

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Genome Sequencing of Rumen Bacteria Involved in Lignocellulose Digestion

A thesis presented in fulfilment of the requirements for the degree of

Master of Science

in the

Institute of Molecular Biosciences

At Massey University, Palmerston North,

New Zealand.



MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA

Nikola Palevich

2011

Abstract

Determining the role of rumen microbes and their enzymes in plant polysaccharide breakdown is fundamental to understanding digestion and maximising productivity in ruminant animals. In order to learn more about lignocellulose degradation in pasture-grazed dairy cows under NZ conditions, twenty representative strains from five major phylotype clusters (*Butyrivibrio fibrisolvens/hungatei* cluster 383, *Pseudobutyrvibrio xylanivorans* clusters 247 and 245, *Selenomonas ruminantium* cluster 212, and *Lachnospiraceae* cluster 121), cultivated directly from the fibre-adherent rumen microbial fraction of dairy cows were selected. Genotypic and phenotypic analysis of these strains led to identification of *Butyrivibrio* sp. MB2003 that adheres to and efficiently degrades the plant fibre. The 3.3 Mb MB2003 genome was sequenced and annotated and found to consist of four replicons: a chromosome (7 contigs, in 1 super scaffold), a chromid (Bhu II), a megaplasmid (pNP144) and a small plasmid (pNP6). A novel feature of the MB2003 genome is the presence of a chromid (Bhu II) which is now the smallest chromid reported for all bacteria. The MB2003 polysaccharide-degrading enzymes, surface structures and predicted strategy for attachment to, and degradation of, complex polysaccharides was found to be comparable to that of the fibrolytic bacterium *Butyrivibrio proteoclasticus* B316. Both MB2003 and B316 are non-motile, despite the presence of flagellar gene clusters, and utilise a range of insoluble plant polysaccharides, but not cellulose. Xylan is the preferred insoluble substrate of MB2003 and its genome encodes a large repertoire of enzymes predicted to metabolise this complex polysaccharide. The MB2003 draft genome produced in this work is the first opportunity to conduct comparative analysis of two rumen bacteria belonging to the same genus. Although both MB2003 and B316 have similar phenotypic characteristics and occupy the same habitat, the genome of MB2003 is much smaller and contains fewer extracellular polysaccharide degrading enzymes. From this comparison it can be concluded that MB2003 is a secondary hemicellulose degrader, offering an alternate view of the genes required for a xylanolytic lifestyle in the rumen, and posing an interesting question about the purpose of the wider range of polysaccharide degrading enzymes found in B316.

Acknowledgements

To my supervisors Dr Graeme Attwood, Dr William (Bill) Kelly and Dr Jasna Rakonjac, thank you for taking me on as a Masters student and for your encouragement throughout this journey. I trust that my work has met your expectations and hope that you have enjoyed teaching me as I have enjoyed learning from all of you.

To the great people of AgResearch for their financial support and making this project possible.

I wish to thank the AgR and FRST for NERF funding for the “Accessing the uncultured rumen microbiome” programme (contract number C10X0803) and financial support for this MSc scholarship.

To the Rumen Nutrition and Microbiology Team particularly Dr Sinead Leahy and Dr Eric Altermann for your guidance with the assembly, annotation and all things computational, Dr Adrian Cookson, Dr Christina Moon and Dr Peter Janssen for your input throughout this project.

To the Rumen microbiology lab, namely Dong Li (aka “the machine”), Dr Gemma Henderson and Samantha Noel thank you for showing me the ropes. I’ve learned many things from you.

To the students at AgResearch thank you for all the laughs and conversations, in the end it’s what gets you through the day. I wish you all the best in what I’m sure will be bright futures.

To my love Faith Cox, thank you for keeping me sane and being my anchor throughout this testing time, even when I was at my most “unappetising” self. I hope you know that the work I do is for our future.

To my mum Desanka Palevich and nana Ruzica Radulovic words cannot express the love and strength that you give me to carry on each and every day. It seems as though there is nothing I can give you in return for your gifts, but I hope you know that...

“Everything I do. I do it for you!” (Bryan Adams)

Table of Contents

Abstract.....	i
Acknowledgements.....	ii
Table of Contents.....	iii
Figures.....	vii
Tables.....	x
Abbreviations.....	xii
Chapter 1 Literature Review.....	1
1.1 The importance of ruminant animals to New Zealand agriculture.....	2
1.2 Ruminants.....	2
1.2.1 The ruminant digestive tract.....	2
1.2.2 Ruminant-microbe symbiosis.....	4
1.2.3 Factors limiting digestion of forages in the rumen	5
1.2.4 The rumen microbial ecosystem	5
1.2.5 The fibre degrading rumen microbiome	6
1.3 The Plant Cell Wall	7
1.3.1 Type-I and Type-II plant cell walls.....	7
1.3.2 Cellulose.....	8
1.3.3 Hemicellulose.....	10
1.3.4 Pectin.....	10
1.3.5 Lignin	11
1.4 Lignocellulose degradation	11
1.4.1 Hemicellulose degradation.....	14
1.4.2 Cellulose degradation.....	14
1.4.3 Pectin degradation	14

