1 [34.8]	1.2 [25.7]	1.7 [37.8]	1.4 [42.7]	1.2 [35.4]	1.6 [47.6]	1.8 [40.1]	1.4 [43.6]	1.5 [43.3]	1.6 [47.7]	1.3 [38.7]	1.3 [38.5]	1.2 [43.5]	1.3 [45.5]	1.3 [47.3]	1.4 [41.2]	1.5 [43.2]
	CCC	0.3 [5.5]	0.2 [5.7]	0.4 [9.4]	0.2 [5.4]	0.2 [5.9]	0.2 [4.5]	0.2 [4.5]	0.13 [3.1]	0.1 [3.9]	0.1 [3.8]	0 [1.3]	0.2 [6.1]	0.2 [6]	0.1 [5]	0.1 [3.6]	0.1 [4.1]	0.1 [3.4]	0.1 [4.2]
	CCG	0.6 [11.8]	0.5 [15.1]	1 [21.7]	0.5 [11.4]	0.3 [10.2]	0.4 [12.8]	0.1 [4.3]	5.8 [5.8]	0.1 [3.1]	0.1 [3.1]	0 [1.2]	0.2 [7.5]	0.3 [7.4]	0.1 [2.6]	0 [1.6]	0 [1.9]	0.1 [2.2]	0.1 [2.8]
	CCU	2.2 [45]	1.4 [44.4]	2.1 [43.3]	2.0 [45.4]	1.3 [41.2]	1.7 [47.3]	1.5 [43.7]	2.3 [51.0]	1.6 [49.4]	1.7 [49.7]	1.7 [49.8]	1.5 [47.7]	1.6 [48]	1.4 [48.9]	1.4 [49.4]	1.2 [46.8]	1.8 [53.2]	1.7 [49.8]
His (H)	CAC	0.8 [32]	0.5 [30.5]	1 [40.5]	0.6 [26.3]	0.4 [22.5]	0.4 [24.9]	0.3 [18.1]	0.6 [28.4]	0.3 [20]	0.3 [19.6]	0.2 [11.4]	0.3 [19.3]	0.3 [18.6]	0.2 [16.9]	0.3 [19.5]	0.3 [20.2]	0.2 [13.2]	0.3 [17.9]

	CAU	1.7 [68]	1.1 [69.5]	1.5 [59.5]	1.7 [73.7]	1.3 [77.5]	1.3 [75.1]	1.4 [81.9]	1.6 [71.6]	1.3 [80]	1.3 [80.4]	1.4 [88.6]	1.3 [80.7]	1.4 [81.4]	1.2 [83.1]	1.2 [80.5]	1.2 [79.8]	1.4 [86.8]	1.4 [82.1]
Asp (D)	GAC	3.2 [29.3]	1.9 [28.0]	4.3 [39.2]	2.3 [23.3]	1.0 [15.9]	1.6 [22.9]	0.8 [12.2]	3.7 [35.2]	0.9 [13]	0.9 [13]	0.5 [7.7]	1.3 [18.6]	1.3 [18.4]	0.7 [12.6]	0.9 [16.9]	0.8 [14.5]	0.6 [9.6]	0.7 [10.8]
	GAU	7.7 [70.7]	4.9 [72.0]	6.6 [60.8]	7.6 [76.7]	5.4 [84.1]	5.5 [77.1]	6.1 [87.8]	6.8 [64.8]	6 [87]	6.2 [87]	6 [92.3]	5.6 [81.4]	5.9 [81.6]	5 [87.4]	4.6 [83.1]	4.7 [85.5]	5.8 [90.4]	5.4 [89.2]
Arg (R)	CGA	0.2 [4.4]	0.1 [2.8]	0.2 [3.8]	0.2 [4.7]	0.1 [3.0]	0.1 [2.4]	0.2 [5]	0.2 [3.5]	0 [2]	0.1 [2.1]	0 [1.8]	0.0 [2.1]	0 [2.1]	0.2 [7.9]	0.2 [6.2]	0.2 [8.4]	0.2 [5]	0.3 [8.6]
	CGC	0.2 [3.6]	0.1 [4.5]	0.3 [6.1]	0.2 [3.6]	0.1 [2.2]	0.1 [2.4]	0 [1.8]	0.2 [3.9]	0 [1.7]	0.1 [1.6]	0 [0.5]	0.0 [1.1]	0 [1.2]	0 [1.6]	0.1 [2.1]	0.1 [2]	0 [1]	0 [1.2]
	CGG	0.1 [1.7]	0.0 [1.1]	0.1 [1.9]	0.1 [1.6]	0 [1.1]	0 [0.9]	0 [1]	0 [0.9]	0 [0.4]	0 [0.5]	0 [0.4]	0.0 [0.8]	0 [0.8]	0 [1.3]	0 [0.9]	0 [1.7]	0 [1]	0.1 [2.3]
	CGU	0.8 [15.6]	0.5 [14.8]	0.7 [13.6]	0.8 [16.6]	0.5 [16.5]	0.5 [15.3]	0.5 [15.8]	0.5 [10.9]	0.5 [13.4]	0.5 [13.8]	0.6 [18.5]	0.5 [15.4]	0.6 [15.9]	0.3 [11.2]	0.4 [13.2]	0.3 [12.2]	0.4 [11.8]	0.7 [24]
	AGA	2.7 [54.6]	1.7 [54.0]	2.3 [46.6]	2.6 [56.5]	2.0 [62.2]	2.0 [61.6]	2.1 [62.9]	2.7 [55.9]	2.4 [69]	2.4 [68.1]	2.1 [67.5]	2.0 [58.1]	2.0 [57.7]	1.8 [64.1]	1.8 [65.7]	1.7 [62.1]	2.1 [65.8]	1.7 [58.4]
	AGG	1.0 [20.1]	0.7 [22.8]	1.4 [28]	0.8 [16.9]	0.5 [15.0]	0.6 [17.4]	0.4 [13.4]	1.2 [24.8]	0.5 [13.5]	0.5 [13.8]	0.3 [11.2]	0.8 [22.5]	0.8 [22.4]	0.4 [14]	0.3 [12]	0.4 [13.6]	0.5 [15.4]	0.2 [5.6]
Ile (I)	AUA	4 [26.8]	2.4 [25.2]	3.9 [26.6]	4 [28.4]	2.3 [24.3]	3 [29.4]	3.2 [28.1]	4.3 [31.5]	2.9 [27.8]	3 [28.1]	3.0 [29.6]	2.8 [29.9]	2.9 [29.7]	4 [39.4]	3.4 [33.5]	3.8 [38.2]	4.2 [37.7]	3.7 [38.7]
	AUC	2.9 [19.5]	2.1 [22.3]	3.9 [26.5]	2.3 [16.2]	1.2 [12.5]	1.3 [12.6]	1 [8.5]	3.3 [24.1]	1.2 [11.4]	1.2 [11.5]	0.6 [6.2]	1.4 [14.9]	1.5 [14.8]	1.2 [11.5]	0.8 [7.6]	0.9 [9.2]	0.9 [7.8]	0.8 [8.6]
	AUU	8 [53.6]	4.9 [52.4]	7 [47]	7.8 [55.4]	6.1 [63.2]	6 [58]	7.3 [63.4]	6.0 [44.4]	6.3 [60.8]	6.4 [60.4]	6.5 [64.2]	5.2 [55.2]	5.4 [55.5]	5 [49]	5.9 [58.9]	5.2 [52.6]	1.4 [54.4]	5.1 [52.7]
Gly (G)	GGA	4.5 [42]	2.7 [40.4]	4.6 [43.1]	4.0 [41.7]	2.7 [42.1]	3.4 [46.3]	3.1 [44]	4.9 [49.4]	2.8 [41.5]	2.9 [40.9]	2.6 [40.8]	2.5 [38.9]	2.6 [38.8]	3.1 [45.7]	2.9 [45.6]	2.9 [43.5]	3.2 [46]	2.8 [44.4]
	GGC	1.2 [11.2]	1.0 [14.7]	2.3 [21.3]	1.0 [10.2]	0.5 [7.5]	0.6 [8.1]	0.4 [5.4]	1.7 [17.6]	0.5 [7.7]	0.5 [7.6]	0.1 [1.9]	0.4 [5.9]	0.4 [5.9]	0.4 [6.1]	0.4 [6.4]	0.3 [4.7]	0.3 [4.5]	0.2 [3]
	GGG	0.6 [5.9]	6.3 [0.4]	0.5 [5]	0.5 [5.1]	0.5 [7.5]	0.4 [5.7]	0.5 [7]	0.6 [6.5]	0.4 [6.1]	0.4 [6.2]	0.2 [3.1]	0.3 [5.1]	0.4 [5.2]	0.6 [8.7]	0.5 [7.1]	0.5 [8]	0.6 [8.5]	0.5 [7.9]
	GGU	4.4 [40.9]	2.6 [38.6]	3.2 [30.6]	4.1 [43.1]	2.8 [43.2]	2.9 [39.9]	3.1 [43.5]	2.6 [26.6]	3.1 [44.8]	3.2 [45.3]	3.5 [54.2]	3.2 [50.1]	3.4 [50.1]	2.7 [39.5]	2.6 [40.8]	2.9 [43.8]	2.9 [41]	2.9 [44.8]
Stop codons																			
ochre	UAA	59	58.1	57.5	62.1	62.1	64.9	64.4	61.5	74.1	73.9	74.9	68.2	68.4	64.9	72.9	66.8	68.2	86.9
amber	UAG	23.6	25.4	24	18.1	19.5	19.7	21.3	23.9	17.3	17	17.3	23.1	22.1	17.5	14.6	17.3	18.4	5.5
opal	UGA	17.4	16.5	18.5	19.8	18.4	15.4	14.3	14.6	8.6	9.1	7.8	8.7	9.5	17.7	12.5	15.9	13.4	7.6
Translation initiator																			
M	AUG	90.9	89.5	89.8	90.6	89.6	91.5	85.8	91.0	90.1	91.7	90.9	92.0	90.7	99.5	99.4	98.7	83.9	99.6
L	GUG	6.1	5	5.7	5.6	5.5	5.5	9.6	6.4	5.8	6.1	6.1	5.7	5.8	0.3	0.2	0.9	12.4	0
V	UUG	2.9	2.5	2.9	3.7	3.4	3.1	3.9	2.6	2.7	2.2	3	2.3	2.8	0.2	0.4	0.4	3.7	0

[#]the number outside square bracket indicates percentage of codon in genome, number in square bracket indicates percentage of codon used for each amino acid.

Table A.6.3. CRISPR associated genes in *Methanobrevibacter* spp.

gene	D5	YE315	ZA-10^T	SM9	HO^T	PS^T	JMR01	M1^T	KM1H5-1P^T	YLM1	SH^T	AbM4	JH1^T	RFM-1^T	RFM-2^T	RFM-3^T	ANOR1	ATM^T	Reference*
cas1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	SERP2463
cas2	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	SERP2462
cas3'	-	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+	-	+	APE1232
cas3''	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	APE1231
cas4	+	+	+	-	+	-	-	-	-	-	-	+	-	-	+	+	-	-	APE1239
cas5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	APE1234
cas6	+	-	+	-	-	+	+	-	+	+	-	+	+	+	+	-	-	+	PF1131
cas6e	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*ygcH*
cas6f	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	y1727
cas7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*ygcJ*
cas8a1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	LA3191
cas8a2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MJ0385
cas8b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MTH1090
cas8c	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	BH0338
cas9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Spy1046
cas10/csm1	-	-	-	+	-	-	+	+	+	+	+	-	-	-	+	+	-	-	MTH326
cas10d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Slr7011
csy1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y1724
csy2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y1725
csy3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y1726
cse1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*ygcL*
cse2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*ygcK*
csc1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	alr1563
csc2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	slr7012
csa5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MJ0380
csn2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	SPy1049
csm2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MTH1081
csm3	-	-	-	+	-	-	+	+	+	+	+	-	-	-	+	+	-	-	MTH1080
csm4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MTH1079
csm5	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	MTH1078
csm6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	APE2256
cmr1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PF1130
cmr3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PF1128
cmr4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PF1126
cmr5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MTH324
cmr6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PF1124
csb1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Balac_1306
csb2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*Balac_1305*
csb3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*Balac_1303*
csx17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Btus_2683
csx14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GSU0052
csx10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Caur_2274
csx16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	VVA1548
csaX	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	SSO1438
csx3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AF1864
csx1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	MJ1666

csx15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	TTE2665
csf1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AFE_1038
csf2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AFE_1039
csf3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AFE_1040
csf4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	AFE_1037

Prediction cutoff: E-04. +Represents the gene is predicted in the particular genome. –indicates the absence of gene in the particular genome. *Reference protein is displayed in either locus tag or gene name within the NCBI protein database

Table A.6.4. Domains of predicted secretome in the *Methanobrevibacter* genomes analysed

Pfam/TIGRfam	**D5**	**YE315**	**ZA-10^T**	**SM9**	**HO^T**	**PS^T**	**JMR01**	**M1^T**	**KM1H5-1P^T**	**YLM1**	**SH^T**	**AbM4**	**JH1^T**	**RFM-1^T**	**RFM-2^T**	**RFM-3^T**	**ATM^T**	**ANOR1**
Repeat domains (containing PF13306: Leucine rich repeats (6 copies), PF13754: Bacterial Ig-like domain (group 3) , PF02369: Bacterial Ig-like domain (group 1), PF09479: Listeria-Bacteroides repeat domain, PF05345: Putative Ig domain, PF02368: Bacterial Ig-like domain (group 2), PF09373: Pseudomurein-binding repeat, TIGR02167: bacterial surface protein 26-residue repeat)	27	19	22	26	5	11	7	37	9	10	3	2	3	12	3	9	2	6
LPxTG motif containing proteins																		
PF01345: DUF11	-	-	1	2	-	2	-	1	1	1	-	1	1	-	-	-	-	-
PF13229: Right handed beta helix region	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Lipo-box containing protein																		
PF00112: peptidase_C1	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
PF00288: GHMP kinases N terminal domain	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
PF00384: Molybdopterin oxidoreductase	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	
PF00497: Bacterial extracellular solute-binding proteins, family 3	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
PF00756: Esterase	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PF01058: NADH ubiquinone oxidoreductase, 20 Kd subunit	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-
PF01345: DUF11	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-		-	-
PF01497: Periplasmic binding protein	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	2	-	-
PF01841: Transglutaminase-like superfamily	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-
PF02655: ATP-grasp domain	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
PF02663: FmdE, Molybdenum formylmethanofuran dehydrogenase operon	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
PF03063: Prismane/CO dehydrogenase family	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
PF03275: UDP-galactopyranose mutase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
PF03577: Peptidase family C69	-	-	-	2	-	-	-	1	-	-	-	-	-	-	-	-	-	-
PF03412: Peptidase C39 family	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
PF05048: Periplasmic copper-binding protein (NosD)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
PF07705: CARDB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
PF09394: Chagasin family peptidase inhibitor I42	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PF09972: DUF2207	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-
PF11824: DUF3344	-	-	-	-	-	-	1	-	1	1	-	-	-	-	-	-	-	-
PF13197: DUF4013	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
PF12695: Alpha/beta hydrolase family	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

PF13190: PDGLE domain	-	-	-	-	1	-	-	-	-	-	1	-	-	-	1	-	1	-	
PF13200: Putative glycosyl hydrolase domain (DUF4015)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	
PF13229: Right handed beta helix region	1	-	1	1	1	-	-	1	-	-	1	1	1	1	1	1	-	-	
PF13290: Chitobiase/beta-hexosaminidase C-terminal domain	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF13360: PQQ-like domain	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13391: HNH endonuclease	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF13425: *O*-antigen ligase like membrane protein	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13459: 4Fe-4S single cluster domain	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	
PF13531: Bacterial extracellular solute-binding protein	-	2	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13620: Carboxypeptidase regulatory-like domain	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF14347: DUF4399	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
Integral membrane protein (3 or more TMH)																			
PF00112: peptidase_C1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF00534: Glycosyl transferase family 1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF00535: Glycosyl transferase family 2	1	1	1	1	1	-	-	1	1	1	-	-	-	1	-	1	1	-	
PF00892: EamA-like transporter family	-	-	1	-	-	-	-	-	1	1	1	1	1	-	-	-	-	-	
PF00893: Small Multidrug Resistance protein	-	1	-	-	-	-	-	2	-	-	-	-	-	-	1	-	-	-	
PF01345: DUF11	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF01569: PAP2 superfamily	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF01595: DUF21	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF01699: Sodium/calcium exchanger protein	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	
PF01891: Cobalt uptake substrate-specific transmembrane region CbiM	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	
PF02163: Peptidase family M50	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF02386: Cation transport protein	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF02516: Oligosaccharyl transferase STT3 subunit	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	
PF02554: Carbon starvation protein CstA	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF02588: DUF161	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF02659: Putative manganese efflux pump (DUF204)	1	1	1	1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	
PF02687: FtsX-like permease family	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF03186: CobD/Cbib protein	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF03553: Na^+/H^+ antiporter family	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF04018: DUF368	-	-	1	1	1	-	-	-	1	-	-	-	-	1	1	-	-	1	
PF04020: Mycobacterial 4 TMS phage holin, superfamily IV	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF04203: Sortase family	-	-	-	-	-	-	-	-	-	-	1	1	1	1	-	-	-	-	
PF04206: H_4MPT S-methyltransferase, subunit E	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
PF04207: H_4MPT S-methyltransferase, subunit D	-	-	-	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	
PF07690: Major Facilitator Superfamily	1	1	-	-	2	1	1	-	-	1	-	-	-	-	-	-	1	-	

PF09376: NurA domain	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF09972: DUF2207	1	1	1	1	1	1	1	3	1	1	1	1	1	-	1	-	1	1	
PF13197: DUF4013	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	
PF13229: Right handed beta helix	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	
PF13439: Glycosyltransferase Family 4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
Dual membrane anchored																			
PF00005: ABC transporter	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF00112: peptidase_C1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF00375: Sodium:dicarboxylate symporter family	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF00482: Type II secretion system (T2SS), protein F	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF00565: Staphylococcal nuclease homologue	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF01345: DUF11	5	-	1	5			-		1	3	1	-	1		1	3	-	3	
PF02514: CobN/Magnesium Chelatase	-	-	-	-	1	1	-	1	-	1	1	-	-	1	1		-	1	
PF02553: Cobalt transport protein component CbiN	-	-	-	-	1		-	-	-	-	1	-	-	-	-	1	-	-	
PF02663: FmdE, Molybdenum formylmethanofuran dehydrogenase operon	1	-	1	-	2	1	-	-	-	-	-	-	1	-	-	-	-	1	
PF02805: Metal binding domain of Ada	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF03412: Peptidase C39 family	-	-	-	-	-	-	-	-	-		-	-	-	-	-	1	-	-	
PF04304: DUF454	-	-	-	-	-	-	-	-	-		-	-	-	1	-	-	-	-	
PF05048: Periplasmic copper-binding protein (NosD)	1	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	-	-	
PF07705: CARDB	3	-	-	-	-	-	-				-	-	-	-	-	-	-	-	
PF09880: DUF2107	1	-	1	-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	
PF05399: Ectropic viral integration site 2A protein (EVI2A)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	
PF13190: PDGLE domain	1	-	1	1	-	-	-	-	-	-	-	1	1	-	-	1	-	-	
PF13229: Right handed beta helix	4	4	1	1	1	3	2	2	1	2	1	2	1	6	-	2	-	1	
PF13360: PQQ-like domain	-	-	-	1	1	1	-	-	-	-	1	1	1	1	1	-	-	-	
PF13620: Carboxypeptidase regulatory-like domain	3	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	
C-terminal anchored																			
PF01345: DUF11	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	1	-	-	
PF02514: CobN/Magnesium Chelatase	-	-	-	1	-	-	-	-	-	-	-	1	1	-	-	-	-	-	
PF02663: FmdE, Molybdenum formylmethanofuran dehydrogenase operon	-	-	1	1	-	1	-	-	1	1	-	-	-	-	-	-	-	-	
PF05048: Periplasmic copper-binding protein (NosD)	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	
PF13229: Right handed beta helix	-	-	-	-	-	-	-	-	1	1	-	1	1	-	1	-	-	-	
PF13620: Carboxypeptidase regulatory-like domain	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
N-terminal anchored																			
PF00112: peptidase_C1	5	2	5	6	1	-	5	7	1	1	1	1	1	-	-	-	-	-	
PF00188: Cysteine-rich secretory protein family	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	-	-	
PF00491: Arginase family	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

PF00496: Bacterial extracellular solute-binding proteins	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF00497: Bacterial extracellular solute-binding proteins, family 3	1	-	-	-	-	1	-	2	1	1	-	-	-	-	-	-	-	-	
PF00545: ribonuclease	1	1	-	1	1	-	1	1	1	1	-	-	-	-	-	-	-	-	
PF00565: Staphylococcal nuclease homologue	2	1	1	1	1	1	1	1	1	-	1	2	1	-	1	1	-	-	
PF00717: Peptidase S24-like	1	-	-	-	-	-	-	2	-	-	1	-	-	-	-	1	1	-	
PF00769: Ezrin/radixin/moesin family	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF00831: Ribosomal L29 protein	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF00999: Sodium/hydrogen exchanger family	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	
PF01145: SPFH domain / Band 7 family	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF01157: Ribosomal protein L21e	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-		-	-	
PF01345: DUF11	2	-	1	-	2	-	1	-	3	1	-	-	-	-	6	6	-	-	
PF01471: Putative peptidoglycan binding domain	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	
PF01497: Periplasmic binding protein	-	1	-	-	-	1		-	-	-	-	-	-	-	-	2	1	-	
PF01547: Bacterial extracellular solute-binding protein	1	-	-	-	-		1	-	-	-	-	-	-	-	-	-	-	-	
PF01696: Adenovirus EB1 55K protein / large t-antigen	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF01789: PsbP	-	-	-		-	-	-	-	-	-	-	1	1	-	-	-	-	-	
PF01841: Transglutaminase-like superfamily	-	-	1	1	1	-	1	1	2	2	2	3	2	8	4	-	1	2	
PF02514: CobN/Magnesium Chelatase	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	
PF02553: Cobalt transport protein component CbiN	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	
PF02663: FmdE, Molybdenum formylmethanofuran dehydrogenase operon	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	
PF02805: Metal binding domain of Ada	-	1	1	1	1	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF03174: Chitobiase/beta-hexosaminidase C-terminal domain	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF03190: DUF255	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	
PF03306: Alpha-acetolactate decarboxylase	1		1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF03412: Peptidase C39 family	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	2	-	1	
PF03413: Peptidase propeptide and YPEB domain	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF04156: IncA protein	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF04289 :DUF447	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF04909: Amidohydrolase	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF04952: Succinylglutamate desuccinylase	-	-	-	-	-	-	-	-	1	1	-	-	-	1	-	1	-	-	
PF05016: Plasmid stabilisation system protein	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF05048: Periplasmic copper-binding protein (NosD)	2	-	-	-	-	-	1	-	-	-	-	-	1	2	3	-	-	-	
PF05738: Cna protein B-type domain	2	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	
PF06009: Laminin Domain II	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF06207: DUF1002	2	1	1	1	2	-	1	1	1	-	-	-	-	-	-	-	-	-	
PF07602: DUF1565	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	1	
PF07685: CobB/CobQ-like glutamine amidotransferase domain	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF07705: CARDB	1	-	-	-	-	-	-	-	-	-	-	-	-	4	2	3	-	2	