1.5	Rumen Microbial Genomics	15
1.5.1	The fibrolytic rumen bacteria.....	15
1.5.2	<i>Butyrivibrio</i> and <i>Pseudobutyrvibrio</i>	18
1.5.3	<i>Butyrivibrio proteoclasticus</i> B316	18
1.6	Project Background	20
1.7	Project Aims	22
Chapter 2	Materials and Methods.....	23
2.1	Materials	24
2.1.1	Bacterial strains	24
2.1.2	Buffers and reagents.....	24
2.1.3	Media and media components.....	27
2.1.4	Enzyme buffers and reaction conditions	31
2.1.5	General Laboratory Equipment.....	32
2.1.6	Software	34
2.2	Methods	35
2.2.1	Bacterial isolate selection.....	35
2.2.2	Growth conditions	35
2.2.3	Bacterial growth curves.....	36
2.2.4	16S rRNA gene sequencing	36
2.2.5	Carbon source utilisation.....	36
2.2.6	Culture Storage.....	37
2.2.7	Percentage Guanine plus Cysteine (%G+C) Content	37
2.2.8	Volatile Fatty Acid (VFA) Analysis	37
2.2.9	Adhesion Assays	37
2.2.10	Scanning (SEM) and Transmission Electron Microscopy (TEM).....	38
2.2.11	Motility Assay	39
2.2.12	Pulsed-field gel electrophoresis (PFGE).....	40

2.2.13	<i>Butyrivibrio</i> sp. MB2003 Genome Project.....	41
2.2.14	Gap Closure.....	44
2.2.15	Annotation of the <i>Butyrivibrio</i> sp. MB2003 genome.....	48
2.2.16	Analysis of sequence data	50
Chapter 3	Results.....	51
3.1	Bacterial isolate selection	52
3.2	Characterisation of candidate isolates	55
3.2.1	Microscopic observations of wet mounts and Gram stains.....	56
3.2.2	16S rRNA gene sequencing	57
3.2.3	Phylogenetic analyses	57
3.2.4	PFGE analysis of genomic DNAs.....	58
3.3	Final strain selection and phenotypic characterisation.....	61
3.3.1	Carbon source utilisation.....	62
3.3.2	Insoluble substrate utilisation.....	63
3.3.3	Growth curves for six candidate isolates on optimal substrate (cellobiose).....	63
3.3.4	VFA analysis	64
3.3.5	Moles %G+C content.....	65
3.3.6	Adherence	66
3.3.7	Taxonomic Assignment	69
3.4	Whole-genome sequencing of MB2003	70
3.4.1	Sequence statistics.....	70
3.5	Comparison of the MB2003 genome to <i>B. proteoclasticus</i> B316.....	77
3.5.1	MUMmer plot	77
3.6	Multi-replicon genome architecture	79
3.7	<i>Butyrivibrio hungatei</i> MB2003 genome annotation.....	79
3.7.1	MB2003 Chromid Bhu II.....	80

3.7.2	MB2003 Plasmid pNP6	81
3.7.3	MB2003 Megaplasmid pNP144.....	82
3.7.4	MB2003 Chromosome	83
3.8	Genes involved in polysaccharide breakdown	84
3.9	Xylan degradation in MB2003	88
3.10	Motility	91
3.11	Transmission Electron Microscopy (TEM).....	93
Chapter 4	Discussion.....	96
4.1	MB2003 general genome features.....	98
4.2	Polysaccharide Degradation	100
4.3	Comparison of polysaccharide-degrading strategies in ruminal bacteria.....	106
4.4	Conclusion.....	108
4.5	Future work	109
Appendix I	Gene Lists.....	110
References	149

Figures

Figure 1.1. A phylogeny of ruminant families belonging to order Artiodactyla.....	3
Figure 1.2. Simplified schematic diagram of the ruminant digestive tract.....	4
Figure 1.3. Bacterial diversity of the rumen microbial ecosystem.....	6
Figure 1.4. Simplified chemical structure of cellulose, glucuronoarabinoxylan, xyloglucan, and mixed linkage glucans.....	9
Figure 1.5. Schematic diagram illustrating the major linkages in the forage plant cell wall and the hemicellulolytic and cellulolytic enzymes important for lignocellulose degradation by ruminal microorganisms.....	13
Figure 1.6. 16S rRNA gene phylogeny using near full-length sequences from Clostridial Cluster I and subcluster XIVa strains, with <i>Bacillus subtilis</i> subsp. <i>subtilis</i> as an outgroup.....	19
Figure 1.7. Ultrastructure of <i>B. proteoclasticus</i> B316 cells.....	20
Figure 2.1 Outline of the pre-processing and assembly of the <i>Butyrivibrio</i> sp. MB2003 gap closure sequences.....	47
Figure 3.1. Phylogenetic tree based on the 16S rRNA gene sequences from rumen bacterial isolates cultivated in the NERF study, plant-adherent rumen microbial communities isolated in the ARC study, and a selection of reference rumen bacteria.....	53
Figure 3.2. Outline of the initial phenotypic characterisation experiments of the 20 original isolates.....	55
Figure 3.3. Phylogenetic tree based on 16S rRNA sequence data showing the 16 remaining isolates.....	59
Figure 3.4. 1% Agarose PFGE gels of 18 candidate isolates digested with <i>Apa1</i>	60
Figure 3.5. 1% Agarose PFGE gel of uncut genomic DNA from 10 selected isolates..	61
Figure 3.6. Bacterial growth curves of six candidate isolates grown on cellobiose.....	64

Figure 3.7. Fluorescence microscopy images of DAPI-stained cultures.....	67
Figure 3.8. Scanning Electron Microscopy images of MB2003 cultures.....	68
Figure 3.9. Scanning Electron Microscopy images of MA3014 cultures.....	69
Figure 3.10. Bar chart of read length frequency	71
Figure 3.11. Distribution of contig length.....	72
Figure 3.12. Diagrammatic representation of the MB2003 draft genome.....	76
Figure 3.13. MUMmer plot of MB2003 versus <i>B. proteoclasticus</i> B316 chromosomes.....	78
Figure 3.14. Circular plot of the MB2003 Bhu II chromid.....	80
Figure 3.15. Circular plot of the MB2003 pNP6 plasmid.....	81
Figure 3.16. Circular plot of the MB2003 pNP144 megaplasmid.....	82
Figure 3.17. Circular plot of the draft MB2003 chromosome.....	83
Figure 3.18. Diagrammatic representation of xylan, showing linkages attacked by hemicellulose-degrading enzymes.....	89
Figure 3.19. Comparison of polysaccharide utilisation loci (PUL) in <i>B. hungatei</i> MB2003 with <i>B. proteoclasticus</i> B316.....	90
Figure 3.20. Motility assays and flagella biosynthesis operons in MB2003, MA3014 and B316.....	92
Figure 3.21. Transmission Electron Microscopy images of cultures grown on RM02 media.....	94
Figure 3.22. Transmission Electron Microscopy images of cultures grown on BY media.....	95
Figure 4.1. Hemicellulose structure and attack by <i>Butyrivibrio</i> enzymes.....	103

Figure 4.2. Predicted domain architectures (PFAM) of secreted polysaccharide-degrading enzymes from <i>B. hungatei</i> MB2003.....	104
Figure 4.3. Predicted domain architectures (PFAM) of secreted polysaccharide-degrading enzymes from <i>B. proteoclasticus</i> B316.....	105

Tables

Table 1.1. Estimated composition of Type-I and Type-II primary and secondary plant cell walls.	8
Table 1.2. Summary of the hemicellulolytic and cellulolytic enzymes important for lignocellulose degradation.	12
Table 1.3. Summary of the major lignocellulose degrading bacteria of the bovine rumen.	16
Table 1.4. Publically available genome sequences from fibrolytic rumen bacteria.....	17
Table 2.1. Bacterial strains.....	24
Table 2.2. GenRFV components.....	28
Table 2.3. Trace Element Solution components.....	29
Table 2.4. Vitamin Solution components.....	30
Table 2.5. BY Medium components	30
Table 2.6. RM02 Medium components.....	31
Table 2.7. Centrifuge specifications and suppliers.....	32
Table 2.8. Centrifuge tubes and suppliers.....	32
Table 2.9. Gel Migration Size Standards.....	33
Table 2.10. Software used.....	34
Table 2.11. PCR assembly of reagents.....	45
Table 2.12. Long Range PCR assembly of reagents.....	45
Table 3.1. Summary of occurrence of all 20 isolates within large clusters of 16S rRNA gene sequences derived from pasture-fed cows.....	54
Table 3.2. Microscopic observations of isolates.	56

Table 3.3. RDP and BLASTN search results for 16S rRNA gene sequences of each bacterial isolate.	58
Table 3.4. Carbon source utilisation of the six candidate isolates.	62
Table 3.5. Wet mount observations of insoluble carbon source utilisation of the six candidate isolates.	63
Table 3.6. VFA analysis for the six candidate isolates.	65
Table 3.7. Moles %G+C content for the six candidate strains.....	65
Table 3.8. MB2003 sequence statistics.....	70
Table 3.9. MB2003 contig statistics.	71
Table 3.10. Arrangement of MB2003 contigs into scaffolds.....	74
Table 3.11. Summary of BLASTx search results of contig ends (800bp) for MB2003 genome..	75
Table 3.12. Structure of <i>Butyrivibrio</i> and <i>Pseudobutyrvibrio</i> genomes.....	79
Table 3.13. Sequence information of the MB2003 Bhu II chromid.....	80
Table 3.14. Sequence information of the MB2003 pNP6 plasmid.....	81
Table 3.15. Sequence information of the MB2003 pNP144 megaplasmid.....	82
Table 3.16. Sequence information of the MB2003 chromosome.....	83
Table 3.17. MB2003 genes predicted to encode secreted or intracellular proteins involved in polysaccharide breakdown.	84
Table 3.18. Secreted polysaccharide degrading genes in MB2003 compared with their homologues in <i>B. proteoclasticus</i> B316.	85
Table 3.19. Intracellular polysaccharide degrading genes in MB2003 compared with their homologues in <i>B. proteoclasticus</i> B316.....	86