PF07790: DUF1628	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF08308: PEGA domain	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF09084: NMT1/THI5 like	1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF09394: Chagasin family peptidase inhibitor I42	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF09587: Bacterial capsule synthesis protein PGA_cap	-	-	-	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF09752: DUF2048	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	
PF09972: DUF2207	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF10670: DUF4198	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF11824: DUF3344	1	-	3	2	-	1	-	2	1	1	-	-	-	-	-	-	-	-	
PF11959: DUF3473	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF12682: Flavodoxin_4	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF12804: MobA-like NTP transferase domain	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	
PF12849: PBP superfamily domain	-	-	1	-	-	-	-	-	-	-	-	1	1	-	1	1	1	-	
PF13098: Thioredoxin-like domain	-	1	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF13200: Putative glycosyl hydrolase domain (DUF4015)	1	-	-	1	-	-	1	-	-	-	1	-	-	-	-	-	-	-	
PF13229: Right handed beta helix	3	5	10	5	7	5	9	12	6	6	2	1	4	27	15	6	3	7	
PF13360: PQQ-like domain	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF13499: EF-hand domain pair	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF13531: Bacterial extracellular solute-binding protein	-	2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13570: PQQ-like domain	1	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	
PF13620: Carboxypeptidase regulatory-like domain	12	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	1	-	
PF14347: DUF4399	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF14947: Winged helix-turn-helix	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13379: NMT1-like family	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	1	-	-	
Secreted																			
PF00112: peptidase_C1	-	-	6	3	-	-	-	2	-	-	-	-	-	-	-	-	-	-	
PF00188: Cysteine-rich secretory protein family	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF00215: Orotidine 5'-phosphate decarboxylase / HUMPS family	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF00534: Glycosyl transferases group 1	-	-	-	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	
PF00590: Tetrapyrrole (Corrin/Porphyrin) Methylases	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF00682: HMGL-like	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF00984: UDP-glucose/GDP-mannose dehydrogenase family	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF01058: NADH ubiquinone oxidoreductase, 20 Kd subunit	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	1	
PF01157: Ribosomal protein L21e	1	-	-	1	-	-	-	-	-	-	-	-	-	1	1	1	1	-	
PF01180: Dihydroorotate dehydrogenase	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	
PF01379: Porphobilinogen deaminase, dipyromethane cofactor binding domain	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	
PF01345: DUF11	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	4	-	-	

PF01370: NAD dependent epimerase/dehydratase family	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	
PF01451: Low molecular weight phosphotyrosine protein phosphatase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	
PF01471: Putative peptidoglycan binding domain	1	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF01497: Periplasmic binding protein		-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	
PF01554: MatE	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF01641: SelR domain	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF01841: Transglutaminase-like superfamily	2	-	2	-	-	1	-	1	3	3		1	1	2	3	3	-	2	
PF01850: PIN domain	-	-	-	-	-	-	-		-	-	-	-	-	-	-	1	-	1	
PF01872: RibD C-terminal domain	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	
PF01888: CbiD	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	
PF01890: Cobalamin synthesis G C-terminus	-	-	-	-	-	-	-	-	-	-	-	-	-		1	-	-	-	
PF01909: Nucleotidyltransferase domain	-	-	-	-	-	-	-	-	-	-	-	-	-	1		-	-	-	
PF02254: TrkA-N domain	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF02514: CobN/Magnesium Chelatase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	
PF02558: Ketopantoate reductase PanE/ApbA	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF02805: Metal binding domain of Ada	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	
PF03412: Peptidase C39 family	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	
PF03900: Porphobilinogen deaminase, C-terminal domain	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	
PF04427: Brix domain	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF04466: Phage terminase large subunit	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF04851: Type III restriction enzyme, res subunit	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF04919: DUF655	-	-	-	-		-	-	-	-	-	-	-	-	-	-	1	-	-	
PF04952: Succinylglutamate desuccinylase	1	-	-	-	1	-	-	-	-	-	-	-	-	1	-	1	1	-	
PF05048: Periplasmic copper-binding protein (NosD)	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	
PF06207: DUF1002	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	
PF07602: DUF1565	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF07705: CARDB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	
PF08011: PD-(D/E)XK nuclease superfamily	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF08245: Mur ligase middle domain	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF08353: DUF1727	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF08502: LeuA allosteric (dimerisation) domain	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF08859: DGC domain	1	1	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	
PF09820: Predicted AAA-ATPase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	
PF09869: DUF2096	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	
PF11760: Cobalamin synthesis G N-terminal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	
PF11797: DUF3324	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13200: Putative glycosyl hydrolase domain (DUF4015)	-	-	1	1	-	-	-	-	-	-	-	-	-	2	-	-	-	-	
PF13229: Right handed beta helix	1	1	2	3	-	-	1	1	-	-	-	-	-	3	-	-	-	-	
PF13290: Chitobiase/beta-hexosaminidase C-terminal domain	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	
PF13307: Helicase C-terminal domain	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	

PF13360: PQQ-like domain	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PF13379: NMT1-like family	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
PF13620: Carboxypeptidase regulatory-like domain	3	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
PF17147: Pyruvate:Fdx oxidoreductase core domain II	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-

Prediction cutoff: E-05.

Table A.6.5. Predicted *Methanobrevibacter* sp. D5 genes within the *Methanobrevibacter* core genome

****

Locus_tag	Predicted gene product	COG category
D5_0001	cell division control protein Cdc6	[L]
D5_0004	ATP phosphoribosyltransferase HisG	[E]
D5_0005	archaeal histone HmfB	[B]
D5_0006	amidohydrolase	[F]
D5_0007	succinate dehydrogenase/fumarate reductase flavoprotein subunit SdhA	[C]
D5_0013	Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit A GatA	[J]
D5_0021	xylose isomerase-like TIM barrel domain-containing protein	[G]
D5_0028	PHP domain-containing protein	[R]
D5_0033	radical SAM domain-containing protein	[R]
D5_0039	dihydroxy-acid dehydratase	[E]
D5_0040	histidinol dehydrogenase	[E]
D5_0042	aspartyl-tRNA synthetase AspS	[J]
D5_0045	glutamate-1-semialdehyde-2,1-aminomutase HemL	[H]
D5_0046	signal peptidase I	[U]
D5_0047	arginyl-tRNA synthetase ArgS	[J]
D5_0050	MATE efflux family protein, matE	[V]
D5_0052	phosphoribosylamine--glycine ligase, PurD	[F]
D5_0054	ornithine carbamoyltransferase ArgF	[E]
D5_0057	acetolactate synthase large subunit IlvB	[E]
D5_0058	acetolactate synthase small subunit IlvN	[E]
D5_0059	carbonic anhydrase Cab	[E]
D5_0061	ketol-acid reductoisomerase IlvC	[E]
D5_0062	methanogenesis marker protein 12	[S]
D5_0063	LSM domain-containing protein	[K]
D5_0065	5'-nucleotidase SurE	[R]
D5_0067	conserved hypothetical transmembrane protein	[S]
D5_0069	threonyl-tRNA synthetase ThrS	[J]
D5_0070	bifunctional formaldehyde-activating enzyme /3- hexulose-6-phosphate synthase Fae/Hps	[G]
D5_0074	NADPH-dependent FMN reductase	[R]
D5_0075	polysaccharide biosynthesis protein	[R]
D5_0076	molybdate transport system regulatory protein ModE	[R]
D5_0078	imidazoleglycerol-phosphate dehydratase HisB	[E]
D5_0080	branched-chain-amino-acid aminotransferase IlvE	[E]
D5_0081	undecaprenyl-diphosphatase	[V]
D5_0084	nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase CobT	[H]
D5_0086	glycoprotease M22 family	[O]
D5_0087	RdgB/HAM1 family non-canonical purine NTP pyrophosphatase	[F]
D5_0088	ribosomal protein S15P Rps15p	[J]
D5_0090	aconitase	[S]
D5_0096	Hef nuclease	[L]
D5_0097	xylose isomerase-like TIM barrel domain-containing protein	[G]
D5_0098	PIN domain-containing protein	[R]
D5_0099	phosphoribosyl-AMP cyclohydrolase HisI	[E]
D5_0100	histidyl-tRNA synthetase HisS	[J]
D5_0101	shikimate 5-dehydrogenase AroE	[E]
D5_0103	replication factor C large subunit RfcL	[K]
D5_0104	replication factor C small subunit RfcS	[L]
D5_0105	peptidase M48 family	[O]
D5_0106	tRNA(1-methyladenosine) methyltransferase	[J]
D5_0107	leucyl-tRNA synthetase LeuS	[J]
D5_0126	Anion-transporting ATPase	[P]
D5_0134	translation-associated GTPase	[J]
D5_0142	methanogenesis marker 13	[C]
D5_0143	imidazole glycerol phosphate synthase	[E]
D5_0144	hydrogenase expression/formation protein, HypE	[O]
D5_0148	hypothetical protein	[S]
D5_0149	amidohydrolase	[Q]
D5_0153	adenylosuccinate lyase PurB	[F]
D5_0154	HTH domain-containing protein	[K]
D5_0164	DNA polymerase sliding clamp subunit PCNA family Pcn	[S]
D5_0167	ribosomal protein S27e Rps27e	[J]
D5_0168	translation initiation factor aIF-2 alpha subunit	[J]
D5_0169	H/ACA RNA-protein complex component Nop10p	[R]
D5_0170	conserved hypothetical protein	[S]
D5_0174	coenzyme F_{420} hydrogenase alpha subunit FrhA	[C]
D5_0175	coenzyme F_{420} hydrogenase delta subunit FrhD	[C]
D5_0176	coenzyme F_{420} hydrogenase gamma subunit FrhG	[C]
D5_0177	coenzyme F_{420} hydrogenase beta subunit FrhB	[C]
D5_0178	methionine aminopeptidase Map	[J]
D5_0179	conserved hypothetical secreted protein	[Unclassified]
D5_0190	molecular chaperone DnaJ	[O]
D5_0191	chaperone protein DnaK	[O]
D5_0192	chaperone protein DnaK	[O]
D5_0193	transcriptional regulator ArsR family	[K]
D5_0194	hydrogenase maturation factor HypF	[O]
D5_0199	polysaccharide/polyol phosphate ABC transporter ATP-binding protein	[G]
D5_0200	polysaccharide/polyol phosphate ABC transporter permease protein	[G]
D5_0201	phosphoribosyl-ATP pyrophosphohydrolase HisE	[E]
D5_0202	CBS domain-containing protein	[K]
D5_0203	Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit B GatB	[J]
D5_0204	radical SAM domain-containing protein	[H]
D5_0205	archaeal Holliday junction resolvase Hjc	[L]
D5_0206	metallo-beta-lactamase superfamily protein	[R]
D5_0207	potassium uptake protein TrkA family	[P]
D5_0208	potassium uptake protein TrkH family	[P]
D5_0210	formylmethanofuran-H_4MPT formyltransferase Ftr	[C]
D5_0213	carbohydrate kinase	[G]
D5_0214	sugar fermentation stimulation protein SfsA	[R]
D5_0217	argininosuccinate synthase ArgG	[E]
D5_0221	succinylglutamate desuccinylase /aspartoacylase	[R]
D5_0225	energy-converting hydrogenase B subunit A EhbA	[P]
D5_0226	energy-converting hydrogenase B subunit B EhbB	[P]
D5_0229	energy-converting hydrogenase B subunit E EhbE	[P]
D5_0230	energy-converting hydrogenase B subunit F EhbF	[C]
D5_0232	energy-converting hydrogenase B subunit H EhbH	[P]
D5_0233	energy-converting hydrogenase B subunit I EhbI	[P]
D5_0235	energy-converting hydrogenase B subunit K EhbK	Not in COGs
D5_0236	energy-converting hydrogenase B subunit L EhbL	[C]
D5_0237	energy-converting hydrogenase B subunit M EhbM	[C]
D5_0238	energy-converting hydrogenase B subunit N EhbN	[C]
D5_0239	energy-converting hydrogenase B, subunit O, EhbO	[C]
D5_0240	energy-converting hydrogenase B subunit P EhbP	Not in COGs
D5_0241	energy-converting hydrogenase B subunit Q EhbQ	[R]
D5_0244	conserved hypothetical protein	[S]
D5_0248	CBS domain-containing protein	[R]
D5_0249	CBS domain-containing protein	[R]

D5_0250	prephenate dehydratase PheA	[E]
D5_0254	phosphopantothenate-cysteine ligase CoaB	[H]
D5_0255	fibrillarin FlpA	[J]
D5_0256	pre-mRNA splicing ribonucleoprotein PRP31	[J]
D5_0258	dihydroorotate dehydrogenase PyrD	[F]
D5_0259	dihydroorotate dehydrogenase electron transfer subunit PyrK	[H]
D5_0260	conserved hypothetical transmembrane protein	[R]
D5_0261	DNA polymerase family B PolB	[L]
D5_0263	L-sulfolactate dehydrogenase ComC	[C]
D5_0264	phosphoribosylformylglycinamidine cyclo-ligase PurM	[F]
D5_0265	RNA-metabolising metallo-beta-lactamase	[R]
D5_0266	proteasome beta subunit	[O]
D5_0267	Met-10 like-protein	[R]
D5_0268	FO synthase subunit 1 CofG	[H]
D5_0270	GTP cyclohydrolase MptA	[S]
D5_0271	transcriptional regulator AsnC family	[K]
D5_0272	N^2,N^2-dimethylguanosine tRNA methyltransferase Trm1	[J]
D5_0273	aminotransferase DegT/DnrJ/EryC1/StrS family	[M]
D5_0275	PP-loop family protein	[D]
D5_0277	CBS domain-containing protein	[R]
D5_0278	conserved hypothetical protein	[Unclassified]
D5_0279	TraB family protein	[S]
D5_0280	2-phosphosulfolactate phosphatase ComB	[H]
D5_0282	methanogenesis marker protein 7	[R]
D5_0283	methanogenesis marker protein 10	[C]
D5_0284	methyl-CoM reductase beta subunit McrB	[H]
D5_0285	methyl-CoM reductase D subunit McrD	[H]
D5_0286	methyl-CoM reductase C subunit McrC	[H]
D5_0288	methyl-CoM reductase alpha subunit McrA	[H]
D5_0289	H_4MPT S-methyltransferase subunit E MtrE	[H]
D5_0290	H_4MPT S-methyltransferase subunit D MtrD	[H]
D5_0291	H_4MPT S-methyltransferase subunit C MtrC	[H]
D5_0292	H_4MPT S-methyltransferase subunit B MtrB	[H]
D5_0294	H_4MPT S-methyltransferase subunit F MtrF	[H]
D5_0295	H_4MPT S-methyltransferase subunit G MtrG	[H]
D5_0297	methanogenesis marker protein 14	[S]
D5_0299	PRC-barrel domain-containing protein	[S]
D5_0302	methyl viologen-reducing hydrogenase delta subunit MvhD	[C]
D5_0303	methyl viologen-reducing hydrogenase gamma subunit MvhG	[C]
D5_0304	methyl viologen-reducing hydrogenase alpha subunit MvhA	[C]
D5_0305	methyl viologen-reducing hydrogenase beta subunit MvhB	[C]
D5_0308	nucleotidyl transferase	[R]
D5_0310	hypothetical protein	Not in COGs
D5_0311	ribosomal protein L10e Rpl10e	[J]
D5_0312	phosphoenolpyruvate synthase PpsA	[G]
D5_0313	L-tyrosine decarboxylase MfnA	[E]
D5_0314	dihydronepterin aldolase MtnD	[S]
D5_0315	sortase family protein	[M]
D5_0317	phosphoenolpyruvate synthase PpsA	[G]
D5_0319	cell shape determining protein, MreB/Mrl family	[D]
D5_0320	ribonuclease HII RnhB	[L]
D5_0322	transporter ExbD/TolR family	[Unclassified]
D5_0323	IMP cyclohydrolase PurO	[F]
D5_0324	F_{420}-*0*:gamma-glutamyl ligase CofE	[S]
D5_0325	LPPG:FO 2-phospho-L-lactate transferase CofD	[S]
D5_0326	GTP cyclohydrolase III ArfA	[S]
D5_0327	tRNA-dihydrouridine synthase	[J]
D5_0328	methyl-CoM reductase component A2 AtwA	[R]
D5_0332	methanogenesis marker domain 9	[K]
D5_0333	siroheme synthase CysG	[H]
D5_0334	glutamyl-tRNA reductase HemA	[H]
D5_0335	ATPase	[R]
D5_0339	glyceraldehyde-3-phosphate dehydrogenase Gap	[G]
D5_0347	DNA topoisomerase VI subunit A	[L]
D5_0348	DNA topoisomerase VI subunit B	[L]
D5_0349	RNA-binding protein	[R]
D5_0350	H_4MPT S-methyltransferase subunit H MtrH2	[H]
D5_0357	serine/threonine protein kinase RIO1 family	[T]
D5_0358	translation initiation factor aIF-1A	[J]
D5_0359	molybdopterin biosynthesis protein MoeA	[H]
D5_0360	transcriptional regulator	[K]
D5_0362	peptidase U62 family	[R]
D5_0364	hydrogenase expression/formation protein HypD	[O]
D5_0369	digeranylgeranylglyceryl phosphate synthase DGGGPS	[H]
D5_0370	myo-inositol-1-phosphate synthase	[I]
D5_0372	pyruvate carboxylase subunit B PycB	[C]
D5_0390	conserved hypothetical protein	[S]
D5_0396	fumarate hydratase FumA	[C]
D5_0398	2-oxoglutarate Fdx oxidoreductase subunit alpha KorA	[C]
D5_0399	2-oxoglutarate Fdx oxidoreductase subunit beta KorB	[C]
D5_0400	2-oxoglutarate Fdx oxidoreductase subunit gamma KorC	[C]
D5_0401	succinyl-CoA synthetase beta subunit SucC	[C]
D5_0402	radical SAM domain-containing protein	Not in COGs
D5_0405	triose-phosphate isomerase TpiA	[G]
D5_0406	phosphoglycerate kinase Pgk	[G]
D5_0408	hydroxyethylthiazole kinase	[H]
D5_0409	DNA-directed RNA polymerase subunit H RpoH	[K]
D5_0411	DNA-directed RNA polymerase subunit B' RpoB1	[K]
D5_0412	DNA-directed RNA polymerase subunit A' RpoA1	[K]
D5_0413	DNA-directed RNA polymerase subunit A" RpoA2	[K]
D5_0414	ribosomal protein L30e Rpl30e	[J]
D5_0415	transcription elongation factor NusA-like protein	[K]
D5_0416	ribosomal protein S12P Rps12p	[J]
D5_0419	translation elongation factor aEF-1 alpha	[J]
D5_0420	ribosomal protein S10P Rps10p	[J]
D5_0423	conserved hypothetical protein	[S]
D5_0425	ribosomal-protein-alanine acetyltransferase RimI	[R]
D5_0428	peptide chain release factor aRF1	[J]
D5_0429	glutamate dehydrogenase GdhA	[E]
D5_0498	UDP-glucose pyrophosphorylase GalU	[M]
D5_0510	H_4MPT *S*-methyltransferase subunit A MtrA	[H]
D5_0517	orotate phosphoribosyltransferase PyrE	[F]
D5_0518	oxidoreductase GFO/IDH/MOCA family	[R]
D5_0519	porphobilinogen deaminase HemC	[H]
D5_0521	bifunctional inositol-1 monophosphatase/ fructose-1,6-bisphosphatase/ATP-NAD kinase	[G]
D5_0522	pyruvoyl-dependent arginine decarboxylase PdaD	[S]
D5_0523	translation initiation factor aIF-5A	[J]
D5_0524	arginase/agmatinase family protein	[E]
D5_0531	conserved hypothetical protein	[S]
D5_0532	adenine deaminase Ade	[F]
D5_0534	phosphosugar isomerase	[M]
D5_0543	GTP-binding protein	[R]
D5_0544	AMMECR1 domain-containing protein	[S]
D5_0545	peptidase U62 family	[R]
D5_0546	GMC oxidoreductase family protein	[E]
D5_0547	aspartate carbamoyltransferase regulatory subunit PyrI	[F]
D5_0549	*N*-acetyl-gamma-glutamyl-phosphate reductase ArgC	[E]
D5_0550	glycerol-3-phosphate cytidylyltransferase	[M]