Abbreviations

bp, kb, Mbp	base pair(s), kilo base(s), mega base pair(s)
CDS	Coding Sequences
COGs	Clusters of Orthologous Groups
DAPI	4', 6-diamidino-2-phenylindole
EC	enzyme classification
EDTA	ethylenediamine tetraacetic acid
EPS	Extracellular Polysaccharides
g	gram(s)
hr	hour(s)
mg	milligram(s)
min	minute(s)
ml	millilitre
mM, M	milliMolar, Molar
mV	milliVolts
NDF	Neutral Detergent Fibre
nt	nucleotide(s)
ORFs	Open Reading Frames
sec	second(s)
TE	Tris-EDTA buffer
v/v	volume : volume ratio
w/v	weight : volume ratio

Chapter 1

Literature Review

1.1 The importance of ruminant animals to New Zealand agriculture

New Zealand agriculture differs from the rest of the world in that livestock are pasture fed, as opposed to being fed with grains or other high energy feed. The difference is in the elevated lignocellulose content of pasture that ruminants find difficult to degrade efficiently. Therefore, a great opportunity exists in New Zealand, to study and improve the efficiency of lignocellulose breakdown of forage-fed ruminant livestock leading to the increased productivity of animal products important to the New Zealand economy. In 2009 almost a third of the world's land area was devoted to agricultural pasture providing forage for more than 1.4 billion cattle, and 2.1 billion sheep and goats (FAOSTAT, 2007). The New Zealand dairy herd is continually expanding, in June 2009 the national herd amounted to 5.9 million head, and with 4.1 million beef cattle this brought the total New Zealand cattle herd to approximately 10 million. In the same year, the New Zealand sheep flock was estimated at 32.4 million head. When the two categories were combined 1.3 million tonnes of meat and approximately 15.2 million litres of milk was produced (FAOSTAT, 2007). The current value of agricultural products, including meat, milk, wool and velvet, to the New Zealand economy is approximately \$13.5 billion, and with the increase in population, demand for meat and milk products can only be expected to increase (FAOSTAT, 2007).

1.2 Ruminants

1.2.1 The ruminant digestive tract

Ruminant animals belong to the order Artiodactyla which includes animals such as cattle, deer, sheep and goats along with many other wild ruminants (Figure 1.1). Ruminants digest plant material by initially chewing it and digesting it within the animal's first two stomachs (the reticulum and rumen, referred to collectively as the reticulorumen, or commonly just as "the rumen"). This is assisted by regurgitating the semi-digested material (known as the cud), and rechewing it. The word "ruminant" in fact comes from the Latin *ruminare*, which means "to chew over again". The cud is most accurately described as a bolus of semi-degraded food regurgitated from the reticulorumen of a ruminant, while the actual process of re-chewing the cud to further break down plant matter and stimulate digestion is called "rumination".

Most rumen microbiology studies to date have focused on the agriculturally important ruminants such as cattle and sheep that belong to the Bovidae family and deer belonging to the Cervidae family, as shown in Figure 1.1 below.

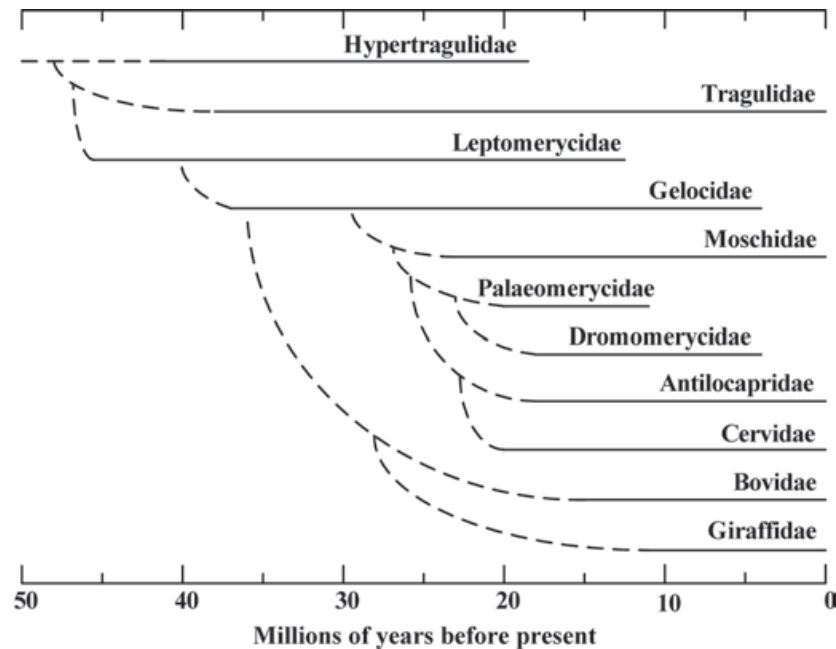
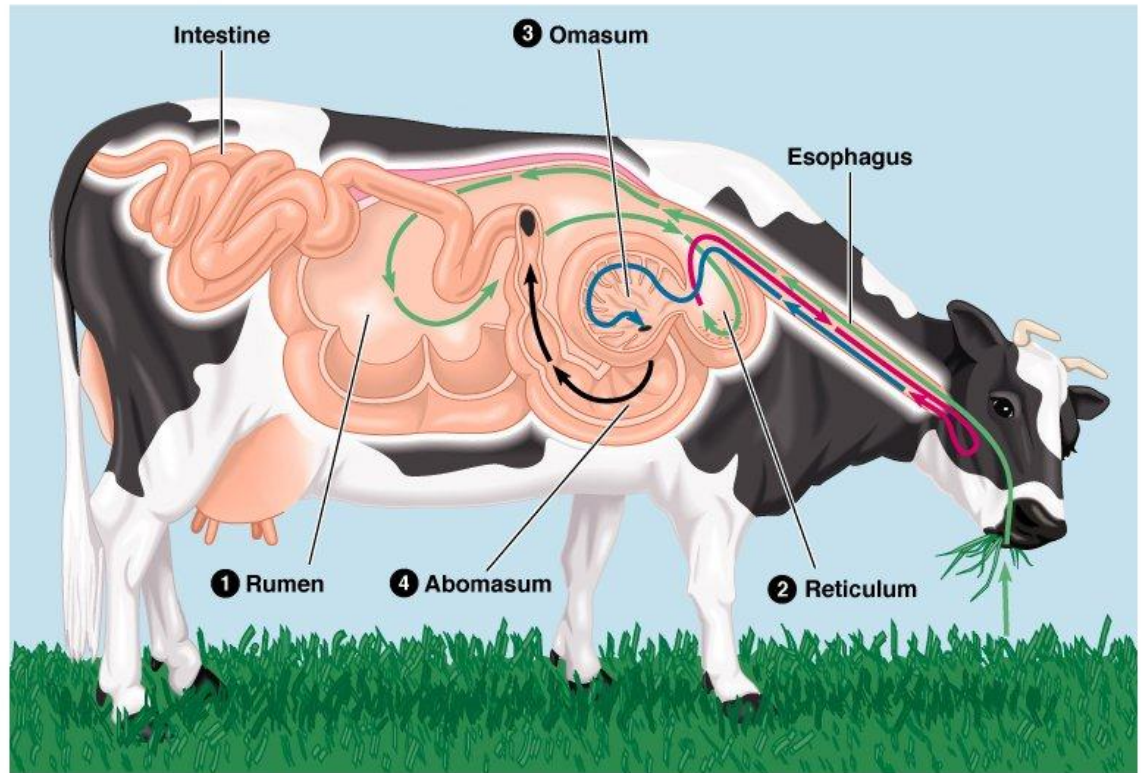


Figure 1.1. A phylogeny of ruminant families belonging to order Artiodactyla. Solid lines indicate age ranges documented in the fossil record and stippled lines indicate inferred age ranges and family relationships (Hackmann and Spain, 2010).

The typical digestive tract of a cow is shown in Figure 1.2. At approximately 100 L in cattle and up to 10 L in sheep, the rumen is the first and largest of the fore-stomach compartments. The reticulum is considered a simple rumen extension, and is connected to the omasum by a short tunnel-like structure called the reticulo-omasal orifice. The abomasum is the ruminant's true stomach; histologically it is similar to that of monogastric mammals (Flint *et al.*, 2008). In the rumen, plant material is mixed with saliva and separates into solid and liquid layers. The solid plant material clumps together and is buoyed up by trapped gas bubbles so that it floats to the surface of the reticulorumen to form a raft of material. The proximal location of the rumen facilitates the regurgitation of partially digested lignocellulose for further mastication. Occasionally, a portion of the solid raft material is regurgitated into the mouth, and chewed to assist in reducing its particle size. This process triggers the flow of copious amounts of saliva from the mouth to the fore-stomach, which under ideal conditions buffers the rumen at approximately pH 6.8 and is essential for efficient rumen function.



©1999 Addison Wesley Longman, Inc.

Figure 1.2. Simplified schematic diagram of the ruminant digestive tract. The rumen is the first and largest fore-stomach compartment, and the primary site of microbial mediated lignocellulose degradation and fermentation. Digesta flows from the rumen to the reticulum, omasum and abomasum, respectively. (www.mun.ca/biology/scarr/Ruminant_Digestion.htm)

1.2.2 Ruminant-microbe symbiosis

A unique symbiosis exists between the ruminant and the microorganisms within the rumen environment. Ruminants are unable to digest lignocellulose, which makes up the main component of their diet. Lignocellulose is the most abundant organic material on Earth and consists of a matrix of cross-linked cellulose and hemicellulose networks, glycosylated proteins, and lignin polymers (Wang *et al.*, 2011). The microbes in the reticulorumen carry out lignocellulose degradation and fermentation within the rumen (Nocek and Russell, 1988), and in doing so generate large quantities of short chain volatile fatty acids (SCVFAs) which in turn support growth and cell turnover of the host animal. The main SCVFAs produced are acetic, propionic and butyric acids in an approximate ratio of 60:30:10 respectively. The host absorbs these SCVFAs across the rumen epithelium and transports them to the liver, where they are metabolised. Acetic acid is the major source of acetyl-CoA for lipid synthesis which is oxidised to generate

ATP in nearly all tissues of the ruminant. In order to produce energy from butyric acid, β -hydroxybutyric acid must be absorbed from the rumen epithelium and then undergoes oxidation in the appropriate tissues. Propionic acid is a major substrate for gluconeogenesis and is thus almost exclusively utilised in the liver. SCVFA production supplies more than 70% of the total energy requirements for the ruminant (Flint *et al.*, 2008). The ruminant obtains over half of its total protein requirements from the microbial biomass that is flushed from the rumen to the small intestine. Some of the ingested protein from plant material can be used directly by the ruminant if it by-passes degradation in the rumen (Ulyatt *et al.*, 1980), but most of the ingested plant protein is quickly broken down to peptides, amino acids and ammonia, which is then assimilated into microbial protein.