D5_0551	phosphoribosylformimino-5- aminoimidazole carboxamide ribotide isomerase, hisA	[E]
D5_0560	conserved hypothetical transmembrane protein	[Unclassified]
D5_0581	tRNA pseudouridine synthase A TruA	[J]
D5_0583	UDP-*N*-acetylglucosamine 2-epimerase	[R]
D5_0585	(4-{4-[2-(γ-L-glutamylamino)ethyl]phenoxymethyl}furan-2-yl)methanamine synthase MfnF	[I]
D5_0586	conserved hypothetical transmembrane protein	[R]
D5_0587	7,8-dihydro-6-hydroxymethylpterin dimethyltransferase	[E]
D5_0589	hypothetical protein	[R]
D5_0592	MiaB-like tRNA modifying enzyme	[J]
D5_0595	histone acetyltransferase ELP3 family	[K]
D5_0598	DEAD/DEAH box helicase domain-containing protein	[R]
D5_0599	thioredoxin	[O]
D5_0603	glycosyl transferase GT4 family	[M]
D5_0604	shikimate kinase AroK	[E]
D5_0606	chorismate mutase, aroH	[E]
D5_0608	aspartate kinase Ask	[E]
D5_0609	dihydrodipicolinate synthase DapA	[E]
D5_0610	dihydrodipicolinate reductase DapB	[E]
D5_0611	aspartate-semialdehyde dehydrogenase Asd	[E]
D5_0614	thermosome subunit	[O]
D5_0619	PRC-barrel domain-containing protein	[S]
D5_0620	orotate phosphoribosyltransferase PyrE	[F]
D5_0621	molybdenum cofactor biosynthesis protein B MoaB	[H]
D5_0625	hypothetical protein	[R]
D5_0627	methanogenesis marker protein 2	[R]
D5_0631	methanogenesis marker protein 3	[O]
D5_0632	methanogenesis marker protein 6	[S]
D5_0633	methanogenesis marker protein 5	[S]
D5_0634	methanogenesis marker protein 15	[I]
D5_0635	methanogenesis marker protein 17	[S]
D5_0636	nitrogenase cofactor biosynthesis protein NifB	[R]
D5_0639	translation initiation factor aIF-2B alpha subunit	[J]
D5_0643	Met-10 like-protein	[R]
D5_0644	diphthine synthase DphB	[J]
D5_0648	archaeosine tRNA-ribosyltransferase TgtA	[J]
D5_0649	CoB--CoM heterodisulfide reductase subunit C HdrC	[C]
D5_0650	CoB--CoM heterodisulfide reductase subunit B HdrB	[C]
D5_0651	conserved hypothetical protein	[S]
D5_0652	conserved hypothetical protein	[S]
D5_0653	calcineurin-like phosphoesterase	[R]
D5_0655	2-phosphoglycerate kinase Pgk	[G]
D5_0656	CBS domain-containing protein	[R]
D5_0657	transporter CDF family	[P]
D5_0658	alanine aminotransferase	[E]
D5_0659	radical SAM domain-containing protein	[C]
D5_0661	phosphopantetheine adenylyltransferase CoaD	[E]
D5_0667	metallo-beta-lactamase superfamily protein	[Unclassified]
D5_0670	conserved hypothetical protein	[S]
D5_0674	fumarate hydratase FumA	[L]
D5_0675	conserved hypothetical protein	[C]
D5_0676	pyridoxal phosphate enzyme	[Unclassified]
D5_0677	biotin-acetyl-CoA-carboxylase ligase BirA	[E]
D5_0678	pyruvate carboxylase subunit A PycA	[H]
D5_0680	conserved hypothetical protein	[J]
D5_0681	ribosomal protein L3P Rpl3p	[S]
D5_0682	ribosomal protein L4p Rpl4p	[J]
D5_0683	ribosomal protein L23P Rpl23p	[J]
D5_0685	ribosomal protein S19P Rps19p	[J]
D5_0686	ribosomal protein L22P Rpl22p	[J]
D5_0687	ribosomal protein S3P Rps3p	[J]
D5_0688	ribosomal protein L29P Rpl29p	[J]
D5_0689	translation initiation factor aSUI1	[J]
D5_0692	ribosomal protein L14P Rpl14p	[J]
D5_0693	ribosomal protein L24P Rpl24p	[J]
D5_0694	ribosomal protein S4e Rps4e	[J]
D5_0695	ribosomal protein L5P Rpl5p	[J]
D5_0696	ribosomal protein S8P Rps8p	[J]
D5_0697	ribosomal protein L6P Rpl6p	[J]
D5_0698	ribosomal protein L32e Rpl32e	[J]
D5_0699	ribosomal protein L19e Rpl19e	[J]
D5_0700	ribosomal protein L18P Rpl18p	[J]
D5_0701	ribosomal protein S5P Rps5p	[J]
D5_0702	ribosomal protein L30P Rpl30p	[J]
D5_0703	ribosomal protein L15P Rpl15p	[J]
D5_0704	preprotein translocase subunit SecY	[U]
D5_0705	adenylate kinase Adk	[F]
D5_0706	conserved hypothetical transmembrane protein	[S]
D5_0708	cytidylate kinase Cmk	[S]
D5_0709	ribosomal protein L14e Rpl14e	[J]
D5_0710	H/ACA RNA-protein complex component Cbf5p	[J]
D5_0719	ribosomal protein S13P Rps13p	[J]
D5_0720	ribosomal protein S4P Rps4p	[J]
D5_0721	ribosomal protein S11P Rps11p	[J]
D5_0722	DNA-directed RNA polymerase subunit D RpoD	[C]
D5_0723	ribosomal protein L18e Rpl18e	[J]
D5_0724	ribosomal protein L13P Rpl13p	[J]
D5_0725	ribosomal protein S9P Rps9p	[J]
D5_0726	DNA-directed RNA polymerase subunit N RpoN	[K]
D5_0727	DNA-directed RNA polymerase subunit K RpoK	[K]
D5_0728	phosphopyruvate hydratase Eno	[G]
D5_0730	ribosomal protein S2P Rps2p	[J]
D5_0731	conserved hypothetical protein	[R]
D5_0732	mevalonate kinase Mvk	[I]
D5_0733	isopentenyl diphosphate kinase	[R]
D5_0734	isopentenyl diphosphate delta-isomerase Fni	[C]
D5_0735	RNA-metabolising metallo-beta-lactamase	[R]
D5_0736	bifunctional short chain isoprenyl diphosphate synthase IdsA	[H]
D5_0741	F_{420}-dependent methylenetetrahydromethanopterin dehydrogenase Mtd	[C]
D5_0744	methyl-CoM reductase II gamma subunit MrtG	[H]
D5_0745	methyl-CoM reductase II alpha subunit MrtA	[H]
D5_0748	transcriptional regulator TrmB family	[K]
D5_0749	hydroxylamine reductase Hcp	[C]
D5_0750	cupin 2 domain-containing protein	[S]
D5_0751	SAM-dependent methyltransferase	[Q]
D5_0754	glutamyl-tRNA synthetase GltX	[J]
D5_0757	diaminopimelate aminotransferase DapL	[E]
D5_0766	adenylosuccinate synthetase PurA	[F]
D5_0785	conserved hypothetical protein	[S]
D5_0786	metallo-beta-lactamase superfamily protein	[R]
D5_0787	chorismate synthase AroC	[E]
D5_0790	ATP:dephospho-CoA triphosphoribosyl transferase CitG	[H]
D5_0800	phenylalanyl-tRNA synthetase alpha subunit PheS	[J]
D5_0801	exodeoxyribonuclease III Xth	[L]
D5_0802	hydrolase HAD superfamily	[R]
D5_0806	ribosomal protein S8e Rps8e	[J]
D5_0807	hydrogenase expression/formation protein HypE	[O]
D5_0808	hypothetical transmembrane protein	[Unclassified]
D5_0811	CRISPR-associated protein Cas4	[Unclassified]
D5_0814	transcriptional regulator MarR family	[K]
D5_0815	ssDNA exonuclease RecJ	[L]
D5_0816	signal recognition particle SRP19 protein	[U]
D5_0819	uroporphyrinogen III synthase HemD	[H]
D5_0855	hypothetical transmembrane protein	[S]
D5_0862	phosphoribosylaminoimidazole-succinocarboxamide synthase PurC	[F]
D5_0863	phosphoribosylformylglycinamidine (FGAM) synthase PurS	[F]

D5_0864	phosphoribosylformylglycinamidine (FGAM) synthase PurQ	[F]
D5_0865	uroporphyrin-III C-methyltransferase CobA	[H]
D5_0868	SAM-dependent methyltransferase	[S]
D5_0870	archaeosine tRNA-ribosyltransferase TgtA	[J]
D5_0872	hydroxymethylglutaryl-CoA synthase	[I]
D5_0873	acetyl-CoA acetyltransferase	[I]
D5_0908	glucosamine-fructose-6-phosphate aminotransferase GlmS	[M]
D5_0911	translation elongation factor aEF-2	[J]
D5_0946	cobyric acid synthase CbiP	[H]
D5_0947	ATP-dependent protease S16 family	[O]
D5_0954	ribose-phosphate diphosphokinase Prs	[F]
D5_0956	excinuclease ABC B subunit UvrB	[L]
D5_0957	ammonium transporter Amt	[P]
D5_0958	nitrogen regulatory protein P-II GlnK	[E]
D5_0959	excinuclease ABC A subunit UvrA	[L]
D5_0961	transporter MIP family	[G]
D5_0964	conserved hypothetical transmembrane protein	[S]
D5_0967	hypothetical protein	[S]
D5_0988	hypothetical protein	[Unclassified]
D5_0989	radical SAM domain-containing protein	[R]
D5_0993	transcriptional regulator	[K]
D5_0995	conserved hypothetical transmembrane protein	[Unclassified]
D5_0997	orotidine 5'-phosphate decarboxylase PyrF	[F]
D5_0998	cobalamin biosynthesis protein CbiM	[P]
D5_0999	cobalt transport protein CbiN	[P]
D5_1000	cobalt ABC transporter permease protein CbiQ	[P]
D5_1001	cobalt ABC transporter ATP-binding protein CbiO	[P]
D5_1002	riboflavin synthase RibC	[H]
D5_1003	glycosyl transferase GT2 family	[M]
D5_1005	CoB--CoM heterodisulfide reductase subunit B HdrB	[C]
D5_1006	iron-sulfur cluster-binding protein	[C]
D5_1007	2-furaldehyde phosphate synthase MfnB	[S]
D5_1008	inosine-5'-monophosphate dehydrogenase GuaB	[F]
D5_1009	ribosomal protein L37Ae Rpl37ae	[J]
D5_1012	prefoldin beta subunit PfdB	[O]
D5_1014	hisA/hisF family protein HisAF	[R]
D5_1025	helicase	[L]
D5_1026	excinuclease ABC C subunit UvrC	[L]
D5_1029	ABC transporter ATP-binding protein	[R]
D5_1031	Fdx	[C]
D5_1032	geranylgeranyl reductase family protein	[C]
D5_1033	UDP-glucose 4-epimerase GalE	[M]
D5_1036	amidophosphoribosyltransferase PurF	[F]
D5_1037	peptidase U32	[O]
D5_1039	4Fe-4S iron sulfur cluster binding protein NifH/FrxC family	[P]
D5_1075	beta-ribofuranosylaminobenzene 5'-phosphate synthase MptG	[R]
D5_1079	F_{390} synthetase FtsA	[H]
D5_1084	hypothetical transmembrane protein	[R]
D5_1283	universal stress protein, UspA	[T]
D5_1311	MatE efflux family protein	[V]
D5_1316	hypothetical transmembrane protein	[S]
D5_1461	NADPH-dependent F_{420} reductase NpdG	[R]
D5_1466	tRNA nucleotidyltransferase Cca	[J]
D5_1468	3-dehydroquinate synthase AroB	[E]
D5_1469	phospho-2-dehydro-3-deoxyheptonate aldolase/fructose-bisphosphate aldolase	[G]
D5_1470	SAM-dependent methyltransferase	[R]
D5_1479	PHP domain-containing protein	[E]
D5_1482	diaminohydroxyphosphoribosylaminopyrimidine reductase RibD	[H]
D5_1483	cell wall biosynthesis protein phospho-*N*-acetylmuramoyl-pentapeptide-transferase family	[M]
D5_1488	conserved hypothetical protein	[S]
D5_1489	methionyl-tRNA synthetase MetG	[J]
D5_1490	DNA primase large subunit PriB	[L]
D5_1495	DNA primase small subunit PriA	[L]
D5_1498	FO synthase subunit 2 CofH	[H]
D5_1507	conserved hypothetical protein	[R]
D5_1509	cyclic 2,3-diphosphoglycerate-synthetase CpgS	[R]
D5_1512	undecaprenyl pyrophosphate synthetase UppS	[I]
D5_1513	hydrolase TatD family	[L]
D5_1523	conserved hypothetical protein	[S]
D5_1524	conserved hypothetical protein	[S]
D5_1525	conserved hypothetical transmembrane protein	Not in COGs
D5_1526	conserved hypothetical protein	[S]
D5_1528	hydrogenase nickel insertion protein HypA	[R]
D5_1531	ATP-dependent DNA helicase UvrD/REP family	[L]
D5_1533	conserved hypothetical protein	[S]
D5_1537	hypothetical protein	[Unclassified]
D5_1540	ATPase	[R]
D5_1542	geranylgeranylglyceryl phosphate synthase	[R]
D5_1543	ribosomal protein L40e Rpl40e	[J]
D5_1546	conserved hypothetical protein	[S]
D5_1549	nicotinamide-nucleotide adenylyltransferase	[H]
D5_1553	conserved hypothetical protein	[S]
D5_1555	glutamyl aminopeptidase PepA	[G]
D5_1558	exonuclease	[L]
D5_1560	pantothenate synthase PanC	[S]
D5_1561	dephospho-CoA kinase CoaE	[H]
D5_1562	ATP-binding protein	[R]
D5_1566	CTP synthase PyrG	[F]
D5_1572	2,3-bisphosphoglycerate-independent phosphoglycerate mutase ApgM	[G]
D5_1575	Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase subunit C GatC	Not in COGs
D5_1588	homoserine *O*-acetyltransferase MetX	[E]
D5_1589	ribonuclease III Rnc	[K]
D5_1590	conserved hypothetical protein	Not in COGs
D5_1596	RNA-binding protein	[J]
D5_1597	creatinine amidohydrolase ArfB	[R]
D5_1599	conserved hypothetical protein	[S]
D5_1626	methenyltetrahydromethanopterin cyclohydrolase Mch	[H]
D5_1635	*O*-acetylhomoserine/*O*-acetylserine sulfhydrylase MetZ/CysK2	[E]
D5_1641	ATPase	[R]
D5_1643	argininosuccinate lyase ArgH	[J]
D5_1644	ribosomal protein S27ae Rps27ae	[J]
D5_1645	ribosomal protein S24e Rps24e	[S]
D5_1647	DNA-directed RNA polymerase subunit E" RpoE2	[K]
D5_1648	DNA-directed RNA polymerase subunit E' RpoE1	[R]
D5_1649	conserved hypothetical protein	[J]
D5_1651	ribosomal protein S6e Rps6e	[J]
D5_1652	translation initiation factor IF-2	[F]
D5_1653	nucleoside diphosphate kinase Ndk	[J]
D5_1654	ribosomal protein L24e Rpl24e	[J]
D5_1655	ribosomal protein S28e Rps28e	[J]
D5_1656	ribosomal protein L7Ae Rpl7ae	[E]
D5_1657	threonine synthase ThrC	[S]
D5_1659	tryptophanyl-tRNA synthetase TrpS	[J]
D5_1660	tRNA intron endonuclease EndA	[K]
D5_1661	iron-dependent repressor	[Unclassified]
D5_1666	conserved hypothetical protein	[S]
D5_1667	conserved hypothetical transmembrane protein	Not in COGs
D5_1669	hydroxymethylglutaryl-CoA reductase (NADPH) HmgA	[I]
D5_1670	succinate-CoA ligase alpha subunit SucD	[C]
D5_1671	conserved hypothetical protein	[S]
D5_1672	3-dehydroquinate dehydratase type I AroD	[E]
D5_1677	HD domain-containing protein	[R]
D5_1678	3-polyprenyl-4-hydroxybenzoate decarboxylase UbiX	[H]

D5_1682	exosome complex RNA-binding protein Rrp42	[J]
D5_1683	exosome complex exonuclease Rrp41	[J]
D5_1684	exosome complex RNA-binding protein Rrp4	[J]
D5_1685	exosome subunit	[J]
D5_1686	proteasome alpha subunit PsmA	[O]
D5_1687	ribonuclease P subunit P14	[J]
D5_1694	archaea-specific RecJ-like exonuclease	[L]
D5_1697	NifU-like FeS cluster assembly scaffold protein	[C]
D5_1698	cysteine desulfurase NifS	[C]
D5_1700	cysteinyl-tRNA synthetase CysS	[E]
D5_1701	conserved hypothetical protein	[J]
D5_1702	serine *O*-acetyltransferase CysE	[R]
D5_1703	cysteine synthase CysKM1	[E]
D5_1704	endonuclease III Nth	[E]
D5_1705	3-phosphoshikimate 1-carboxyvinyltransferase AroA	[L]
D5_1708	valyl-tRNA synthetase ValS	[J]
D5_1712	conserved hypothetical transmembrane protein	Not in COGs
D5_1713	phenylalanyl-tRNA synthetase subunit beta PheT	[J]
D5_1714	conserved hypothetical protein	[S]
D5_1719	ribose 5-phosphate isomerase A RpiA	[G]
D5_1720	hypothetical protein	[S]
D5_1721	NAD(P)-dependent glycerol-1-phosphate dehydrogenase EgsA	[C]
D5_1722	prolyl-tRNA synthetase ProS	[J]
D5_1723	2-phospho-L-lactate guanylyltransferase CofC	[S]
D5_1724	phosphomethylpyrimidine kinase ThiD	[H]
D5_1732	carboxymuconolactone decarboxylase family protein	[S]
D5_1748	exosome subunit	[J]
D5_1749	ribosomal protein L15e Rpl15e	[J]
D5_1751	tungsten formylmethanofuran dehydrogenase subunit E FwdE	[C]
D5_1766	CBS domain-containing protein	[R]
D5_1768	carbohydrate kinase PfkB family	[G]
D5_1769	formylmethanofuran- H_4MPT formyltransferase Ftr1	[C]
D5_1771	energy-converting hydrogenase A subunit Q EhaQ	[C]
D5_1772	energy-converting hydrogenase A subunit P EhaP	[C]
D5_1773	energy-converting hydrogenase A subunit O EhaO	[C]
D5_1774	energy-converting hydrogenase A subunit N EhaN	[C]
D5_1778	energy-converting hydrogenase A subunit J EhaJ	[C]
D5_1779	energy-converting hydrogenase A subunit I EhaI	[Unclassified]
D5_1780	energy-converting hydrogenase A subunit H EhaH	[S]
D5_1781	energy-converting hydrogenase A subunit G EhaG	[S]
D5_1782	energy-converting hydrogenase A subunit F EhaF	[S]
D5_1783	energy-converting hydrogenase A subunit E EhaE	[S]
D5_1784	energy-converting hydrogenase A subunit D EhaD	[S]
D5_1785	energy-converting hydrogenase A subunit C EhaC	[S]
D5_1786	energy-converting hydrogenase A subunit B EhaB	[S]
D5_1787	energy-converting hydrogenase A subunit A EhaA	[S]
D5_1791	HTH and cupin domain-containing protein	[K]
D5_1792	acetyl-CoA synthetase AcsA	[I]
D5_1796	glutamyl-tRNA^{Gln} amidotransferase subunit D GatD	[E]
D5_1797	glutamyl-tRNA^{Gln} amidotransferase subunit E GatE	[J]
D5_1799	ferrous iron transport protein B FeoB	[P]
D5_1809	hypothetical protein	Not in COGs
D5_1810	GMP synthase subunit B GuaAb	[F]
D5_1811	conserved hypothetical protein	[S]
D5_1812	conserved hypothetical protein	[M]
D5_1813	conserved hypothetical protein	[S]
D5_1815	RNA methylase	[L]
D5_1818	HTH domain-containing protein	[K]
D5_1819	conserved hypothetical protein	[S]
D5_1820	conserved hypothetical transmembrane protein	Not in COGs
D5_1821	conserved hypothetical transmembrane protein	Not in COGs
D5_1822	cell wall biosynthesis protein Mur ligase family	[M]
D5_1823	cell wall biosynthesis protein phospho-*N*-acetylmuramoyl-pentapeptide-transferase family	[M]
D5_1824	D-alanine--D-alanine ligase	[F]
D5_1827	nickel responsive transcriptional regulator NikR	[K]
D5_1828	conserved hypothetical protein	[Unclassified]
D5_1843	HEAT repeat-containing protein	[C]
D5_1844	isohomocitrate dehydrogenase AksF	[C]
D5_1847	acetylglutamate kinase ArgB	[E]
D5_1862	bifunctional ornithine acetyltransferase/*N*-acetylglutamate synthase protein ArgJ	[E]
D5_1864	conserved hypothetical transmembrane protein	Not in COGs
D5_1865	conserved hypothetical protein	Not in COGs
D5_1867	nascent polypeptide-associated complex protein	[K]
D5_1883	cation/acetate symporter, actP	[E]
D5_1886	ACT domain-containing protein	[R]
D5_1888	hypothetical protein	Not in COGs
D5_1889	indolepyruvate Fdx oxidoreductase beta subunit IorB	[C]
D5_1890	indolepyruvate Fdx oxidoreductase alpha subunit IorA	[C]
D5_1899	succinate dehydrogenase/fumarate reductase iron-sulfur protein, sdhB	[C]
D5_1900	RNA methyltransferase TrmH family	[J]
D5_1902	deoxycytidine triphosphate deaminase Dcd	[F]
D5_1903	glycyl-tRNA synthetase GlyS	[J]
D5_1905	hydrolase TatD family	[R]
D5_1907	conserved hypothetical protein	[S]
D5_1908	conserved hypothetical transmembrane protein	[S]
D5_1910	ZPR1 zinc-finger domain-containing protein	[R]
D5_1916	uridylate kinase PyrH	[F]
D5_1918	hypothetical transmembrane protein	[S]
D5_1929	glycosyl transferase GT2 family	[M]
D5_1934	DNA primase DnaG	[L]
D5_1938	A_1A_0 archaeal ATP synthase subunit D AhaD	[C]
D5_1939	A_1A_0 archaeal ATP synthase subunit B AhaB	[C]
D5_1940	A_1A_0 archaeal ATP synthase subunit A AhaA	[C]
D5_1941	A_1A_0 archaeal ATP synthase subunit F AhaF	[C]
D5_1942	A_1A_0 archaeal ATP synthase subunit C AhaC	[C]
D5_1943	A_1A_0 archaeal ATP synthase subunit E AhaE	[C]
D5_1944	A_1A_0 archaeal ATP synthase subunit K AhaK	[C]
D5_1945	A_1A_0 archaeal ATP synthase subunit I AhaI	[C]
D5_1946	A_1A_0 archaeal ATP synthase subunit H AhaH	[C]
D5_1951	2-methylcitrate synthase/citrate synthase II PrpC/CitZ	[C]
D5_1952	fumarate hydratase FumA	[C]