1.2.3 Factors limiting digestion of forages in the rumen

Four main factors have been recognised as restricting lignocellulose degradation in the rumen (Cheng *et al.*, 1991): (i) host factors that mediate the availability of nutrients through mastication, salivation and digesta kinetics (i.e. the fibre intake rate and extent of mastication by ruminants affects the rate of passage through the digestive tract, with higher intakes resulting in lower overall fibre digestion), (ii) the chemical composition and degradability of the insoluble component of plant material, (iii) the composition of the rumen microbial ecosystem and (iv) the polysaccharolytic capability of the rumen microbial ecosystem.

1.2.4 The rumen microbial ecosystem

Culture-dependent studies on rumen microbiota have identified five fungal species, at least 370 bacterial and archaeal species, and 40 protozoal species that work in concert to degrade and ferment lignocellulosic biomass (Orpin and Joblin, 1997). On the other hand, cultivation-independent molecular surveys looking at the microbial diversity within the rumen ecosystem have revealed that only 11% of the total bacterial taxa have been cultured (Tajima *et al.*, 1999; Edwards *et al.*, 2004; Shin *et al.*, 2004a; Shin *et al.*, 2004b; Shin *et al.*, 2004c; Cho *et al.*, 2006; Skillman *et al.*, 2006; Wright *et al.*, 2007). It is believed that the number of uncultivated individual species within the rumen range from some hundreds to thousands, with the majority categorised as being Gram-positive *Firmicutes* or the Gram-negative *Bacteroidetes* (Figure 1.3). Studies to date have shown

that the current cultured and completely characterised rumen microorganisms do not fully represent the principal and metabolically significant components of the rumen microbiota (Attwood *et al.*, 2008). As a result, our understanding of rumen fibre degradation and the enzyme systems involved in lignocellulose degradation is not fully developed. The discovery and complete characterisation of novel rumen microbes that are proficient at degrading lignocellulosic material will provide a much needed understanding of the metabolic processes involved in plant cell wall degradation.

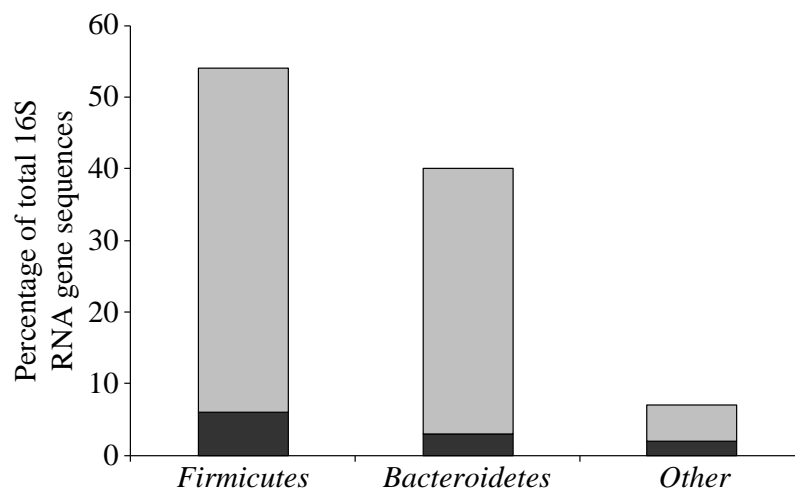


Figure 1.3. Bacterial diversity of the rumen microbial ecosystem. The percentage of 16S rRNA gene sequences represented by cultivated (black) and uncultivated (grey) bacteria is shown for the two predominant phyla, and minor or unknown phyla. Adapted from Edwards *et al.*, (2008).

1.2.5 The fibre degrading rumen microbiome

The term rumen microbiome describes the set of rumen microbes, their genomes, and environmental interactions, and is composed of: (i) microbes found free living in the rumen liquor (planktonic phase), (ii) microbes loosely or tightly associated with ingested plant material (adherent phase) and (iii) microbes associated with the rumen epithelium. The adherence of microbes to plant material is crucial in order to maximise efficient degradation of plant cell wall polysaccharides (McAllister *et al.*, 1994). It has been shown that bacteria attach to plant material within minutes of ingestion (Bonhomme, 1990; Koike *et al.*, 2003a). Of the total microbial biomass in the rumen, ~80% are fibre adherent bacteria; they account for 90% of the total cellulase and xylanase activities, as well as ~70% of the amylase and protease activities in the rumen (Akin, 1980; Craig *et al.*, 1987; Minato *et al.*, 1993). Molecular surveys have demonstrated that *Bacteroidetes* mainly occupy the planktonic phase whereas

Firmicutes account for up to 75% of the adherent bacterial population (Tajima *et al.*, 1999; Cho *et al.*, 2006).