D5_1953	conserved hypothetical protein	[Unclassified]
D5_1959	DNA-binding protein	[R]
D5_1960	HTH domain-containing protein	[K]
D5_1962	conserved hypothetical protein	Not in COGs
D5_1964	phosphoglycerate dehydrogenase SerA	[H]
D5_1965	conserved hypothetical protein	[R]
D5_1966	$tRNA^{His}$ guanylyltransferase ThgL	[S]
D5_1967	L-aspartate dehydrogenase	[R]
D5_1969	tRNA binding domain-containing protein	[R]
D5_1974	lactaldehyde dehydrogenase CofA	[C]
D5_1998	Xaa-Pro aminopeptidase	[E]
D5_1999	peptidase M50 family	[R]
D5_2000	methanogenesis marker protein 1	[S]
D5_2001	TfuA-like protein	[S]
D5_2004	CBS domain-containing protein	[R]
D5_2006	amidohydrolase	[F]
D5_2008	carbamoyl-phosphate synthase large subunit CarB	[E]
D5_2009	carbamoyl-phosphate synthase small subunit CarA	[E]
D5_2012	nicotinate-nucleotide pyrophosphorylase NadC	[H]
D5_2013	ribonuclease Z Rnz	[R]
D5_2015	quinolinate synthetase A protein NadA	[H]
D5_2019	homoserine *O*-acetyltransferase MetX3	[E]
D5_2020	conserved hypothetical protein	[L]
D5_2024	HEAT repeat-containing protein	[C]
D5_2026	DEAD/DEAH box helicase domain-containing protein	[R]
D5_2029	5-formaminoimidazole-4-carboxamide-1-β-D- ribofuranosyl 5'-monophosphate-formate ligase PurP	[R]
D5_2030	phosphodiesterase MJ0936 family	[R]
D5_2031	ribosomal RNA large subunit methyltransferase J, rrmJ	[J]
D5_2033	MCM family protein	[L]
D5_2034	translation initiation factor aIF-2 beta subunit	[J]
D5_2035	NMD3 family protein	[J]
D5_2036	tyrosyl-tRNA synthetase TyrS	[J]
D5_2038	thymidylate kinase Tmk	[F]
D5_2041	DNA mismatch repair ATPase MutS family	[L]
D5_2109	transcription initiation factor TFIIB Tfb	[K]
D5_2110	5,10-methylenetetrahydromethanopterin reductase Mer	[C]
D5_2111	radical SAM domain-containing protein	[R]
D5_2112	conserved hypothetical transmembrane protein	[Unclassified]
D5_2116	thiamine biosynthesis protein ThiS	[H]
D5_2117	PP-loop family protein	[D]
D5_2121	dihydropteroate synthase-related protein	[H]
D5_2122	pyruvate Fdx oxidoreductase gamma subunit PorC	[C]
D5_2123	pyruvate Fdx oxidoreductase delta subunit PorD	[C]
D5_2124	pyruvate Fdx oxidoreductase alpha subunit PorA	[C]
D5_2125	pyruvate Fdx oxidoreductase beta subunit PorB	[C]
D5_2126	pyruvate Fdx oxidoreductase-associated	PorE [C]
D5_2128	fumarate hydratase FumA	[C]
D5_2141	conserved hypothetical transmembrane protein	[S]
D5_2142	cobalamin biosynthesis protein CbiM	[P]
D5_2145	cobalt ABC transporter ATP-binding protein CbiO	[P]
D5_2146	hypothetical protein	[Unclassified]
D5_2147	ferrous iron transport protein A FeoA	[P]
D5_2200	translation elongation factor aEF-1 beta	[J]
D5_2202	delta 1-pyrroline-5-carboxylate synthetase	[R]
D5_2203	peptidyl-tRNA hydrolase	[S]
D5_2204	ATPase RIL	[R]
D5_2207	aspartate aminotransferase	[E]
D5_2208	DNA repair and recombination protein RadB	[L]
D5_2211	Sua5/YciO/YrdC/YwlC family translation factor	[J]
D5_2212	phosphatidylglycerophosphate synthase PgsA	[I]
D5_2213	conserved hypothetical protein	[S]
D5_2218	fructose 1,6-bisphosphatase Fbp	[G]
D5_2219	thiamine biosynthesis ATP pyrophosphatase ThiI	[H]
D5_2220	alanyl-tRNA synthetase AlaS	[J]
D5_2221	ribosomal protein L12P Rpl12p	[J]
D5_2222	acidic ribosomal protein P0 RplPO	[J]
D5_2223	ribosomal protein L1P Rpl1p	[J]
D5_2224	ribosomal protein L11P Rpl11p	[J]
D5_2225	ribosomal protein L24 family	[K]
D5_2226	preprotein translocase subunit SecE	[U]
D5_2227	cell division protein FtsZ	[D]
D5_2229	F_{420}-*0*:gamma-glutamyl ligase CofE	[S]
D5_2231	HD domain-containing protein	[R]
D5_2232	RNA-binding protein	[J]
D5_2233	conserved hypothetical protein	[R]
D5_2234	tryptophan-binding regulator TrpY	[R]
D5_2235	hydrogenase assembly chaperone HypC	[O]
D5_2236	glutathione-disulfide reductase Gor	[C]
D5_2237	cell wall biosynthesis protein UDP-glycosyltransferase family	[Unclassified]
D5_2241	prephenate dehydrogenase TyrA	[J]
D5_2242	cell division control protein Cdc48	[O]
D5_2244	methanogenesis marker protein 8	[S]
D5_2246	ATP-dependent DNA ligase DnlI	[L]
D5_2251	phosphoglucosamine mutase GlmM	[G]
D5_2252	TPR repeat-containing protein	[R]
D5_2253	pyruvate formate-lyase-activating enzyme PflA	[O]
D5_2254	histidinol-phosphate aminotransferase HisC	[E]
D5_2255	acetyltransferase	[R]
D5_2256	UDP-*N*-acetylglucosamine diphosphorylase/ glucosamine-1-phosphate *N*-acetyltransferase GlmU	[M]
D5_2257	rubredoxin	[C]
D5_2258	phosphoglucosamine mutase GlmM	[G]
D5_2259	2,3-bisphosphoglycerate-independent phosphoglycerate mutase ApgM	[G]
D5_2260	conserved hypothetical transmembrane protein	[S]
D5_2269	ribosomal protein S3Ae Rps3ae	[J]
D5_2272	dinitrogenase iron-molybdenum cofactor biosynthesis protein	[Unclassified]
D5_2274	methylthioadenosine phosphorylase MtnP	[F]
D5_2275	conserved hypothetical protein	[S]
D5_2277	conserved hypothetical protein	[S]
D5_2279	cdc6 family replication initiation protein Cdc6-2	[L]
D5_2280	6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase MptE	[R]
D5_2285	aminotransferase class V family	[E]
D5_2338	signal recognition particle receptor FtsY	[U]
D5_2339	prefoldin alpha subunit PfdA	[O]
D5_2340	ribosomal protein LX RplX	[J]
D5_2341	translation initiation factor aIF-6	[J]
D5_2342	ribosomal protein L31e Rpl31e	[J]
D5_2343	ribosomal protein L39e Rpl39e	[J]
D5_2345	DNA-binding protein	[R]
D5_2348	ribonuclease P subunit RPR2	[J]
D5_2352	GTP-binding protein	[R]
D5_2361	DNA topoisomerase I TopA	[L]
D5_2363	phosphoserine phosphatase SerB	[S]
D5_2364	TATA-box binding protein Tbp	[K]
D5_2367	adenylate cyclase CyaA	[F]
D5_2368	homocitrate synthase AksA	[E]
D5_2372	conserved hypothetical protein	[H]
D5_2373	flap endonuclease Fen	[L]
D5_2375	*S*-adenosyl-L-homocysteine hydrolase AhcY	[H]

D5_2378	glutamine synthetase GlnA	[E]
D5_2379	conserved hypothetical protein	[S]
D5_2381	conserved hypothetical protein	[S]
D5_2382	tungsten formylmethanofuran dehydrogenase subunit C FwdC	[C]
D5_2383	tungsten formylmethanofuran dehydrogenase subunit A FwdA	[C]
D5_2384	tungsten formylmethanofuran dehydrogenase subunit B FwdB	[C]
D5_2385	tungsten formylmethanofuran dehydrogenase subunit D FwdD	[C]
D5_2386	tungsten formylmethanofuran dehydrogenase subunit G FwdG	[C]
D5_2387	tungsten formylmethanofuran dehydrogenase subunit F FwdF	[C]
D5_2388	tungsten formylmethanofuran dehydrogenase subunit H FwdH	[Unclassified]
D5_2389	molybdopterin-guanine dinucleotide biosynthesis protein B MobB	[H]
D5_2391	formate dehydrogenase beta subunit FdhB	[C]
D5_2393	formate/nitrite transporter FdhC	[P]
D5_2394	formate dehydrogenase accessory protein, FdhD	[C]
D5_2395	3-hexulose-6-phosphate isomerase Phi1	[M]
D5_2396	transcriptional regulator LysR family	[K]
D5_2397	carbohydrate kinase PfkB family	[G]
D5_2398	thiamine biosynthesis protein ThiC	[H]
D5_2399	proteasome-activating nucleotidase	[O]
D5_2400	lysyl-tRNA synthetase LysS	[J]
D5_2402	CMP/dCMP deaminase	[F]
D5_2404	DNA polymerase large subunit DP2 PolD	[L]
D5_2405	anaerobic ribonucleoside-triphosphate reductase NrdD	[F]
D5_2410	ribosomal protein L21e Rpl21e	[J]
D5_2411	DNA-directed RNA polymerase subunit F RpoF	[S]
D5_2412	RNA-binding protein	[J]
D5_2414	SAM-dependent methyltransferase HemK-related	[J]
D5_2416	diaminopimelate epimerase DapF	[E]
D5_2417	diaminopimelate decarboxylase LysA	[E]
D5_2420	acetylornithine aminotransferase ArgD	[E]
D5_2421	peptidyl-prolyl cis-trans isomerase	[O]
D5_2423	MFS transporter	[G]
D5_2424	8-oxoguanine DNA-glycosylase Ogg	[L]
D5_2425	imidazoleglycerol-phosphate synthase cyclase subunit HisF	[E]
D5_2426	preprotein translocase subunit SecG	[U]
D5_2427	molybdenum cofactor biosynthesis protein C MoaC	[H]
D5_2428	pseudouridylate synthase	[J]
D5_2429	signal recognition particle SRP54 protein	[U
D5_2430	adenine phosphoribosyltransferase Apt	[F]
D5_2431	diphthamide biosynthesis protein	[J]
D5_2432	exosome complex RNA-binding protein Csl4	[J]
D5_2433	DNA-directed RNA polymerase subunit L RpoL	[K]
D5_2435	transcription factor S Tfs	[K]
D5_2452	$F_{420}H_2$ oxidase FprA	[C]
D5_2454	peptidase M50 family	[M]
D5_2455	molybdopterin biosynthesis protein MoeA	[H]
D5_2456	phosphoribosylformylglycinamidine (FGAM) synthase II PurL	[F]
D5_2457	isoleucyl-tRNA synthetase IleS	[J]
D5_2463	dihydromethanopterin reductase (acceptor) DmrX	[C]
D5_2464	serine hydroxymethyltransferase GlyA	[E]
D5_2467	CoB--CoM heterodisulfide reductase subunit A HdrA	[C]
D5_2469	DNA repair and recombination protein RadA	[L]
D5_2470	replication factor A	[L]
D5_2471	exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase	[M]
D5_2472	glycosyl transferase GT2 family	[R]
D5_2473	glycosyl transferase GT2 family	[R]
D5_2478	dTDP-glucose 4,6-dehydratase RfbB	[M]
D5_2479	dTDP-4-dehydrorhamnose 3,5-epimerase RfbC	[M]
D5_2480	glucose-1-phosphate thymidylyltransferase RfbA	[M]
D5_2483	dTDP-4-dehydrorhamnose reductase RfbD	[M]
D5_2484	UDP-*N*-acetyl-D-mannosaminuronate dehydrogenase	[M]
D5_2485	UbiD family decarboxylase	[H]
D5_2486	phosphoribosylaminoimidazole carboxylase PurE	[F]
D5_2488	cell wall biosynthesis glycosyl transferase GT2 family	[M]
D5_2489	methanogenesis marker protein 11	[R]
D5_2490	6,7-dimethyl-8-ribityllumazine synthase, RibH	[H]
D5_2494	3-isopropylmalate dehydratase large subunit LeuC	[E]
D5_2498	cobalamin biosynthesis protein CbiX	[S]
D5_2500	conserved hypothetical protein	[R]
D5_2503	divalent cation transporter mgtE family	[P]
D5_2504	TrkA domain-containing protein	[S]
D5_2506	conserved hypothetical protein	[Unclassified]
D5_2507	DNA polymerase small subunit DP1 PolD1	[L]
D5_2508	fuculose 1-phosphate aldolase FucA	[G]
D5_2510	conserved hypothetical protein	[S]

only one gene from D5 for each gene family is represented

Table A.6.6. Predicted *Mbb. ruminantium* $M1^T$ genes conserved within the *Mbb. ruminantium* clade

****

Locus_tag	Predicted gene product	COG category
MRU_RS00630	hypothetical protein	[S]
MRU_RS00655	hypothetical protein	Not in COGs
MRU_RS00920	protein disulfide isomerase	[C]
MRU_RS00945	4Fe-4S Fdx-binding domain-containing protein	[C]
MRU_RS00980	hypothetical protein	Not in COGs
MRU_RS01255	hypothetical protein	[unclassified]
MRU_RS01265	hypothetical protein	Not in COGs
MRU_RS01270	hypothetical protein	[R]
MRU_RS01300	hypothetical protein	[unclassified]
MRU_RS01360	adhesin-like protein	[P]
MRU_RS01675	adhesin-like protein	Not in COGs
MRU_RS01680	adhesin-like protein	Not in COGs
MRU_RS01685	hypothetical protein	Not in COGs
MRU_RS01725	hypothetical protein	Not in COGs
MRU_RS01780	hypothetical protein	[S]
MRU_RS02075	Na^+-dependent transporter SNF family	[R]
MRU_RS02080	hypothetical protein	Not in COGs
MRU_RS02090	hypothetical protein	Not in COGs
MRU_RS02095	hypothetical protein	Not in COGs
MRU_RS02180	hypothetical protein	Not in COGs
MRU_RS02220	7-cyano-7-deazaguanine synthase	[R]
MRU_RS02225	queuosine biosynthesis protein QueD	[H]
MRU_RS02230	7-cyano-7- deazaguanosine biosynthesis protein QueE	[O]
MRU_RS02545	hypothetical protein	Not in COGs
MRU_RS02685	hypothetical protein	[S]
MRU_RS02750	hypothetical protein	[unclassified]
MRU_RS02770	hypothetical protein	[unclassified]
MRU_RS02875	hypothetical protein	[I]
MRU_RS02990	peptidase U32	[O]
MRU_RS03050	hypothetical protein	Not in COGs
MRU_RS03090	hypothetical protein	Not in COGs
MRU_RS03095	hypothetical protein	Not in COGs
MRU_RS03150	mechanosensitive ion channel protein	[M]

MRU_RS03165	hypothetical protein	Not in COGs
MRU_RS03300	transcriptional regulator	[K]
MRU_RS03385	PP-loop family protein	[R]
MRU_RS03490	2-methylcitrate dehydratase	[R]
MRU_RS03520	hypothetical protein	Not in COGs
MRU_RS03600	ATP-grasp domain-containing protein	[R]
MRU_RS03635	hypothetical protein	[unclassified]
MRU_RS03645	hypothetical protein	[unclassified]
MRU_RS03695	RNA-binding protein	[J]
MRU_RS03700	hypothetical protein	Not in COGs
MRU_RS03705	hypothetical protein	Not in COGs
MRU_RS03710	hypothetical protein	Not in COGs
MRU_RS03715	hypothetical protein	Not in COGs
MRU_RS03735	hypothetical protein	Not in COGs
MRU_RS03740	hypothetical protein	Not in COGs
MRU_RS03980	hypothetical protein	[unclassified]
MRU_RS03985	ion transporter	[P]
MRU_RS04070	hypothetical protein	Not in COGs
MRU_RS04075	hypothetical protein	Not in COGs
MRU_RS04080	hypothetical protein	Not in COGs
MRU_RS04110	MarR family transcriptional regulator	[K]
MRU_RS04470	precorrin-2 C20-methyltransferase CbiL	[H]
MRU_RS04475	cobalamin biosynthesis protein CbiD	[H]
MRU_RS04485	cobalamin biosynthesis protein CbiG	[H]
MRU_RS04490	precorrin-3B C17-methyltransferase	[H]
MRU_RS04495	precorrin-6X reductase CbiJ	[H]
MRU_RS04500	precorrin-6Y C5,15-methyltransferase (decarboxylating) CbiET	[H]
MRU_RS04505	cobyrinic acid a,c-diamide synthase	[H]
MRU_RS04510	precorrin-8X methylmutase	[H]
MRU_RS04515	sirohydrochlorin cobaltochelatase	[H]
MRU_RS04540	hypothetical protein	Not in COGs
MRU_RS04555	hypothetical protein	[R]
MRU_RS04705	SAM-dependent methyltransferase	[R]
MRU_RS04720	hypothetical protein	Not in COGs
MRU_RS04760	hypothetical protein	Not in COGs
MRU_RS04875	hypothetical protein	[unclassified]
MRU_RS04900	hypothetical protein	Not in COGs
MRU_RS04910	hypothetical protein	Not in COGs
MRU_RS05140	hypothetical protein	Not in COGs
MRU_RS05165	hypothetical protein	Not in COGs
MRU_RS05220	RNA methylase	[R]
MRU_RS05245	hypothetical protein	Not in COGs
MRU_RS05460	hypothetical protein	Not in COGs
MRU_RS05465	hypothetical protein	[R]
MRU_RS05520	hypothetical protein	Not in COGs
MRU_RS05540	hypothetical protein	[S]
MRU_RS05555	DNA double-strand break repair protein Rad50	[L]
MRU_RS05570	hypothetical protein	[L]
MRU_RS05630	hypothetical protein	Not in COGs
MRU_RS05650	hypothetical protein	Not in COGs
MRU_RS05695	restriction endonuclease	[V]
MRU_RS05700	hypothetical protein	Not in COGs
MRU_RS05740	hypothetical protein	[S]
MRU_RS05800	hypothetical protein	Not in COGs
MRU_RS05900	CRISPR-associated protein TIGR02710 family	[unclassified]
MRU_RS05955	hypothetical protein	Not in COGs
MRU_RS05965	hypothetical protein	[S]
MRU_RS06085	hypothetical protein	Not in COGs
MRU_RS06095	adenosylcobinamide amidohydrolase CbiZ	[S]
MRU_RS06100	hypothetical protein	[unclassified]
MRU_RS06110	cobalt ABC transporter permease	Not in COGs
MRU_RS06140	hypothetical protein	Not in COGs
MRU_RS06155	hypothetical protein	Not in COGs
MRU_RS06165	hypothetical protein	[unclassified]
MRU_RS06170	hypothetical protein	[unclassified]
MRU_RS06185	hypothetical protein	[unclassified]
MRU_RS06220	hypothetical protein	[unclassified]
MRU_RS06390	molybdopterin-guanine dinucleotide biosynthesis protein A MobA2	[H]
MRU_RS06590	hypothetical protein	Not in COGs
MRU_RS06735	hypothetical protein	[unclassified]
MRU_RS06790	adhesin-like protein	[unclassified]
MRU_RS06925	hypothetical protein	Not in COGs
MRU_RS07080	adhesin-like protein	[unclassified]
MRU_RS07145	NAD synthetase	[H]
MRU_RS07295	hypothetical protein	[unclassified]
MRU_RS07345	hypothetical protein	[R]
MRU_RS07465	hypothetical protein	[S]
MRU_RS07495	hypothetical protein	[E]
MRU_RS07515	hypothetical protein	Not in COGs
MRU_RS07540	hydrolase alpha/beta fold family	[R]
MRU_RS07545	pyrroline-5-carboxylate reductase ProC	[E]
MRU_RS07675	hypothetical protein	Not in COGs
MRU_RS07715	hypothetical protein	[M]
MRU_RS07815	hypothetical protein	[unclassified]
MRU_RS07890	6-*O*-methylguanine DNA methyltransferase Ogt	[L]
MRU_RS07965	hypothetical protein	Not in COGs
MRU_RS07990	hypothetical protein	Not in COGs
MRU_RS08030	adhesin-like protein with transglutaminase domain	[E]
MRU_RS08035	hypothetical protein	[unclassified]
MRU_RS08285	hypothetical protein	[unclassified]
MRU_RS08385	peptidase S49	[O]
MRU_RS08410	hypothetical protein	[S]
MRU_RS08490	hypothetical protein	[R]
MRU_RS08530	NADH pyrophosphatase NudC	[L]
MRU_RS08655	heat-shock protein Hsp20/alpha crystallin family	[O]
MRU_RS08725	cell wall biosynthesis protein Mur ligase family	[M]
MRU_RS08760	hypothetical protein	Not in COGs
MRU_RS08810	ArsR family transcriptional regulator	[K]
MRU_RS08815	hypothetical protein	Not in COGs
MRU_RS08970	hypothetical protein	Not in COGs
MRU_RS08975	hypothetical protein	Not in COGs
MRU_RS08985	hypothetical protein	Not in COGs
MRU_RS09185	hypothetical protein	[unclassified]
MRU_RS09225	hypothetical protein	Not in COGs
MRU_RS09370	hypothetical protein	Not in COGs
MRU_RS09420	hypothetical protein	[unclassified]
MRU_RS09430	hypothetical protein	[unclassified]
MRU_RS09455	hypothetical protein	Not in COGs
MRU_RS09510	hypothetical protein	[unclassified]
MRU_RS09740	hypothetical protein	Not in COGs
MRU_RS09875	hypothetical protein	[S]
MRU_RS09920	hypothetical protein	Not in COGs
MRU_RS09965	hypothetical protein	[unclassified]
MRU_RS10040	energy-converting hydrogenase B subunit J EhbJ	[unclassified]
MRU_RS10055	energy-converting hydrogenase B subunit G EhbG	[unclassified]
MRU_RS10090	hypothetical protein	[R]
MRU_RS10115	hypothetical protein	[P]
MRU_RS10255	6-carboxyhexanoate--CoA ligase	[H]
MRU_RS10260	8-amino-7-oxononanoate synthase	[H]
MRU_RS10315	adhesin-like protein	Not in COGs
MRU_RS10380	hypothetical protein	Not in COGs
MRU_RS10410	hypothetical protein	[S]
MRU_RS10425	formate dehydrogenase alpha chain FdhA2	[C]
MRU_RS10475	adenosylmethionine--8-amino-7-oxononanoate aminotrans ferase BioA	[H]

MRU_RS10485	dithiobiotin synthetase	[H]
MRU_RS10635	hypothetical protein	Not in COGs
MRU_RS10645	hypothetical protein	[R]
MRU_RS10730	hypothetical protein	Not in COGs
MRU_RS10875	NADH oxidase	[C]
MRU_RS11055	hypothetical protein	[unclassified]
MRU_RS11060	hypothetical protein	[unclassified]
MRU_RS11095	hypothetical protein	[H]
MRU_RS11185	hypothetical protein	[unclassified]
MRU_RS11200	asparagine synthase (glutamine-hydrolyzing) AsnB	[E]
MRU_RS11210	hypothetical protein	Not in COGs
MRU_RS11245	hypothetical protein	[L]
MRU_RS11250	hypothetical protein	[R]
MRU_RS11260	type II restriction endonuclease	[unclassified]

only one gene from M1^T for each gene family is represented

Table A.6.7. Predicted *Methanobrevibacter* sp. AbM4 genes conserved within the *Mbb. wolinii* clade

****

Locus_tag	**Predicted gene product**	**COG category**
2540853212	flavodoxin domain-containing protein	[C]
2540853237	PQQ repeat-containing cell surface protein	[S]
2540853282	Protein of unknown function (DUF4013)	[unclassified]
2540853310	ABC transporter permease protein	[V]
2540853362	hypothetical protein	[T]
2540853369	glutamine amidotransferase class II	[R]
2540853371	hypothetical protein	[unclassified]
2540853438	OB fold nucleic acid binding domain-containing protein	[R]
2540853473	NADPH-dependent FMN reductase [R]	
2540853489	transposase	Not in COGs
2540853495	hypothetical protein	Not in COGs
2540853502	adhesin-like protein	[unclassified]
2540853507	hypothetical protein	[unclassified]
2540853509	dolichyl-phosphate-mannose-protein mannosyltransferase	[unclassified]
2540853511	dolichyl-phosphate-mannose-protein mannosyltransferase	[unclassified]
2540853512	hypothetical protein	Not in COGs
2540853516	hexapeptide repeat-containing acetyltransferase	[R]
2540853526	methanogenesis marker protein 5	[S]
2540853571	aspartate racemase Asp	[M]
2540853577	hypothetical protein	Not in COGs
2540853578	hypothetical protein	[D]
2540853579	hypothetical protein	Not in COGs
2540853582	adhesin-like protein	Not in COGs
2540853597	hypothetical protein	Not in COGs
2540853602	hypothetical protein	[unclassified]
2540853610	MFS transporter	[P]
2540853623	prpD 2-methylcitrate dehydratase PrpD	[R]
2540853628	hypothetical protein	Not in COGs
2540853629	hypothetical protein	Not in COGs
2540853639	hypothetical protein	Not in COGs
2540853642	hypothetical protein	Not in COGs
2540853646	uncharacterized conserved protein	[S]
2540853654	hypothetical protein	[unclassified]
2540853661	putative heavy-metal-binding	[M]
2540853686	transporter	[R]
2540853689	heat shock protein Hsp20	[O]
2540853693	hypothetical protein	[unclassified]
2540853707	predicted ATP-utilizing enzyme (ATP-grasp superfamily)	[R]
2540853721	hsdS2 type I restriction-modification system S subunit HsdS2	[V]
2540853723	hsdS3 type I restriction-modification system S subunit HsdS3	[V]
2540853731	deoxyribose-phosphate aldolase DeoC	[F]
2540853735	Short repeat of unknown function (DUF308)	[unclassified]
2540853741	hypothetical protein	Not in COGs
2540853753	hypothetical protein	[unclassified]
2540853769	transcriptional regulator, TetR family	[K]
2540853772	two-component system histidine kinase	[T]
2540853777	adhesin-like protein	[R]
2540853780	adhesin-like protein with PMBR domains	[unclassified]
2540853781	zinc-ribbon domain	[S]
2540853784	uncharacterized protein conserved in bacteria	Not in COGs
2540853787	hypothetical protein	Not in COGs
2540853793	adhesin-like protein	[unclassified]
2540853810	hypothetical protein	Not in COGs
2540853818	hypothetical protein	[unclassified]
2540853827	hypothetical protein	Not in COGs
2540853833	4Fe-4S Fdx iron-sulfur binding domain-containing protein	[C]
2540853834	hypothetical protein	Not in COGs
2540853835	TauE family transporter	[R]
2540853841	SAM-dependent methyltransferase	[R]
2540853853	adhesin-like protein	[unclassified]
2540853854	adhesin-like protein	Not in COGs
2540853855	adhesin-like protein	[unclassified]
2540853868	ATPase	[R]
2540853871	TPR domain-containing protein	[R]
2540853873	TPR domain-containing protein	[R]
2540853874	TPR domain-containing protein	[R]
2540853875	TPR domain-containing protein	[R]
2540853876	TPR domain-containing protein	[R]
2540853882	hypothetical protein	Not in COGs
2540853885	flavoprotein HI0933 family	[R]
2540853891	pyrroline-5-carboxylate reductase ProC	[E]
2540853898	acetyl esterase	[R]
2540853942	hypothetical protein	Not in COGs
2540853945	hypothetical protein	Not in COGs
2540853947	hypothetical protein	Not in COGs
2540853956	adenosylcobinamide amidohydrolase CbiZ	[H]
2540853959	uncharacterized conserved protein	[S]
2540853966	hypothetical protein	[unclassified]
2540853992	hypothetical protein	[S]
2540854005	double zinc ribbon	[S]
2540854007	asnB asparagine synthase	[E]
2540854015	oxidoreductase aldo/keto reductase family	[R]
2540854017	adhesin-like protein	[unclassified]
2540854030	ogt 6-*O*-methylguanine DNA methyltransferase Ogt	[L]
2540854031	hypothetical protein	Not in COGs
2540854033	hypothetical protein	Not in COGs
2540854043	predicted integral membrane protein	[S]
2540854044	predicted integral membrane protein	[S]
2540854045	adhesin-like protein	[unclassified]
2540854049	adhesin-like protein	[unclassified]
2540854051	MFS transporter	[P]
2540854052	transcriptional regulator, TetR family	[K]

2540854058	hypothetical protein	Not in COGs
2540854064	adhesin-like protein with transglutaminase and PMBR domains	[E]
2540854067	hypothetical protein	Not in COGs
2540854068	hypothetical protein	Not in COGs
2540854070	cell wall biosynthesis protein Mur ligase family	[M]
2540854076	hypothetical protein	Not in COGs
2540854077	DNA double-strand break repair protein Rad50	[L]
2540854080	NurA domain-containing protein	[S]
2540854084	hypothetical protein	Not in COGs
2540854101	carbohydrate kinase PfkB family	[G]
2540854110	Protein of unknown function (DUF3194)	[unclassified]
2540854128	hypothetical protein	Not in COGs
2540854135	energy-convertinghydrogenase A subunit R EhaR	[S]
2540854154	hypothetical protein	[unclassified]
2540854163	hypothetical protein	[unclassified]
2540854188	hypothetical protein	Not in COGs
2540854191	uncharacterized protein conserved in archaea	[S]
2540854209	hypothetical protein	[M]
2540854226	hypothetical protein	Not in COGs
2540854228	hydrolase HAD superfamily	[R]
2540854231	hypothetical protein	Not in COGs
2540854234	Domain of unknown function (DUF955)	[E]
2540854239	hydrolase HAD superfamily	[R]
2540854243	hydrolase HAD superfamily	[R]
2540854244	hydrolase HAD superfamily	[R]
2540854247	hypothetical protein	Not in COGs
2540854249	hypothetical protein	Not in COGs
2540854262	hypothetical protein	Not in COGs
2540854271	hypothetical protein	Not in COGs
2540854323	uncharacterized conserved protein (DUF2304)	Not in COGs
2540854368	short repeat of unknown function (DUF308)	[S]
2540854383	hypothetical protein	Not in COGs
2540854396	adhesin-like protein	[P]
2540854413	hypothetical protein	Not in COGs
2540854414	hypothetical protein	Not in COGs
2540854482	hypothetical protein	[unclassified]
2540854485	hypothetical protein	[unclassified]
2540854509	polysaccharide biosynthesis protein	[R]
2540854538	CAAX amino terminal protease family protein	[unclassified]
2540854548	SIR2 family protein	[K]
2540854560	mechanosensitive ion channel protein	[P]
2540854565	zinc-ribbon domain	[O]
2540854574	uncharacterized protein conserved in archaea	[S]
2540854575	flavodoxin	[C]
2540854576	hypothetical protein	[R]
2540854593	peptidase U32 family	[O]
2540854595	hypothetical protein	[L]
2540854597	trpD anthranilate phosphoribosyltransferase	[E]
2540854598	trpA tryptophan synthase, alpha chain	[E]
2540854600	trpF phosphoribosylanthranilate isomerase	[E]
2540854601	trpC indole-3-glycerol phosphate synthase	[E]
2540854603	trpE anthranilate synthase, component I"	[E]
2540854604	hypothetical protein	[unclassified]
2540854615	NADP oxidoreductase coenzyme F_{420}-dependent	[I]
2540854625	ppx exopolyphosphatase Ppx	[F]
2540854633	phosphate uptake regulator PhoU1	[P]
2540854639	hypothetical protein	[unclassified]
2540854643	hypothetical protein	[unclassified]
2540854649	4Fe-4S Fdx iron-sulfur binding domain-containing protein	[C]
2540854660	CAAX amino terminal protease family protein	[R]
2540854722	metallophosphoesterase	[R]
2540854729	hypothetical protein	Not in COGs
2540854730	metallo-beta-lactamase superfamily protein	[R]
2540854754	hypothetical protein	[unclassified]
2540854755	hypothetical protein	[unclassified]
2540854763	CRISPR-associated protein Cas4	[L]
2540854787	energy-converting hydrogenase B subunit J EhbJ	[unclassified]
2540854802	acetyltransferase GNAT family	[R]
2540854809	EamA-like transporter family protein	[P]
2540854810	EamA-like transporter family protein	[P]
2540854812	hypothetical protein	[unclassified]
2540854893	predicted membrane protein	[S]
2540854907	predicted membrane protein	[S]
2540854915	zinc-ribbon domai	[T]
2540854923	cobalamin biosynthesis protein CbiG	[H]
2540854924	cobalamin biosynthesis protein CbiB	[H]

only one gene from AbM4 for each gene family is represented

Table A.6.8. Predicted *Methanobrevibacter* sp. D5 genes unique to the *Methanobrevibacter* sp. D5 genome

****

Locus_tag	Predicted gene product	COG category
D5_0009	hypothetical transmembrane protein	[S]
D5_0020	hypothetical protein	Not in COGs
D5_0034	hypothetical secreted protein	[unclassified]
D5_0035	hypothetical transmembrane protein	Not in COGs
D5_0036	adhesin-like protein	[unclassified]
D5_0038	adhesin-like protein	[unclassified]
D5_0051	adhesin-like protein	[unclassified]
D5_0060	hypothetical protein	Not in COGs
D5_0073	hypothetical protein	[R]
D5_0082	hypothetical protein	[unclassified]
D5_0083	hypothetical protein	[unclassified]
D5_0093	pyridoxamine 5'-phosphate oxidase family protein	[S]
D5_0110	hypothetical protein	Not in COGs
D5_0111	adhesin-like protein	[unclassified]
D5_0115	NADH oxidase Nox	[R]
D5_0123	hypothetical protein	[unclassified]
D5_0124	hypothetical protein	[unclassified]
D5_0125	hypothetical protein	Not in COGs
D5_0181	hypothetical protein	[unclassified]
D5_0198	hypothetical protein	Not in COGs
D5_0211	conserved hypothetical secreted protein	[unclassified]
D5_0218	adhesin-like protein	[unclassified]
D5_0219	adhesin-like protein	[unclassified]
D5_0223	glycosyl transferase GT2 family	[M]
D5_0262	adhesin-like protein	Not in COGs
D5_0306	hypothetical protein	Not in COGs
D5_0340	hydrolase alpha/beta fold family	[R]
D5_0344	aldo/keto reductase family protein	[R]
D5_0345	aldo/keto reductase family protein	[R]
D5_0346	aldo/keto reductase family protein	[R]
D5_0354	hypothetical protein	[S]
D5_0355	hypothetical protein	Not in COGs
D5_0378	conserved hypothetical secreted protein	Not in COGs
D5_0387	hypothetical protein	[R]

D5_0392	adhesin-like protein with cysteine protease domain	Not in COGs
D5_0393	adhesin-like protein with cysteine protease domain	[E]
D5_0394	adhesin-like protein	[E]
D5_0431	hypothetical protein	[J]
D5_0432	hypothetical transmembrane protein	[G]
D5_0433	hypothetical protein	Not in COGs
D5_0437	hypothetical protein	Not in COGs
D5_0438	hypothetical protein	Not in COGs
D5_0443	transposase	[L]
D5_0445	hypothetical transmembrane protein	Not in COGs
D5_0449	hypothetical transmembrane protein	[unclassified]
D5_0450	hypothetical transmembrane protein	Not in COGs
D5_0451	hypothetical protein	Not in COGs
D5_0452	adhesin-like protein	[unclassified]
D5_0454	hypothetical protein	Not in COGs
D5_0455	hypothetical protein	Not in COGs
D5_0456	hypothetical protein	Not in COGs
D5_0457	hypothetical transmembrane protein	[S]
D5_0458	adhesin-like protein	Not in COGs
D5_0459	adhesin-like protein	[M]
D5_0460	conserved hypothetical transmembrane protein	[S]
D5_0462	glycosyl transferase GT2 family	[M]
D5_0463	glycosyl transferase	[M]
D5_0476	hypothetical protein	[L]
D5_0478	adhesin-like protein	Not in COGs
D5_0480	hypothetical protein	[L]
D5_0482	non-ribosomal surfactin synthetase SrfAA	[Q]
D5_0483	hypothetical protein	Not in COGs
D5_0484	glycosyl transferase GT2 family	[M]
D5_0486	hypothetical protein	Not in COGs
D5_0487	hypothetical protein	[unclassified]
D5_0488	glycosyl transferase GT2 family	[M]
D5_0491	hypothetical protein	Not in COGs
D5_0492	hypothetical protein	[unclassified]
D5_0494	hypothetical protein	[M]
D5_0495	hypothetical protein	[unclassified]
D5_0496	hypothetical protein	Not in COGs
D5_0497	transposase	[unclassified]
D5_0499	acetyltransferase	[R]
D5_0502	hypothetical transmembrane protein	Not in COGs
D5_0507	phosphatase PAP2 family	[I]
D5_0512	type I restriction-modification enzyme S subunit HsdS	[V]
D5_0513	type I restriction-modification system M subunit, HsdM	[V]
D5_0514	hypothetical protein	[unclassified]
D5_0528	conserved hypothetical secreted protein	Not in COGs
D5_0529	hypothetical protein	Not in COGs
D5_0530	hypothetical transmembrane protein	Not in COGs
D5_0541	hypothetical protein	Not in COGs
D5_0605	conserved hypothetical protein	[R]
D5_0613	hypothetical protein	[E]
D5_0618	exodeoxyribonuclease VII large subunit, XseA	[L]
D5_0637	phosphodiesterase, MJ0936 family	[R]
D5_0640	adhesin-like protein	Not in COGs
D5_0645	conserved hypothetical transmembrane protein	Not in COGs
D5_0671	conserved hypothetical protein	[L]
D5_0715	adhesin-like protein with cysteine protease domain	[unclassified]
D5_0716	conserved hypothetical transmembrane protein	Not in COGs
D5_0756	ATPase	Not in COGs
D5_0767	acetyltransferase GNAT family	[K]
D5_0788	hypothetical protein	[K]
D5_0793	MATE efflux family protein	[V]
D5_0798	nitroreductase family protein	[C]
D5_0821	adhesin-like protein	[M]
D5_0826	hypothetical protein	[unclassified]
D5_0827	hypothetical protein	Not in COGs
D5_0828	ROK family protein	[K]
D5_0841	sialic acid synthase	[M]
D5_0842	UDP-*N*-acetylglucosamine 2-epimerase	[M]
D5_0843	oxidoreductase domain-containing protein	[R]
D5_0844	oxidoreductase GFO/IDH/MOCA family	[R]
D5_0845	cytidyltransferase-related domain-containing protein	[M]
D5_0846	hypothetical protein	[R]
D5_0847	aminotransferase DegT/DnrJ/EryC1/StrS family	[M]
D5_0849	glycosyl transferase	[unclassified]
D5_0850	hypothetical protein	[unclassified]
D5_0857	hypothetical protein	Not in COGs
D5_0858	hypothetical protein	Not in COGs
D5_0861	hypothetical protein	Not in COGs
D5_0869	hypothetical protein	Not in COGs
D5_0876	hypothetical protein	[unclassified]
D5_0884	polysaccharide biosynthesis protein	[R]
D5_0885	hypothetical protein	Not in COGs
D5_0886	conserved hypothetical protein	[M]
D5_0888	hypothetical protein	Not in COGs
D5_0890	conserved hypothetical protein	[unclassified]
D5_0891	conserved hypothetical protein	[unclassified]
D5_0895	hypothetical protein	[C]
D5_0897	hypothetical protein	[T]
D5_0898	hypothetical protein	[M]
D5_0900	hypothetical protein	Not in COGs
D5_0901	hypothetical protein	[L]
D5_0902	hypothetical protein	Not in COGs
D5_0905	hypothetical protein	[M]
D5_0907	hypothetical protein	[T]
D5_0910	hypothetical protein	[O]
D5_0913	hypothetical protein	Not in COGs
D5_0917	hypothetical protein	Not in COGs
D5_0918	hypothetical protein	Not in COGs
D5_0922	hypothetical protein	Not in COGs
D5_0925	hypothetical protein	Not in COGs
D5_0926	hypothetical protein	[unclassified]
D5_0927	hypothetical protein	Not in COGs
D5_0928	hypothetical protein	[M]
D5_0962	acetyltransferase GNAT family	[K]
D5_0963	hypothetical protein	Not in COGs
D5_0965	adhesin-like protein	Not in COGs
D5_0971	adhesin-like protein	[M]
D5_0973	adhesin-like protein	[unclassified]
D5_0974	adhesin-like protein with cysteine protease domain	[O]
D5_0976	adhesin-like protein	[unclassified]
D5_0978	adhesin-like protein	[unclassified]
D5_0979	adhesin-like protein	[unclassified]
D5_1015	hypothetical transmembrane protein	[S]
D5_1016	CAAX amino terminal protease family protein	[R]
D5_1020	hypothetical protein	Not in COGs
D5_1022	hypothetical protein	Not in COGs
D5_1023	TPR repeat-containing protein	[R]
D5_1024	hypothetical protein	Not in COGs
D5_1035	hypothetical protein	Not in COGs
D5_1038	hypothetical protein	Not in COGs
D5_1040	phosphatidylglycerophosphate synthase, PgsA	[I]
D5_1041	hypothetical transmembrane protein	Not in COGs
D5_1042	succinate dehydrogenase/fumarate reductase flavoprotein subunit, SdhA2	[C]
D5_1044	succinate dehydrogenase/fumarate reductase iron-sulfur protein, SdhB1	[C]
D5_1049	hypothetical transmembrane protein	[G]
D5_1054	short-chain dehydrogenase family protein	[I]
D5_1063	hypothetical protein	[unclassified]
D5_1066	hypothetical transmembrane protein	[V]
D5_1071	adhesin-like protein	[R]
D5_1087	hypothetical transmembrane protein	[unclassified]
D5_1101	hypothetical transmembrane protein	Not in COGs
D5_1105	hypothetical protein	[unclassified]
D5_1106	hypothetical protein	Not in COGs
D5_1107	hypothetical protein	Not in COGs