1.3 The Plant Cell Wall

The plant cell wall is the foremost source of nutrition for forage fed ruminants and constitutes approximately 80% of the dry weight of plant material. The cell wall is composed of a complex network of cellulose fibres surrounded by a matrix of non-cellulosic structural polysaccharides, proteins, phenolic compounds and lignin. The structures of the major cell wall polysaccharides are shown in Figure 1.4. Chemical composition varies greatly, depending on the plant species, age and the tissue type (Carpita and Gibeaut, 1993; Carpita, 1996).

1.3.1 Type-I and Type-II plant cell walls

The primary cell wall of flowering plants can be separated into two classes, which differ significantly in their structural features and chemical composition. Type-I cell walls are made up of mainly cellulose microfibrils fixed in a xyloglucan (XyG)-rich hemicellulose, along with significant amounts of pectin and structural proteins (Table 1.1) (Carpita, 1996). They are abundant in both dicotyledonous and noncommelinoid monocotyledonous plants as well as in gymnosperms. Type-II cell walls are found only in commelinoid monocotyledonous plants (i.e. grasses) and differ in that they are comprised of cellulose embedded in very significant amounts of glucuronoarabinoxylans (GAX) with mixed linkage glucans (MLGs), and modest levels of phenolic compounds (Carpita, 1996). In contrast to Type-I cell walls, Type II cell walls contain low levels of pectin and structural proteins. The secondary cell wall of grasses is mainly composed of cellulose, GAX, and lignin; it can contribute up to 50% of dry weight (Brown and Saxena, 2000). The two predominant plant types found in New Zealand pastures are perennial ryegrass and legumes (Woodfield, 1999). Perennial ryegrass is composed of Type-II cell walls with high levels of GAX and ferulic acid crosslinks, whereas legumes (clover and lucerne) have Type-I cell walls that are high in pectin and xyloglucans (Varner and Lin, 1989).

1.3.2 Cellulose

The plant cell wall is comprised of approximately 40% cellulose, an energy rich structural polysaccharide (O'Sullivan, 1997). Cellulose is composed of β 1-4 linked glucose units that spontaneously form insoluble microfibrils (approximately 3-5 nm wide) of single strands of up to 15,000 glucose monomers (Figure 1.4, Vogel, 2008). Individual microfibrils contain variable numbers of cellulose chains and may also associate with other microfibrils to form a larger single microfiber (Pizzi and Eaton, 1985). These structures are held together via numerous inter- and intra-chain hydrogen bonds and van der Waals forces (Ioelovich, 2008). The exterior of cellulose microfibrils is much less rigid compared to the core regions due to the imperfections in bonds of the individual cellulose chains, the actual thickness of the microfibril or possibly through the interactions with other molecules surrounding the polymers (Ioelovich, 2008). The ability of cellulose to form strong bonds with a variety of non-cellulosic cell wall components accounts for the structural rigidity and tensile strength of the plant cell wall (Baskin, 2001).

Table 1.1 Estimated composition of Type-I and Type-II primary and secondary plant cell walls.

Cell wall component	Primary wall ^a		Secondary wall ^a	
	Type-I	Type-II	Type-I	Type-II
Cellulose	15-30 ^{b,c,d}	20-30 ^{b,e}	45-50 ^b	35-45 ^{b,f}
Hemicellulose				
Glucuronoarabinoxylans	5 ^b	20-40 ^c	20-30 ^{b,g}	40-50 ^{b,g}
Mixed linkage glucans	None	10-30 ^c	None	Minor
Xyloglucans	20-25	1-5 ^{b,c}	Minor	Minor
Mannans/glucomannans	5-10 ^c	Minor	3-5 ^g	Minor
Pectin	20-35 ^c	5 ^b	0.1 ^b	0.1 ^b
Structural proteins	10 ^{c,d}	1 ^c	Minor	Minor
Phenolic compounds	Minor	1-5 ^{b,c}	Minor	0.5-1.5 ^b
Lignin	Minor	Minor	7-10 ^b	20 ^b
Silica	None	None	Variable	5-15 ^b

^a Values are the percentage of plant cell-wall dry weight. The table was adapted from (Vogel, 2008). Estimates were obtained from a number of sources as detailed. ^b(Ishii, 1997), ^c(O'Neil and York, 2003), ^d(Zablackis *et al.*, 1995), ^e(Mitchell *et al.*, 2007), ^f(Hatfield *et al.*, 1999) and ^g(Ebringerova *et al.*, 2005).