D5_1108	hypothetical protein	Not in COGs
D5_1109	hypothetical protein	Not in COGs
D5_1110	hypothetical protein	Not in COGs
D5_1112	hypothetical protein	Not in COGs
D5_1113	hypothetical protein	Not in COGs
D5_1118	hypothetical protein	Not in COGs
D5_1119	hypothetical transmembrane protein	[unclassified]
D5_1132	hypothetical transmembrane protein	[unclassified]
D5_1147	hypothetical protein	Not in COGs
D5_1151	hypothetical protein	[H]
D5_1154	hypothetical protein	[C]
D5_1157	WD40 repeat-containing protein	Not in COGs
D5_1158	hypothetical protein	Not in COGs
D5_1159	hypothetical protein	Not in COGs
D5_1160	hypothetical protein	Not in COGs
D5_1162	hypothetical protein	[R]
D5_1165	ADP-ribosylglycohydrolase family protein	[O]
D5_1166	hypothetical protein	Not in COGs
D5_1167	hypothetical protein	Not in COGs
D5_1168	hypothetical protein mru	Not in COGs
D5_1169	conserved hypothetical protein	Not in COGs
D5_1174	hypothetical protein	Not in COGs
D5_1178	hypothetical protein	Not in COGs
D5_1180	hypothetical protein	[I]
D5_1182	hypothetical protein	Not in COGs
D5_1183	conserved hypothetical protein	[S]
D5_1185	conserved hypothetical protein	Not in COGs
D5_1187	hypothetical protein	Not in COGs
D5_1188	hypothetical protein	Not in COGs
D5_1190	hypothetical protein	Not in COGs
D5_1192	hypothetical protein	Not in COGs
D5_1193	hypothetical protein	Not in COGs
D5_1194	hypothetical protein	Not in COGs
D5_1195	hypothetical protein	[C]
D5_1197	hypothetical protein	[S]
D5_1200	hypothetical protein	[unclassified]
D5_1207	ion transport protein	[P]
D5_1212	hypothetical protein	Not in COGs
D5_1215	hypothetical protein	Not in COGs
D5_1216	conserved hypothetical protein	[L]
D5_1217	hypothetical protein	Not in COGs
D5_1218	ATPase	[V]
D5_1219	hypothetical protein	Not in COGs
D5_1222	DNA-cytosine methyltransferase, Dcm	[L]
D5_1223	hypothetical protein	Not in COGs
D5_1224	hypothetical protein	[D]
D5_1226	hypothetical protein	Not in COGs
D5_1227	hypothetical protein	Not in COGs
D5_1229	hypothetical protei	Not in COGs
D5_1230	hypothetical transmembrane protein	Not in COGs
D5_1233	hypothetical protein	Not in COGs
D5_1235	exodeoxyribonuclease VII large subunit, XseA	[L]
D5_1237	hypothetical protein	Not in COGs
D5_1238	hypothetical protein	[D]
D5_1239	hypothetical protein	Not in COGs
D5_1240	ATP-dependent DNA helicase UvrD/REP family	Not in COGs
D5_1244	TPR domain-containing protein	[R]
D5_1245	hypothetical protein	Not in COGs
D5_1246	hypothetical protein	Not in COGs
D5_1247	ATP-dependent DNA helicase UvrD/REP family	[L]
D5_1249	hypothetical transmembrane protein	Not in COGs
D5_1250	hypothetical protein	Not in COGs
D5_1251	hypothetical protein	Not in COGs
D5_1252	hypothetical protein	Not in COGs
D5_1253	hypothetical protein	Not in COGs
D5_1254	hypothetical protein	Not in COGs
D5_1256	hypothetical protein	Not in COGs
D5_1259	hypothetical protein	[D]
D5_1260	hypothetical protein	[S]
D5_1262	hypothetical protein	[S]
D5_1264	hypothetical protein	Not in COGs
D5_1266	hypothetical protein	Not in COGs
D5_1268	hypothetical transmembrane protein	Not in COGs
D5_1270	TPR repeat-containing protein	[R]
D5_1274	hypothetical protein	[unclassified]
D5_1280	hypothetical protein	Not in COGs
D5_1282	hypothetical protein	[O]
D5_1285	hypothetical transmembrane protein	[S]
D5_1289	hypothetical protein	[S]
D5_1297	glycosyl transferase GT2 family	Not in COGs
D5_1303	hypothetical protein	[E]
D5_1305	LPS biosynthesis protein	[M]
D5_1306	glycosyl transferase GT2 family	Not in COGs
D5_1314	hypothetical protein	[M]
D5_1322	hypothetical protein	Not in COGs
D5_1337	hypothetical transmembrane protein	[unclassified]
D5_1343	CAAX amino terminal protease family protein	[S]
D5_1349	ACT domain-containing protein	[S]
D5_1366	DNA mismatch repair ATPase MutS family	[unclassified]
D5_1390	hypothetical protein	Not in COGs
D5_1393	hypothetical transmembrane protein	[O]
D5_1394	hypothetical transmembrane protein	[O]
D5_1406	adhesin-like protein	Not in COGs
D5_1410	hypothetical transmembrane protein	[E]
D5_1412	adhesin-like protein	[unclassified]
D5_1413	adhesin-like protein	[unclassified]
D5_1416	adhesin-like protein	[O]
D5_1417	hypothetical transmembrane protein	[unclassified]
D5_1423	adhesin-like protein	[unclassified]
D5_1424	adhesin-like protein	[unclassified]
D5_1425	chaperone protein DnaJ, *dnaJ*	[O]
D5_1428	hypothetical protein	Not in COGs
D5_1429	hypothetical protein	Not in COGs
D5_1430	hypothetical protein	Not in COGs
D5_1431	hypothetical protein	Not in COGs
D5_1432	hypothetical protein	Not in COGs
D5_1433	hypothetical protein	[unclassified]
D5_1434	hypothetical protein	Not in COGs
D5_1435	hypothetical protein	Not in COGs
D5_1436	hypothetical protein	Not in COGs
D5_1437	hypothetical protein	Not in COGs
D5_1438	hypothetical protein	Not in COGs
D5_1439	hypothetical protein	Not in COGs
D5_1440	hypothetical protein	Not in COGs
D5_1441	hypothetical protein	Not in COGs
D5_1442	hypothetical protein	Not in COGs
D5_1443	hypothetical protein	Not in COGs
D5_1444	hypothetical protein	Not in COGs
D5_1445	hypothetical protein	Not in COGs
D5_1446	hypothetical protein	Not in COGs
D5_1447	hypothetical protein	Not in COGs
D5_1448	hypothetical protein	Not in COGs
D5_1449	hypothetical protein	Not in COGs
D5_1450	hypothetical protein	Not in COGs
D5_1451	hypothetical protein	[R]
D5_1452	hypothetical protein	Not in COGs
D5_1453	hypothetical transmembrane protein	[unclassified]
D5_1454	hypothetical protein	Not in COGs
D5_1455	hypothetical protein	[unclassified]
D5_1456	hypothetical protein	Not in COGs
D5_1457	hypothetical protein	[L]
D5_1458	hypothetical protein	Not in COGs
D5_1459	hypothetical protein	Not in COGs
D5_1463	hypothetical protein	Not in COGs
D5_1464	hypothetical transmembrane protein	[M]
D5_1473	adhesin-like protein	[unclassified]
D5_1476	adhesin-like protein	Not in COGs
D5_1480	hypothetical protein	Not in COGs
D5_1481	CAAX amino terminal protease family protein	[R]
D5_1486	hypothetical protein	[R]
D5_1491	hypothetical transmembrane protein	[unclassified]
D5_1494	adhesin-like protein	Not in COGs
D5_1515	WD40 repeat-containing protein	Not in COGs
D5_1554	hypothetical protein	Not in COGs
D5_1582	hypothetical protein	Not in COGs

D5_1603	hypothetical transmembrane protein	[R]
D5_1604	hypothetical transmembrane protein	Not in COGs
D5_1605	hypothetical transmembrane protein	[S]
D5_1606	hypothetical transmembrane protein	[O]
D5_1618	adhesin-like protein	[unclassified]
D5_1619	adhesin-like protein	[unclassified]
D5_1625	hypothetical protein	Not in COGs
D5_1627	hypothetical protein	Not in COGs
D5_1628	hypothetical protein	[R]
D5_1629	hypothetical protein	Not in COGs
D5_1630	hypothetical protein	Not in COGs
D5_1631	hypothetical protein	Not in COGs
D5_1632	hypothetical protein	[M]
D5_1633	hypothetical transmembrane protein	Not in COGs
D5_1634	hypothetical protein	Not in COGs
D5_1637	hypothetical protein	Not in COGs
D5_1638	hypothetical transmembrane protein	Not in COGs
D5_1707	hypothetical transmembrane protein	Not in COGs
D5_1711	hypothetical protein	[unclassified]
D5_1735	conserved hypothetical protein	[T]
D5_1736	repeat domain containing protein	[M]
D5_1737	hypothetical protein	Not in COGs
D5_1738	hypothetical transmembrane protein	[S]
D5_1742	nitroreductase family protein	[C]
D5_1744	hypothetical	[unclassified]
D5_1745	hypothetical transmembrane protein	Not in COGs
D5_1764	adhesin-like protein	Not in COGs
D5_1800	hypothetical protein	Not in COGs
D5_1801	hypothetical transmembrane protein	Not in COGs
D5_1802	hypothetical protein	Not in COGs
D5_1833	hypothetical protein	[K]
D5_1837	hypothetical protein	Not in COGs
D5_1838	conserved hypothetical transmembrane protein	[S]
D5_1841	5,10-methenyltetrahydromethanopterin hydrogenase Hmd	[C]
D5_1879	TPR repeat-containing protein	[R]
D5_1884	hypothetical transmembrane protein	[unclassified]
D5_1892	hypothetical transmembrane protein	Not in COGs
D5_1896	adhesin-like protein	[unclassified]
D5_1912	adhesin-like protein	Not in COGs
D5_1930	hypothetical protein	Not in COGs
D5_1970	hypothetical protein	Not in COGs
D5_1971	hypothetical protein	Not in COGs
D5_1972	hypothetical protein	Not in COGs
D5_1973	hypothetical transmembrane protein	[S]
D5_1981	conserved hypothetical	Not in COGs
D5_1982	hypothetical transmembrane protein	Not in COGs
D5_1983	hypothetical transmembrane protein	[S]
D5_1985	hypothetical transmembrane protein	Not in COGs
D5_1988	hypothetical transmembrane protein	Not in COGs
D5_1996	homoserine *O*-acetyltransferase, MetX	[E]
D5_2016	short chain dehydrogenase	[I]
D5_2025	hypothetical transmembrane protein	Not in COGs
D5_2046	hypothetical protein	[E]
D5_2050	hypothetical protein	Not in COGs
D5_2055	hypothetical protein	Not in COGs
D5_2056	hypothetical protein	[unclassified]
D5_2057	hypothetical protein	[D]
D5_2059	conserved hypothetical protein	[S]
D5_2063	hypothetical protein	Not in COGs
D5_2065	hypothetical protein	Not in COGs
D5_2074	hypothetical protein	Not in COGs
D5_2078	restriction endonuclease	[V]
D5_2082	DNA repair photolyase	[L]
D5_2083	type I site-specific deoxyribonuclease HsdR family, HsdR	[V]
D5_2084	type I restriction endonuclease subunit S, HdsS	[V]
D5_2085	type I restriction endonuclease subunit M, HsdM	[V]
D5_2086	GIY-YIG catalytic domain-containing endonuclease	[unclassified]
D5_2093	hypothetical transmembrane protein	Not in COGs
D5_2099	hypothetical transmembrane protein	Not in COGs
D5_2120	adhesin-like protein	Not in COGs
D5_2164	type III restriction endonuclease res subunit	[unclassified]
D5_2165	type III restriction endonuclease methylation subunit	[L]
D5_2166	hypothetical protein	Not in COGs
D5_2171	hypothetical transmembrane protein	Not in COGs
D5_2173	adhesin-like protein	[M]
D5_2174	hypothetical transmembrane protein	[M]
D5_2175	adhesin-like protein	[unclassified]
D5_2176	adhesin-like protein	Not in COGs
D5_2193	hypothetical protein	[unclassified]
D5_2194	hypothetical transmembrane protein	Not in COGs
D5_2209	hypothetical transmembrane protein	[S]
D5_2210	hypothetical protein	[R]
D5_2247	metallo-beta-lactamase superfamily protein	[R]
D5_2248	hypothetical transmembrane protein	[unclassified]
D5_2265	adhesin-like protein	[unclassified]
D5_2266	adhesin-like protein	[unclassified]
D5_2268	hypothetical protein	Not in COGs
D5_2281	hypothetical protein	Not in COGs
D5_2282	hypothetical transmembrane protein	Not in COGs
D5_2289	hypothetical protein	Not in COGs
D5_2293	hypothetical transmembrane protein	Not in COGs
D5_2294	hypothetical transmembrane protein	Not in COGs
D5_2295	hypothetical protein	[M]
D5_2296	hypothetical transmembrane protein	Not in COGs
D5_2297	hypothetical protein	Not in COGs
D5_2300	hypothetical protein	Not in COGs
D5_2301	hypothetical transmembrane protein	[unclassified]
D5_2309	hypothetical transmembrane protein	[unclassified]
D5_2311	repeat domain containing protein	[M]
D5_2312	potassium channel protein	Not in COGs
D5_2313	potassium channel protein	[P]
D5_2316	conserved hypothetical transmembrane protein	[S]
D5_2317	hypothetical protein	[unclassified]
D5_2322	hypothetical protein	[C]
D5_2323	hypothetical protein	Not in COGs
D5_2325	conserved hypothetical protein	[Q]
D5_2326	hypothetical protein	Not in COGs
D5_2328	hypothetical protein	[S]
D5_2329	hypothetical transmembrane protein	[unclassified]
D5_2330	conserved hypothetical protein	[unclassified]
D5_2331	hypothetical protein	Not in COGs
D5_2335	adhesin-like protein with cysteine protease domain	[unclassified]
D5_2353	hypothetical protein	[unclassified]
D5_2354	adhesin-like protein	[O]
D5_2355	conserved hypothetical protein	[O]
D5_2356	adhesin-like protein	Not in COGs
D5_2357	adhesin-like protein	[unclassified]
D5_2358	hypothetical protein	Not in COGs
D5_2359	hypothetical protein	Not in COGs
D5_2365	hypothetical protein	[T]
D5_2371	conserved hypothetical protein	[H]
D5_2406	hypothetical protein	Not in COGs
D5_2407	adhesin-like protein with cysteine protease domain	[O]
D5_2415	hypothetical transmembrane protein	Not in COGs
D5_2419	hypothetical protein	Not in COGs
D5_2445	hypothetical protein	Not in COGs
D5_2446	hypothetical protein	Not in COGs
D5_2451	hypothetical transmembrane protein	Not in COGs
D5_2460	hypothetical protein	[unclassified]
D5_2461	hypothetical protein	Not in COGs
D5_2465	transposase	[L]

only one gene from D5 for each gene family is represented

Table A.6.9. Genes involved in methanogenesis and energy generation in *Methanobrevibacter* spp. genomes analysed

	D5	**YE315**	**ZA-10^{T}***	**SM9**	**HO^{T}***	**PS^{T}**	**JMR01***	**M1^{T}**	**KM1H5-1P^{T}***	**YLM1$^{\#}$**	**SH^{T}***	**AbM4***	**JH1^{T}***	**RFM-1^{T}***	**RFM-2^{T}***	**RFM-3^{T}***	**ATM^{T}***	**ANOR1***
Formate metabolism																		
fdhC	D5_2393	TL18_00915	2595204654	mbs_2229	2657040815	Msm_1403	2620725525	MRU_RS01705	2595158485	YLM1_0029	2558933326	2540853337	2553937917	Ga0074191_10501	Ga0074190_111836	Ga0078795_10806	Ga0078798_108220	2618697655, 2618697091
fdhA	D5_2392	TL18_00920 (pseudo)	2595204653	mbs_2228	2657040816	Msm_1404	2620725524	MRU_RS01710	2595158484	YLM1_0030	2558933327	2540853338	2553937918	Ga0074191_10502	Ga0074190_111837	Ga0078795_10805	Ga0078798_108219	2618697656
fdhB	D5_2391	TL18_00925	2595204652	mbs_2227	2657040817	Msm_1405	2620725523	MRU_RS01715	2595158483	YLM1_0031	2558933328	2540853339	2553937919	Ga0074191_10503	Ga0074190_111838	Ga0078795_10804	Ga0078798_108218	2618697657
fdhD	D5_0772, D5_2394	TL18_00875, TL18_07080	2595204656, 2595203780	mbs_1636, mbs_2231	2657040813	Msm_1392, Msm_0295	2620725176	MRU_RS03460, MRU_RS09705	2595158965	YLM1_0328	2558934145	2540854360	2553936931	Ga0074191_10513	Ga0074190_111835	Ga0078795_10809	Ga0078798_121524	2618697650, 2618698474
flpE-like	D5_0737	TL18_01665 (pseudo)	2595205493	mbs_1376	-	Msm_1464	2620726189	-	-	-	-	-		-	-	-	-	-
flpA	D5_0738	TL18_01670	2595205494	mbs_1375	-	Msm_1463	2620726188	MRU_RS10425	2595158368, 2595159808	YLM1_1118, YLM1_1559	-	2540854057	2553937228	-	-	-	-	-
flpB	D5_0739	TL18_01675	2595205495	mbs_1374	2657041195	Msm_1462	2620726187	MRU_RS10430	2595158369, 2595159809	YLM1_1117, YLM1_1558	-	2540854056	2553937229	-	-	-	-	-
flpD	D5_0740	TL18_01680	2595205496	mbs_1373	2657041194	Msm_1461	2620726186	MRU_RS10435	2595158370, 2595159810	YLM1_1116, YLM1_1557	-	2540854055	2553937230	-	-	-	-	-
Tungsten formylmethanofuran dehydrogenase																		
fwdE	D5_1751	TL18_00895	2595204655	mbs_2230	2657040814	Msm_1396	2620724795, 2620726394	MRU_RS01355	2595158492	YLM1_0022	2558933343	2540853352	2553937932	Ga0074191_109211	Ga0074190_11026	Ga0078795_10642	Ga0078798_121528	2618697653
fwdC	D5_2382	TL18_00990	2595204643	mbs_2217	2657040826	Msm_1414	2620725513	MRU_RS01770	2595158473	YLM1_0041	2558933338	2540853350	2553937930	Ga0074191_105012	Ga0074190_104511	Ga0078795_13898	Ga0078798_10829	2618697660

fwdA	D5_2383	TL18_00985	2595204644	mbs_2218	2657040825	Msm_1413	2620725514	MRU_RS01765	2595158474	YLM1_0040	2558933337	2540853349	2553937929	Ga0074191_105011	Ga0074190_104510	Ga007879 5_13897	Ga0078798_108210	2618697661
fwdB	D5_2384	TL18_00980	2595204645	mbs_2219	2657040824	Msm_1412	2620725515	MRU_RS01760	2595158475	YLM1_0039	2558933336	2540853348	2553937928	Ga0074191_105010	Ga0074190_10459	Ga007879 5_13896	Ga0078798_108211	2618697662
fwdD	D5_2385	TL18_00975	2595204646	mbs_2220	2657040823	Msm_1411	2620725516	MRU_RS01755	2595158476	YLM1_0038	2558933335	2540853347	2553937927	Ga0074191_10509	Ga0074190_10458	Ga007879 5_13895	Ga0078798_108212	2618697663
fwdG	D5_2386	TL18_00970	2595204647	mbs_2221	2657040822	Msm_1410	2620725517	MRU_RS01750	2595158477	YLM1_0037	2558933334	2540853346	2553937926	Ga0074191_10508	Ga0074190_10457	Ga007879 5_13894	Ga0078798_108213	2618697664
fwdF	D5_2387	TL18_00965	2595204648	mbs_2222	2657040821	Msm_1409	2620725518	MRU_RS01745	2595158478	YLM1_0036	2558933333	2540853345	2553937925	Ga0074191_10507	Ga0074190_10456	Ga007879 5_13893	Ga0078798_108214	2618697666
fwdH	D5_2388	TL18_00960	2595204649	mbs_2223	2657040820	Msm_1408	2620725519	MRU_RS01740	2595158479	YLM1_0035	2558933332	2540853344	2553937924	Ga0074191_10506	Ga0074190_10455	Ga007879 5_13892	Ga0078798_108215	2618697667
Formylmethanofuran- H_4MPT formyltransferase																		
ftr	D5_0210, D5_1769	TL18_03515, TL18_09675	2595204243, 2595205618	mbs_0743, mbs_2147	2657039708, 2657040050	Msm_0308, Msm_1092	2620724610, 2620725209	MRU_RS06965, MRU_RS10125	2595159281, 2595159591	YLM1_0902, YLM1_1480	2558933862, 2558934229	2540854134, 2540854788	2553937147, 2553937646	Ga0074191_107324, Ga0074191_111353	Ga0074190_111131, Ga0074190_110921	Ga007879 5_102513, Ga007879 5_116217	Ga0078798_125770, Ga0078798_109535	2618697292, 2618698615
Methenyltetrahydromethanopterin cyclohydrolase																		
mch	D5_1626	TL18_04360	2595205554	mbs_1355	2657040207	Msm_1723	2620724969, 2620726285	MRU_RS08105	2595158371	YLM1_1115	2558933616	2540854399	2553936892	Ga0074191_109714	Ga0074190_113227	Ga007879 5_14478	Ga0078798_106416	2618698730
H_2 forming methenyltetrahydromethanopterin dehydrogenase																		
hmd	D5_2132	TL18_01650, TL18_01655	2595205021, 2595204101	mbs_0405, mbs_1161	2657040574	Msm_0572	2620725265	MRU_RS02670	2595159020	YLM1_0194	2558933466	2540853216	2553937792	-	-	-	-	-
F_{420}-dependent methylenetetrahydromethanopterin dehydrogenase Mtd																		
mtd	D5_0079, D5_0741	TL18_10280	2595203961, 2595204702	mbs_0086, mbs_1377	2657039932	Msm_1204	2620724741	MRU_RS10770	2595159382	YLM1_1696	2558934910	2540853248	2553937824	Ga0074191_10304	Ga0074190_106626	Ga007879 5_11587	Ga0078798_1095116	2618697459
5,10-methylenetetrahydromethanopterin reductase Mer																		
mer	D5_2110	TL18_02180	2595205053	mbs_0427	2657040596	Msm_0542	2620726500	MRU_RS02895	2595159063	YLM1_0234	2558934989	2540854611	2553937467	Ga0074191_111924	Ga0074190_11272	Ga007879 5_12443	Ga0078798_11093	2618697110
H_4MPT *S*-methyltransferase																		
mtrE	D5_0289	TL18_09273	2595205127	mbs_2065	2657040125	Msm_1014	2620724514	MRU_RS09625	2595158116	YLM1_1394	2558933935	2540854715	2553937572	Ga0074191_103411	Ga0074190_108988	Ga007879 5_108120	Ga0078798_1257141	2618697767
mtrD	D5_0290	TL18_09274	2595205128	mbs_2064	2657040126	Msm_1013	2620724513	MRU_RS09620	2595158115	YLM1_1393	2558933936	2540854714	2553937571	Ga0074191_103412	Ga0074190_108989	Ga007879 5_108119	Ga0078798_1257142	2618697768
mtrC	D5_0291	TL18_09275	2595205129	mbs_2063	2657040127	Msm_1012	2620724512	MRU_RS09615	2595158114	YLM1_1392	2558933937	2540854713	2553937570	Ga0074191_103413	Ga0074190_108990	Ga007879 5_108118	Ga0078798_1257143	2618697769
mtrB	D5_0292	TL18_09276	2595205130	mbs_2062	2657040128	Msm_1011	2620724511	MRU_RS09610	2595158113	YLM1_1391	2558933938	2540854712	2553937569	Ga0074191_103414	Ga0074190_108991	Ga007879 5_108117	Ga0078798_1257144	2618697770

mtrA	D5_0293	TL18_09277	2595205131	mbs_2061	2657040129	Msm_1010	2620724510	MRU_RS09605	2595158112	YLM1_1390	2558933939	2540854711	2553937568	Ga0074191_103415	Ga0074190_108992	Ga007879_108116	Ga0078798_1257145	2618697771
mtrF	D5_0294	TL18_09278	2595205132	mbs_2060	2657040130	Msm_1009	2620724509	MRU_RS09600	2595158111	YLM1_1389	2558933940	2540854710	2553937567	Ga0074191_103416	Ga0074190_108993	Ga007879_108115	Ga0078798_1257146	2618697772
mtrG	D5_0295	TL18_09279	2595205133	mbs_2059	2657040131	Msm_1008	2620724508	MRU_RS09595	2595158110	YLM1_1388	2558933941	2540854709	2553937566	Ga0074191_103417	Ga0074190_108994	Ga007879_108114	Ga0078798_1257147	2618697773
mtrH	D5_0296, D5_0350	TL18_09280	2595205134, 2595204141	mbs_2058, mbs_1954	2657040132, 2657040180	Msm_1007	2620724507	MRU_RS09590	2595158109, 2595158108	YLM1_1387, YLM1_1386	2558933942, 2558934060	2540854708	2553937565	Ga0074191_103418	Ga0074190_108995	Ga007879_108113	Ga0078798_1257148	2618697774
Methyl-CoM reductase I/II																		
mcrB	D5_0284	TL18_09340	2595205122	mbs_2070	2657040120	Msm_1019	2620724520	MRU_RS09650	2595158121	YLM1_1399	2558933930	2540854720	2553937577	Ga0074191_10346	Ga0074190_108983	Ga007879_108125	Ga0078798_1257136	2618697762
mcrD	D5_0285	TL18_09335	2595205123	mbs_2069	2657040121	Msm_1018	2620724519	MRU_RS09645	2595158120	YLM1_1398	2558933931	2540854719	2553937576	Ga0074191_10347	Ga0074190_108984	Ga007879_108124	Ga0078798_1257137	2618697763
mcrC	D5_0286	TL18_09330	2595205124	mbs_2068	2657040122	Msm_1017	2620724518	MRU_RS09640	2595158119	YLM1_1397	2558933932	2540854718	2553937575	Ga0074191_10348	Ga0074190_108985	Ga007879_108123	Ga0078798_1257138	2618697764
mcrG	D5_0287	TL18_09325	2595205125	mbs_2067	2657040123	Msm_1016	2620724517	MRU_RS09635	2595158118	YLM1_1396	2558933933	2540854717	2553937574	Ga0074191_10349	Ga0074190_108986	Ga007879_108122	Ga0078798_1257139	2618697765
mcrA	D5_0288	TL18_09320	2595205126	mbs_2066	2657040124	Msm_1015	2620724515(C-terminal)-2620724516(N-terminal)	MRU_RS09630	2595158117	YLM1_1395	2558933934	2540854716	2553937573	Ga0074191_103410	Ga0074190_108987	Ga007879_108121	Ga0078798_1257140	2618697766
atwA	D5_0328	TL18_09125	2595203934	mbs_1973, mbs_1409	2657040164, 2657041025	Msm_0971, Msm_1698	2620724442	MRU_RS09260	2595158296	YLM1_0536, YLM1_0551	2558933974	2540854674, 2540853976	2553937531	Ga0074191_105912, Ga0074191_11579	Ga0074190_116229, Ga0074190_10729	Ga007879_132016, Ga007879_12083	Ga0078798_1257238, Ga0078798_1257557	2618697892
mrtB	D5_0742	-	2595203957	mbs_1381	2657041190	Msm_0905	2620726409	-	2595158366	YLM1_1120	-	-	-	-	-	-	-	-
mrtD	D5_0743	-	2595203958	mbs_1380	2657041191	Msm_0904	2620726410	-	2595158365	YLM1_1121	-	-	-	-	-	-	-	-
mrtG	D5_0744	-	2595203959	mbs_1379	2657041192	Msm_0903	2620726411	-	2595158364	YLM1_1122	-	-	-	-	-	-	-	-
mrtA	D5_0745	-	2595203960	mbs_1378	2657041193	Msm_0902	2620726412	-	2595158363	YLM1_1123	-	-	-	-	-	-	-	-
CoB--CoM heterodisulfide reductase																		
hdrA	D5_1502,	TL18_00375, TL18_05450	2595203696, 2595204522	mbs_1003,	2657039812, 2657040761	Msm_0082, Msm_1336	2620726078, 2620725595	MRU_RS00595	2595158659, 2595159533	YLM1_0883, YLM1_1573	2558934522, 2558934537	2540853897, 2540854877	2553937394, 2553937738	Ga0074191_103422, Ga0074191_114662	Ga0074190_118171, Ga0074190_11912	Ga007879_12357, Ga007879_117114	Ga0078798_1257522, Ga0078798_101962	2618697374, 2618697781

	D5_2467			mbs_2297														
hdrB	D5_0650, D5_1005, D5_1503	TL18_04030, TL18_05455, TL18_07635	259520 3697, 259520 3633	mbs_1002, mbs_1427, mbs_1746	265703 9811, 265703 9550	Msm_0083, Msm_0795	262072 6076, 262072 4228	MRU_RS04120	259515 9493, 259515 9534	YLM1_0884, YLM1_0408, YLM1_0627	255893 4521, 255893 3761	254085 3896, 254085 3960, 254085 4504	255393 7395, 255393 6788, 255393 7325	Ga0074191_103423, Ga0074191_112813	Ga0074190_100425, Ga0074190_118170	Ga007879 5_11535, Ga007879 5_12358	Ga0078798_109744, Ga0078798_125752	261869 6945, 261869 7782
hdrC	D5_0649, D5_1504	TL18_05460, TL18_07640	259520 3698, 259520 3632	mbs_1001, mbs_1747	265703 9810, 265703 9549	Msm_0084, Msm_0796	262072 6075, 262072 4229	MRU_RS04115	259515 9494	YLM1_0885, YLM1_0407	255893 4520, 255893 3762	254085 3895, 254085 4505	255393 7396, 255393 6787	Ga0074191_112814, Ga0074191_11427	Ga0074190_100426, Ga0074190_118169	Ga007879 5_11536, Ga007879 5_12359	Ga0078798_109745, Ga0078798_1257520	261869 6946, 261869 7783
Methyl viologen-reducing hydrogenase																		
mvhD	D5_0302, D5_0740	TL18_09250, TL18_01680	259520 5140	mbs_1373, mbs_2052	265704 0138, 265704 1194	Msm_1001, Msm_1461	262072 4501	MRU_RS09550, MRU_RS10435	259515 8102, 259515 8370	YLM1_1380, YLM1_1116	255893 3948	254085 4702, 254085 4055	255393 7559, 255393 7230	Ga0074191_10499,	Ga0074190_10208	Ga007879 5_11467	Ga0078798_1257158	261869 7786
mvhG	D5_0303	TL18_09245	259520 5141	mbs_2051	265704 0139	Msm_1000	262072 4500	MRU_RS09545	259515 8101	YLM1_1379	255893 3949	254085 4701	255393 7558	Ga0074191_10498	Ga0074190_10207	Ga007879 5_11468	Ga0078798_1257159	261869 7787
mvhA	D5_0304	TL18_09240	259520 5142	mbs_2050	265704 0140	Msm_0999	262072 4499	MRU_RS09540	259515 8100	YLM1_1378	255893 3950	254085 4700	255393 7557	Ga0074191_10497	Ga0074190_10206	Ga007879 5_11469	Ga0078798_1257160	261869 7788
mvhB	D5_0305	TL18_09235	259520 5143	mbs_2049	265704 0141	Msm_0998	262072 4498	MRU_RS09535	259515 8099	YLM1_1377	255893 3951	254085 4699	255393 7556	Ga0074191_10496	Ga0074190_10205	Ga007879 5_114610	Ga0078798_1257161	261869 7789
Coenzyme F_{420} hydrogenase																		
frhA	D5_0174	TL18_09880	259520 5289	mbs_0195	265704 0009	Msm_1124	262072 4656	MRU_RS10375	259515 9549	YLM1_1825	255893 4591	254085 4823	255393 7682	Ga0074191_10899	Ga0074190_113617	Ga007879 5_123415	Ga0078798_125732	261869 7332
frhD	D5_0175	TL18_09875	259520 5288	mbs_0196, mbs_0194	265704 0010	Msm_1123	262072 4655	MRU_RS10370	259515 9550	YLM1_1826	255893 4592	254085 4822	255393 7681	Ga0074191_10898	Ga0074190_113618	Ga007879 5_123414	Ga0078798_125733	261869 7331
frhG	D5_0176	TL18_09870	259520 5287	mbs_0197,	265704 0011	Msm_1122	262072 4654	MRU_RS10365	259515 9551	YLM1_1827	255893 4593	254085 4821	255393 7680	Ga0074191_10897	Ga0074190_113619	Ga007879 5_123413	Ga0078798_125734	261869 7330
frhB	D5_0177	TL18_09865	259520 5286	mbs_0198, mbs_1479	265704 0012	Msm_1121	262072 4653	MRU_RS10360	259515 9552	YLM1_1828	255893 4594	254085 4820	255393 7679	Ga0074191_10896	Ga0074190_113620	Ga007879 5_123412	Ga0078798_125735	261869 7329
Energy-converting hydrogenase A																		
ehaA	D5_1770	TL18_03425	259520 4261	mbs_0725	265703 9726	Msm_0326	262072 5227	MRU_RS07055	259515 9300	YLM1_0921	255893 4211	254085 4152	255393 7129	Ga0074191_107342	Ga0074190_11093	Ga007879 5_10112	Ga0078798_109517	261869 8633
ehaB	D5_1771	TL18_03430	259520 4260	mbs_0726	265703 9725	Msm_0325	262072 5226	MRU_RS07050	259515 9299	YLM1_0920	255893 4212	254085 4151	255393 7130	Ga0074191_107341	Ga0074190_11094	Ga007879 5_10113	Ga0078798_109518	261869 8632

ehaC	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1772	03435	4259	0727	9724	324	5225	S07045	9298	_0919	4213	4150	7131	_107340	_11095	5_10114	109519	8631
ehaD	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1773	03440	4258	0728	9723	323	5224	S07040	9297	_0918	4214	4149	7132	_107339	_11096	5_10115	109520	8630
ehaE	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1774	03445	4257	0729	9722	322	5223	S07035	9296	_0917	4215	4148	7133	_107338	_11097	5_10116	109521	8629
ehaF	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1775	03450	4256	0730	9721	321	5222	S07030	9295	_0916	4216	4147	7134	_107337	_11098	5_10117	109522	8628
ehaG	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1776	03455	4255	0731	9720	320	5221	S07025	9294	_0915	4217	4146	7135	_107336	_11099	5_10118	109523	8627
ehaH	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1777	03460	4254	0732	9719	319	5220	S07020	9293	_0914	4218	4145	7136	_107335	_110910	5_10119	109524	8626
ehaI	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1778	03465	4253	0733	9718	318	5219	S07015	9292	_0913	4219	4144	7137	_107334	_110911	5_101110	109525	8625
ehaJ	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1779	03470	4252	0734	9717	317	5218	S07010	9291	_0912	4220	4143	7138	_107333	_110912	5_101111	109526	8624
ehaK	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1780	03475	4251	0735	9716	316	5217	S07005	9290	_0911	4221	4142	7139	_107332	_110913	5_116227	109527	8623
ehaL	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1781	03480	4250	0736	9715	315	5216	S07000	9289	_0910	4222	4141	7140	_107331	_110914	5_116226	109528	8622
ehaM	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1782	03485	4249	0737	9714	314	5215	S06995	9288	_0909	4223	4140	7141	_107330	_110915	5_116225	109529	8621
ehaN	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1783	03490	4248	0738	9713	313	5214	S06990	9287	_0908	4224	4139	7142	_107329	_110916	5_116224	109530	8620
ehaO	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1784	03495	4247	0739	9712	312	5213	S06985	9286	_0907	4225	4138	7143	_107328	_110917	5_116223	109531	8619
ehaP	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1785	03500	4246	0740	9711	311	5212	S06980	9285	_0906	4226	4137	7144	_107327	_110918	5_116222	109532	8618
ehaQ	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1786	03505	4245	0741	9710	310	5211	S06975	9283	_0905	4227	4136	7145	_107326	_110919	5_116221	109533	8617
ehaR	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	1787	03510	4244	0742	9709	309	5210	S06970	9282	_0904	4228	4135	7146	_107325	_110920	5_116220	109534	8616
ehaS-	D5_	TL18_	259520	mbs_	265703	Msm_0	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
like	1766	03530	4240	0746	9705	305	5206	S06950	9278	_0899	4232	4131	7150	_107321	_110924	5_116214	109538	8612
ehaT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Energy-converting hydrogenase B																		
ehbA	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0225	09630	5602	2131	0060	076	4599	S10085	8190	_1471	3869	4780	7638	_109041	_106936	5_109118	125778	7708
ehbB	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0226	09625	5601	2130	0061	075	4598	S10080	8189	_1470	3870	4779	7637	_109040	_106937	5_109117	125779	7707
ehbC	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0227	09620	5600	2129	0062	074	4597	S10075	8188	_1469	3871	4778	7636	_109039	_106938	5_109116	125780	7706
ehbD	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	-	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0228	09615	5599	2128	0063	073	4596	S10070	8187	_1468		4777	7635	_109038	_106939	5_109115	125781	7705
ehbE	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0229	09610	5598	2127	0064	072	4595	S10065	8186	_1467	3872	4776	7634	_109037	_106940	5_109114	125782	7704
ehbF	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0230	09605	5597	2126	0065	071	4594	S10060	8185	_1466	3873	4775	7633	_109036	_106941	5_109113	125783	7703

ehbG	D5_0231	TL18_09600	2595205596	mbs_2125	2657040066	Msm_1070	2620724593	MRU_RS10055	2595158184	YLM1_1465	2558933874	2540854774	2553937632	Ga0074191_109035	Ga0074190_106942	Ga007879 5_109112	Ga0078798_125784	2618697702
ehbH	D5_0232	TL18_09595	2595205595	mbs_2124	2657040067	Msm_1069	2620724592	MRU_RS10050	2595158183	YLM1_1464	2558933875	2540854773	2553937631	Ga0074191_109034	Ga0074190_106943	Ga007879 5_109111	Ga0078798_125785	2618697701
ehbI	D5_0233	TL18_09590	2595205594	mbs_2123	2657040068	Msm_1068	2620724591	MRU_RS10045	2595158182	YLM1_1463	2558933876	2540854772	2553937630	Ga0074191_109033	Ga0074190_106944	Ga007879 5_109110	Ga0078798_125786	2618697700
ehbJ	D5_0234	TL18_09585	2595205593	mbs_2122	2657040069	Msm_1067	2620724590	MRU_RS10040	2595158181	YLM1_1462	2558933877	2540854771	2553937629	Ga0074191_109032	Ga0074190_106945	Ga007879 5_10919	Ga0078798_125787	2618697699
ehbK	D5_0235	TL18_09580	2595205592	mbs_2121	2657040070	Msm_1066	2620724589	MRU_RS10035	2595158180	YLM1_1461	2558933878	2540854770	2553937628	Ga0074191_109031	Ga0074190_106946	Ga007879 5_10918	Ga0078798_125788	2618697698
ehbL	D5_0236	TL18_09575	2595205591	mbs_2120	2657040071	Msm_1065	2620724588	MRU_RS10030	2595158179	YLM1_1460	2558933879	2540854769	2553937627	Ga0074191_109030	Ga0074190_106947	Ga007879 5_10917	Ga0078798_125789	2618697697
ehbM	D5_0237	TL18_09570	2595205590	mbs_2119	2657040072	Msm_1064	2620724587	MRU_RS10025	2595158178	YLM1_1459	2558933880	2540854768	2553937626	Ga0074191_109029	Ga0074190_106948	Ga007879 5_10916	Ga0078798_125790	2618697696
ehbN	D5_0238	TL18_09565	2595205589	mbs_2118	2657040073	Msm_1063	2620724586	MRU_RS10020	2595158177	YLM1_1458	2558933881	2540854767	2553937625	Ga0074191_109028	Ga0074190_106949	Ga007879 5_10915	Ga0078798_125791	2618697695
ehbO	D5_0239	TL18_09560	2595205588	mbs_2117	2657040074	Msm_1062	2620724585	MRU_RS10015	2595158176	YLM1_1457	2558933882	2540854766	2553937624	Ga0074191_109027	Ga0074190_106950	Ga007879 5_10914	Ga0078798_125792	2618697694
ehbP	D5_0240	TL18_09555	2595205587	mbs_2116	2657040075	Msm_1061	2620724584	MRU_RS10010	2595158175	YLM1_1456	2558933883	2540854765	2553937623	Ga0074191_109026	Ga0074190_106951	Ga007879 5_10913	Ga0078798_125793	2618697693
ehbQ	D5_0241	TL18_09550	2595205586	mbs_2115	2657040076	Msm_1060	2620724583	MRU_RS10005	2595158174	YLM1_1455	2558933884	2540854764	2553937622	Ga0074191_109025	Ga0074190_106952	Ga007879 5_10912	Ga0078798_125794	2618697692
Hydrogenase helper genes																		
hypA	D5_1528	TL18_05565	2595203723	mbs_0965	2657039642	Msm_0108	2620724989	MRU_RS08165	2595158367, 2595158351	YLM1_1119, YLM1_1135	2558934039	2540853801	2553938447	Ga0074191_11337	Ga0074190_113218	Ga007879 5_12974	Ga0078798_1257424	2618698226
hypB	D5_1527	TL18_05560	2595203722	mbs_0966	2657039641	Msm_0107	2620724988	MRU_RS08160	2595158352	YLM1_1134	2558934038	2540853802	2553938448	Ga0074191_11336	Ga0074190_113219	Ga007879 5_12975	Ga0078798_1257425	2618698227
hypC	D5_2235	TL18_01440	2595204947	mbs_0341	2657040511	Msm_0636	2620725430	MRU_RS02400	2595158889	YLM1_0147	2558933425	2540853449	2553938025	Ga0074191_116429	Ga0074190_118520	Ga007879 5_11845	Ga0078798_111535	2618697216
hypD	D5_0364	TL18_08960	2595203981	mbs_1938	2657040189	Msm_0945	2620724406	MRU_RS09385	2595158065	YLM1_1342	2558933475	2540853492	2553938068	Ga0074191_11584	Ga0074190_11722	Ga007879 5_10657	Ga0078798_1257206	2618697858
hypE	D5_0144, D5_0807	TL18_10050, TL18_06970	2595205183, 2595204603	mbs_0155, mbs_1610	2657039979, 2657041146	Msm_1492, Msm_1158	2620724691, 2620725807	MRU_RS07745, MRU_RS00955	2595159121, 2595159866	YLM1_1048, YLM1_1506	2558933321, 2558933573	2540853330, 2540854339	2553937910, 2553936955	Ga0074191_11186, Ga0074191_112842	Ga0074190_10275, Ga0074190_113232	Ga007879 5_12916, Ga007879 5_128722	Ga0078798_101912, Ga0078798_102420	2618697386, 2618698237
hypF	D5_0194	TL18_09780	2595205183	mbs_2165	2657040028	Msm_1106	2620724639	MRU_RS10220	2595159572	YLM1_1493	2558933845	2540854804	2553937662	Ga0074191_113543	Ga0074190_11084	Ga007879 5_11564	Ga0078798_125751	2618697316
hycI	D5_1825	TL18_03185	2595205815	mbs_0680	2657039397	Msm_0362	2620726449	MRU_RS05225	2595159150	YLM1_0589	2558934177	2540854184	2553937097	Ga0074191_10534	Ga0074190_100948	Ga007879 5_107142	Ga0078798_102455	2618698684
Alternative reducing potential from alcohol																		
npdG	D5_1461	TL18_05575, TL18_05310	2595205557, 2595205861	mbs_1352	2657041207	Msm_0049	2620724964, 2620726112	MRU_RS07215, MRU_RS04990	2595159766, 2595158790	YLM1_0542, YLM1_0957	2558934599, 2558934474	2540854816, 2540853851	2553937440, 2553937675	Ga0074191_11677, Ga0074191_117027	Ga0074190_11184	Ga007879 5_10972	Ga0078798_11359	2618698283

adh (NADP dependent)	-	TL18_05985	-	-	-	-	-	MRU_RS09245	2595158215	YLM1_1244	-	-	-	-	-	-	-	-
A_1A_0 ATP synthase																		
ahaD	D5_1938	TL18_02715	2595204423	mbs_0591	2657040901	Msm_0433	2620725948	MRU_RS03565	2595158944	YLM1_0349	25583934772	2540853630	2553938215	Ga0074191_10809	Ga0074190_10443	Ga007879_5_124816	Ga0078798_1257337	2618698018
ahaB	D5_1939	TL18_02710	2595204424	mbs_0590	2657040902	Msm_0434	2620725947	MRU_RS03560	2595158945	YLM1_0348	25583934773	2540853631	2553938214	Ga0074191_108010	Ga0074190_10444	Ga007879_5_124815	Ga0078798_1257336	2618698017
ahaA	D5_1940	TL18_02705	2595204425	mbs_0589	2657040903	Msm_0435	2620725946	MRU_RS03555	2595158946	YLM1_0347	25583934774	2540853632	2553938213	Ga0074191_108011	Ga0074190_10445	Ga007879_5_124814	Ga0078798_1257335	2618698016
ahaF	D5_1941	TL18_02700	2595204426	mbs_0588	2657040904	Msm_0436	2620725945	MRU_RS03550	2595158947	YLM1_0346	25583934775	2540853633	2553938212	Ga0074191_108012	Ga0074190_10446	Ga007879_5_124813	Ga0078798_1257334	2618698015
ahaC	D5_1942	TL18_02695	2595204427	mbs_0587	2657040905	Msm_0437	2620725944	MRU_RS03545	2595158948	YLM1_0345	25583934776	2540853634	2553938211	Ga0074191_108013	Ga0074190_10447	Ga007879_5_124812	Ga0078798_1257333	2618698014
ahaE	D5_1943	TL18_02690	2595204428	mbs_0586	2657040906	Msm_0438	2620725943	MRU_RS03540	2595158949	YLM1_0344	25583934777	2540853635	2553938210	Ga0074191_108014	Ga0074190_10448	Ga007879_5_124811	Ga0078798_1257332	2618698013
ahaK	D5_1944	TL18_02685	2595204429	mbs_0585	2657040907	Msm_0439	2620725942	MRU_RS03535	2595158950	YLM1_0343	25583934778	2540853636	2553938209	Ga0074191_108015	Ga0074190_10449	Ga007879_5_124810	Ga0078798_1257331	2618698012
ahaI	D5_1945	TL18_02680	2595204430	mbs_0584	2657040908	Msm_0440	2620725941	MRU_RS03530	2595158951	YLM1_0342	25583934779	2540853637	2553938208	Ga0074191_10667	Ga0074190_104410	Ga007879_5_12489	Ga0078798_1257330	2618698011
ahaH	D5_1946	TL18_02675	2595204431	mbs_0583	2657040909	Msm_0441	2620725940	MRU_RS03525	2595158952	YLM1_0341	25583934780	2540853638	2553938207	Ga0074191_10666	Ga0074190_104411	Ga007879_5_12488	Ga0078798_1257329	2618698010
Methanogenesis Marker proteins																		
mmp1	D5_2000	TL18_02440	2595204464	mbs_0549	2657040939	Msm_0480	2620725905	MRU_RS03395	2595158676	YLM1_0311	25589334966	2540853605	2553938182	Ga0074191_11773	Ga0074190_115372	Ga007879_5_115314	Ga0078798_1257304	2618697962
mmp4	D5_1511	TL18_05490	2595203707	mbs_0994	2657039804	Msm_0095	2620724975	MRU_RS05055	2595159186	YLM1_0550	25589334502	2540853879	2553937412	Ga0074191_111612	Ga0074190_114937	Ga007879_5_11997	Ga0078798_1257534	2618698760
mmp17	D5_0635	TL18_07705	2595203606	mbs_1767	2657040725	Msm_0809	2620724242	MRU_RS08850	2595158217	YLM1_1246	25589333774	2540854515	2553936776	Ga0074191_109125	Ga0074190_105952	Ga007879_5_14603	Ga0078798_12313	2618696982
mmp15	D5_0634	TL18_07710	2595203605	mbs_1768	2657040724	Msm_0810	2620724243	MRU_RS08855	2595158218	YLM1_1247	25589333775	2540854516	2553936775	Ga0074191_109115	Ga0074190_105953	Ga007879_5_14604	Ga0078798_12312	2618696983
mmp5	D5_0633, D5_0171	TL18_07715, TL18_09905	2595203604	mbs_1769, mbs_0189	2657040723	Msm_0811, Msm_1130	2620724244, 2620725242	MRU_RS08860	2595158219	YLM1_1248	25589333776	2540854517, 2540853526	2553936774	Ga0074191_109114	Ga0074190_105954	Ga007879_5_14605	Ga0078798_12311	2618696984
mmp6	D5_0632	TL18_07720	2595203603	mbs_1770	2657040722	Msm_0812	2620724245	MRU_RS08865	2595158220	YLM1_1249	25589333777	2540854518	2553936773	Ga0074191_109113	Ga0074190_105958	Ga007879_5_14606	Ga0078798_122813	2618696985
mmp3	D5_0631	TL18_07725	2595203602	mbs_1771	2657040721	Msm_0813	2620724246	MRU_RS08870	2595158221	YLM1_1250	25589333778	2540854519	2553936772	Ga0074191_109112	Ga0074190_105959	Ga007879_5_14607	Ga0078798_122812	2618696986
mmp2	D5_0627	TL18_07740	2595203600	mbs_1772	2657040718	Msm_0814	2620724248	MRU_RS08890	2595158225	YLM1_1251	25589333779	2540854520	2553936771	Ga0074191_109111	Ga0074190_105960	Ga007879_5_14608	Ga0078798_122811	2618696987
mmp7	D5_0282	TL18_09350	2595205120	mbs_2072	2657040118	Msm_0815,	2620724523	MRU_RS09665	2595158123	YLM1_1401	25589333927	2540854723	2553937580	Ga0074191_10327	Ga0074190_106920	Ga007879_5_12915	Ga0078798_1257133	2618697751

						Msm_1021											
mmp10	D5_0283	TL18_09345	2595205121	mbs_2071	2657040119	Msm_1020	2620724521	MRU_RS09655	2595158122	YLM1_1400	2558933929	2540854721	2553937578	Ga0074191_10345	Ga0074190_11181	Ga0078795_108126	Ga0078981257135	2618697761
mmp8	D5_2244	TL18_01400	2595205337	mbs_0333	2657040449	Msm_0643	2620724850	MRU_RS02215	2595159237	YLM1_0106	2558934421	2540854313	2553936981	Ga0074191_11034	Ga0074190_111411	Ga0078795_115416	Ga0078981257629	2618697231
mmp9	D5_0332	TL18_09115	2595203941	mbs_1970	2657040166	Msm_0969	2620724439	MRU_RS09265	2595158297	YLM1_1321	2558933975	2540854673	2553937530	Ga0074191_105911	Ga0074190_116230	Ga0078795_132015	Ga0078981257237	2618697893
mmp11	D5_2489	TL18_00260	2595204773	mbs_2335	2657040737	Msm_1295	2620725630	MRU_RS00495	2595158642	YLM1_1590	2558934693	2540854901	2553937763	Ga0074191_10153	Ga0074190_111844	Ga0078795_11337	Ga0078981101995	2618697533
mmp12	D5_0062	TL18_10390	2595204725	mbs_0064	2657039909	Msm_1221	2620726351	MRU_RS10605	2595159458	YLM1_1730	2558933290	2540853273	2553937850	Ga0074191_10707	Ga0074190_11352	Ga0078795_113010	Ga0078981095141	2618697480
mmp13	D5_0142	TL18_10060	2595204601	mbs_0153	2657039977	Msm_1160	2620724693	MRU_RS00910	2595159111	YLM1_1516	2558933319	2540853328	2553937908	Ga0074191_11272	Ga0074190_11112	Ga0078795_10302	Ga0078981101915	2618697387
mmp14	D5_0297	TL18_09275	2595205135	mbs_2057	2657040133	Msm_1006	2620724506	MRU_RS09585	2595158107	YLM1_1385	2558933943	2540854707	2553937564	Ga0074191_103419	Ga0074190_102011	Ga0078795_10818	Ga0078981257149	2618697777
mmp16	D5_0150	TL18_05465, TL18_10010	2595203699, 2595205319	mbs_1000, mbs_0163	2657039809, 2657039985	Msm_0085	2620724679, 2620726074	MRU_RS09680	2595158125, 2595159536	YLM1_0886, YLM1_1403	2558933494, 2558934519	2540853519, 2540853894	2553937397, 2553938099	Ga0074191_107212	Ga0074190_118166	Ga0078795_123512	Ga0078981257523	2618697195
F_{420} biosynthesis																		
cofA	D5_1974	TL18_02565	2595204453	mbs_0557	2657040931	Msm_0467	2620725913	MRU_RS03415	2595158976	YLM1_0316	2558934971	2540853609	2553938186	Ga0074191_117710	Ga0074190_109324	Ga0078795_12159	Ga0078981257306	2618697968
fucA	D5_2508	TL18_00160	2595204800	mbs_2363	2657039858	Msm_1270	2620726296	MRU_RS11110	2595158600	YLM1_1634	2558934713	2540854921	2553937781	Ga0074191_102614	Ga0074190_103430	Ga0078795_11833	Ga0078981019112	2618697520
cofC	D5_1723	TL18_06430	2595205421	mbs_0786	2657040342	Msm_0288	2620725166	MRU_RS04805	2595158831	YLM1_0517	2558933627	2540854406	2553936885	Ga0074191_114657	Ga0074190_105627	Ga0078795_118516	Ga0078981257210	2618698829
cofH	D5_1498	TL18_05430	2595205699	mbs_1010	2657039815	Msm_0079	2620726081	MRU_RS06335	2595159655	YLM1_0646	2558934853	2540853974	2553937310	Ga0074191_10448	Ga0074190_115329	Ga0078795_12238	Ga0078981257555	2618698143
cofD	D5_0325	TL18_09140	2595203931	mbs_1976	2657040161	Msm_0974	2620724446	MRU_RS09230	2595158292	YLM1_1316	2558933971	2540854677	2553937534	Ga0074191_105915	Ga0074190_11626	Ga0078795_12644	Ga0078981257243	2618697889
cofE	D5_0324, D5_0503, 2229	TL18_01485, TL18_07135, TL18_09145	2595203908, 2595203930, 2595204954	mbs_0348, mbs_1590, mbs_1977	2657039374, 2657040160, 2657040520	Msm_0630, Msm_0975	2620724447, 2620725418	MRU_RS02435, MRU_RS09220, MRU_RS04775	2595158290, 2595158896	YLM1_0154, YLM1_1314	2558933432, 2558933970	2540853457, 2540854680	2553937537, 2553938033	Ga0074191_116414	Ga0074190_11627	Ga0078795_12643	Ga0078981111526	2618697207, 2618697900
cofG	D5_0268	TL18_09420	2595205679	mbs_2086	2657040104	Msm_1035	2620724550	MRU_RS09880	2595158151	YLM1_1429	2558933912	2540854737	2553937594	Ga0074191_103211	Ga0074190_106618	Ga0078795_11188	Ga0078981257118	2618697743
cofF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F_{430} biosynthesis																		
ftsA	D5_1885	TL18_02950	2595204385	mbs_0629	2657039354	Msm_0387	2620725990	MRU_RS08925	2595158232	YLM1_1261	2558933786	2540854528	2553936761	Ga0074191_113716, Ga0074191_115715	Ga0074190_10593, Ga0074190_10936	Ga0078795_12463, Ga0078795_11315	Ga0078981228 3, Ga0078981095 44	2618696998

CoM biosynthesis																		
comA	D5_1506	TL18_05470	259520_3700	mbs_0999	265703_9808	Msm_0086	262072_6073	-	259515_9537	YLM1_0887	255893_4518	254085_3893	255393_7398	Ga0074191_116413	Ga0074190_10706	Ga007879_11527	Ga0078798_1257524	2618697206
comB	D5_0280	TL18_09360	259520_5118	mbs_2074	265704_0116	Msm_1023	262072_4525	MRU_RS09755	259515_8138	YLM1_1416	255893_3925	254085_4725	255393_7582	Ga0074191_10325	Ga0074190_118153	Ga007879_12133	Ga0078798_1257131	2618697749
comC	D5_0263	TL18_09445	259520_5671	mbs_2091	265704_0099	Msm_1040	262072_4555	MRU_RS09915	259515_8156	YLM1_1434	255893_3907	254085_4742	255393_7599	Ga0074191_103217	Ga0074190_112510	Ga007879_11183	Ga0078798_1257113	2618697728
comD	D5_1500	TL18_05435	259520_5698	mbs_1009	265703_9814	Msm_0080	262072_6080	-	259515_9268	YLM1_0891	255893_4526	254085_3900	255393_7391	Ga0074191_103218	Ga0074190_118172	Ga007879_12353	Ga0078798_1257526	2618697729
comE	D5_1501	TL18_05440	259520_5697	mbs_1008	265703_9813	Msm_0081	262072_6079	-	259515_9269	YLM1_0892	255893_4525	254085_3899	255393_7392	Ga0074191_103219	Ga0074190_118174	Ga007879_12352	Ga0078798_1257525	2618697730 + 2618697731 (interupted)
CoB biosynthesis																		
aksA	D5_2368	TL18_01070	259520_4630	mbs_2204	265704_0840	Msm_0722	262072_5500	MRU_RS01960	259515_8457	YLM1_0056	255893_3368	254085_3386	255393_7967	Ga0074191_10333	Ga0074190_108982	Ga007879_5_108016	Ga0078798_111593	2618697687
aksD	D5_2369	TL18_01065	259520_4631	mbs_2205	265704_0839	Msm_0723	262072_5501	MRU_RS01955	259515_8458	YLM1_0055	255893_3367	254085_3385	255393_7966	Ga0074191_10334	Ga0074190_108981	Ga007879_5_108015	Ga0078798_111594	2618697686
aksE	D5_0588	TL18_07930	259520_3567	mbs_1805	265704_0684	Msm_0847	262072_5628	MRU_RS08455	259515_9789	YLM1_1170	255893_4092	254085_3716	255393_8362	Ga0074191_11133	Ga0074190_115326	Ga007879_5_12235	Ga0078798_1257397	2618698130
aksF	D5_1844	TL18_03130	259520_5711	mbs_0656	265703_9388	Msm_0373	262072_6405	MRU_RS05195	259515_9156	YLM1_0583	255893_4387	254085_4197	255393_7084	Ga0074191_117019	Ga0074190_108054	Ga007879_5_121413	Ga0078798_102445	2618697536
MF biosynthesis																		
mfnA	D5_0313	TL18_09200	259520_3920	mbs_1987	265704_0149	Msm_0987	262072_4462	MRU_RS09490	259515_8090	YLM1_1368	255893_3959	254085_4691	255393_7548	Ga0074191_11761	Ga0074190_108932	Ga007879_5_115117	Ga0078798_1257174	2618697806
mfnB	D5_1007	TL18_04040	259520_5086	mbs_1425	265704_1043	Msm_1628	262072_5653	MRU_RS06050	259515_8980	YLM1_0629	255893_4836	254085_3962	255393_7323	Ga0074191_11033	Ga0074190_10735	Ga007879_5_120812	Ga0078798_12062	2618698343
mfnD	D5_0584	TL18_07950	259520_3561	mbs_1811	265704_0680	Msm_0852	262072_4294	MRU_RS08490	259515_9782	YLM1_1178	255893_4099	254085_3707	255393_8352	Ga0074191_10818	Ga0074190_105611	Ga007879_5_10753	Ga0078798_1257390	2618698123
mfnF	D5_0585	TL18_07945	259520_3562	mbs_1810	265704_0681	Msm_0851	262072_4292	MRU_RS08485	259515_9783	YLM1_1177	255893_4098	254085_3708	255393_8353	Ga0074191_10817	Ga0074190_105612	Ga007879_5_10752	Ga0078798_1257391	2618698125
mfnE	D5_2202	TL18_01695	259520_4993	mbs_0377	265704_0545	Msm_0604	262072_5298	MRU_RS02630	259515_8925	YLM1_0186	255893_3993	254085_4652	255393_7509	Ga0074191_10592	Ga0074190_111418	Ga007879_5_11576	Ga0078798_11155	2618697171
H_4MPT biosynthesis																		
arfA/mptA	D5_0270	TL18_09410	259520_5681	mbs_2084	265704_0106	Msm_1033	262072_4548	MRU_RS09820	259515_8148	YLM1_1426	255893_3914	254085_4735	255393_7592	Ga0074191_10329	Ga0074190_106930	Ga007879_5_111810	Ga0078798_1257120	2618697741
arfB	D5_1597	TL18_05910	259520_3798	mbs_0884	265704_1003	Msm_0184	262072_5062	MRU_RS06415	259515_9602	YLM1_0803	255893_4885	254085_3998	255393_7285	Ga0074191_109215	Ga0074190_108012	Ga007879_5_11255	Ga0078798_1257631	2618698554
mptD	D5_0314	TL18_09195	259520_3921	mbs_1986	265704_0150	Msm_0985	262072_4461	MRU_RS09165	259515_8280	YLM1_1304	255893_3960	254085_4690	255393_7547	Ga0074191_10223	Ga0074190_116224	Ga007879_5_11217	Ga0078798_1257256	2618697910
mptE	D5_2280	TL18_01270	259520_5894	mbs_0300	265704_0423	Msm_0672	262072_5465	MRU_RS02140	259515_8424	YLM1_0089	255893_3394	254085_3417	255393_7993	Ga0074191_10356	Ga0074190_116117	Ga007879_5_100722	Ga0078798_111566	2618697259
mptG	D5_1075	TL18_04245	259520_3621	mbs_1758	265704_0225	Msm_0848	262072_6063	MRU_RS08460	259515_9788	YLM1_1171	255893_4095	254085_3715	255393_8360	Ga0074191_11132	Ga0074190_11174	Ga007879_5_12234	Ga0078798_1257396	2618698129
mptH	-	-	-	-	-	-	-	-	-	-	-	-	-					-

citG	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	0790	07015	5783	1619	1157	477	5818	S07785	9704	_1056	3585	4369	6922	_11497	_100941	5_107148	102414	8691
dmrX	D5_	TL18_	259520	mbs_	265704	Msm_1	262072	MRU_R	259515	YLM1	255893	254085	255393	Ga0074191	Ga0074190	Ga007879	Ga0078798_	261869
	2463	00390	4524	2294	0764	338	5593	S00625	8661	_1571	4539	4875	7736	_114660	_111464	5_13164	101960	7576

*JGI IMG gene ID number is shown instead of locus_tag. # Temporary locus_tag is displayed as this draft genome has not been submitted. Temporary locus_tag is subject to change once genome completes. Formate dehydrogenase (*fdhAB/flpAB*), formate transporter (*fdhC*), formate dehydrogenase accessory protein (*fdhD*), formylmethanofuran dehydrogenase (*fwdA-H*), formylmethanofuran-H_4MPT formyl transferase (*ftr*), N^5,N^{10}-methenyl-H_4MPT cyclohydrolase (*mch*), F_{420}-dependent methylene-H_4MPT dehydrogenase (*mtd*), F_{420} reducing hydrogenase (*frhABDG*), methylene-H_4MPT reductase (*mer*), H_4MPT S-methyltransferase (*mtrA-H*), H_2-dependent methylene-H_4MPT dehydrogenase (*hmd*), methyl-CoM reductase (*mcr*/*mrtA-G*), heterodisulfide reductase (*hdrABC*), methyl-viologen hydrogenase (*mvhADG*), energy conserving hydrogenase (*eha/ehbA-T*), hydrogenase nickel insertion protein (*hypA*), hydrogenase accessory protein (*hypB*), hydrogenase assembly chaperone (*hypC*), hydrogenase expression/formation protein (*hypDEI*), hydrogenase maturation factor (*hypF*), NADP-dependent F_{420} reductase (*npdG*), alcohol dehydrogenase (*adh*), Na^+/H^+ antiporter (*nha*), A_1A_0 ATP synthase subunits are present (*ahaA-K*), methanogenesis marker protein (*mmp1-17*), lactaldehyde dehydrogenase (*cofA*), L-fuculose phosphate aldolase (*fucA*), 2-phospho-L-lactate guanylyltransferase (*cofC*), LPPG:FO 2-phospho-L-lactate transferase (*cofD*), F_{420}-*O*:gamma-glutamyl ligase (*cofE*), γ-F_{420}-2:α-L-glutamate ligase (*cofF*), FO synthase (*cofGH*), F_{390} synthetase (*ftsA*), (2*R*)-phospho-3-sulfolactate synthase (*comA*), 2-phosphosulfolactate phosphohydrolase (*comB*), (2*R*)3-sulfolactate dehydrogenase (*comC*), sulfopyruvate decarboxylase (*comDE*), (*R*)-homocitrate synthase (*aksA*), (*R*)-homocitrate dehydratase (*aksDE*), *Threo*-isohomocitrate dehydrogenase (*aksF*), aspartate 1-decarboxylase/L-tyrosine decarboxylase (*mfnA*), 2-furaldehyde phosphate synthase (*mfnA*), tyramine—L-glutamate ligase (*mfnD*), (4-{4-[2-(γ-L-glutamylamino)ethyl]phenoxymethyl}furan-2-yl)methanamine synthase (*mfnF*), [5-(aminomethyl)furan-3-yl]methyl phosphate kinase (*mfnE*), GTP cyclohydrolase III/IV (*arfA/mptA*), 2-amino-5-formylamino-6-ribosylaminopyrimidin-4(3H)-one 5'-monophosphate deformylase (*arfB*), dihydroneopterin aldolase (*mptD*), 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (*mptE*), beta-ribofuranosylaminobenzene 5-phosphate synthase (*mptG*), dihydropteroate synthase (*mptH*), triphosphoribosyl-dephospho-CoA synthase (*citG*), dihydromethanopterin reductase (*dmrX*).

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Yang Li

Name/Title of Principal Supervisor: Jasna Rakonjac

Name of Published Research Output and full reference:

Li, Yang, Sinead C. Leahy, Jeyamalar Jeyanathan, Gemma Henderson, Faith Cox, Eric Altermann, William J. Kelly et al. "The complete genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the methylotrophic lifestyle of a ruminal representative of the Methanomassiliicoccales." Standards in Genomic Sciences 11, no. 1 (2016): 59.

In which Chapter is the Published Work: Three

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  YL conducted most of the experimental work in the Published Work. He did the microbial culturing,YL performed electron microscopy, assembled the genome, closed sequence gaps, annotated the genome and drafted the manuscript. YL analysed the data and wrote the first draft of the manuscript.

Yang Li Digitally signed by Yang Li DN: cn=Yang Li, o=AgResearch Ltd., ou, email=yang.li@agresearch.co.nz, c=NZ Date: 2016.11.30 12:19:08 +13'00'

Candidate's Signature

30/11/2016

Date

Principal Supervisor's signature

30/11/2016

Date