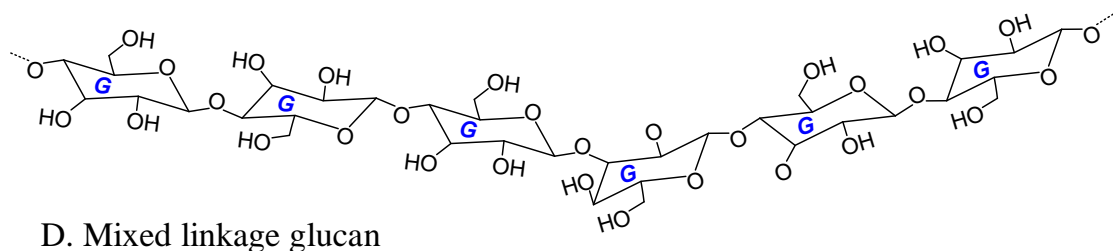
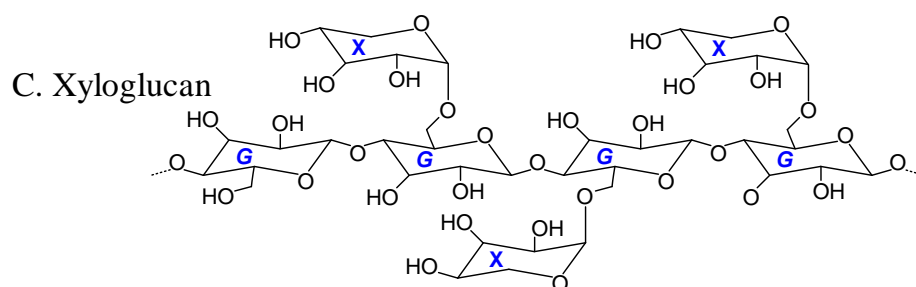
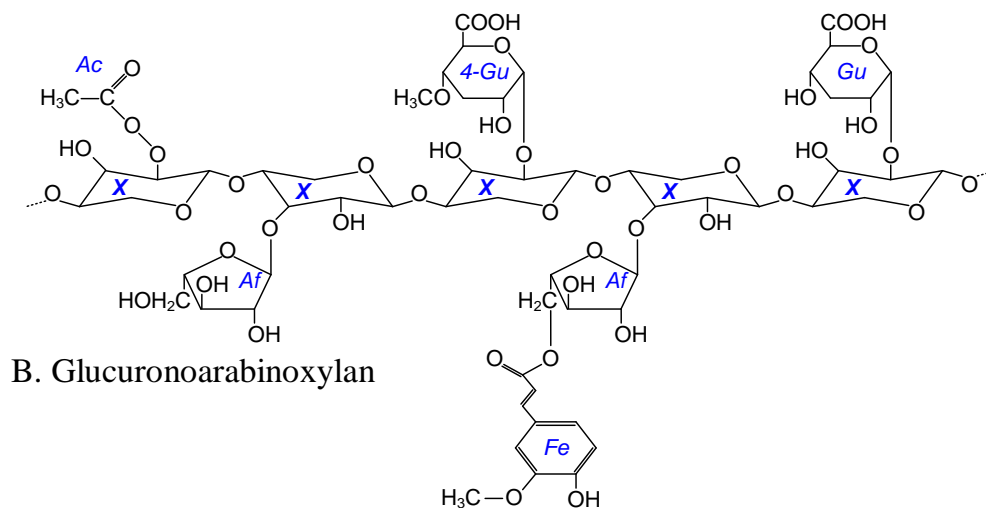
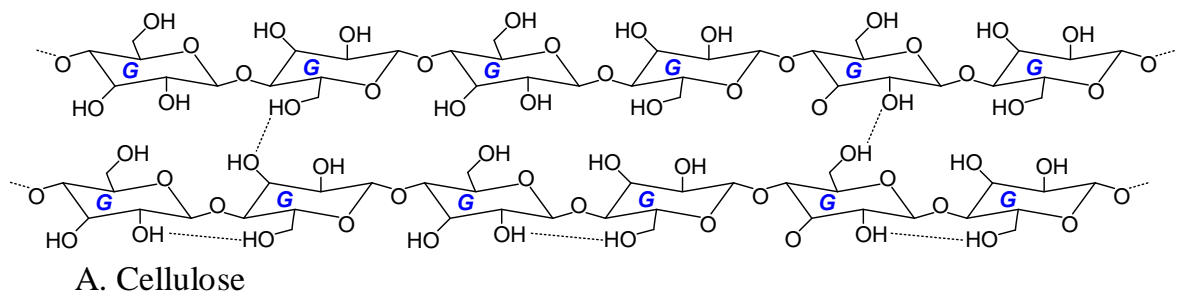


Figure 1.4. Simplified chemical structure of cellulose, glucuronoarabinoxylan, xyloglucan, and mixed linkage glucans. (A) Two chains of cellulose consisting of (1→4)-β-D-glucose monomers with intra- and inter-chain hydrogen bonding (dotted lines). (B) The homopolymeric xylan backbone composed of (1→4)-β-D-xylopyranose monomers with examples of the substituent side chains found in the cell wall of grasses. (C) The four residue repeating pattern of xyloglucan with examples of xylopyranose substituent groups. (D) A mixed linkage glucan chain comprised of two (1→4)-β-cellobiose units connected by a (1→3)-β-linkage. Ac, O-acetyl; Af, α-L-arabinofuranose; Fe, ferulic acid; G, glucose; 4-Gu, 4-O-methyl-D-glucuronic acid; Gu, glucuronic acid; X, xylopyranose (xylose). Figure adapted from Vogel, 2008.

1.3.3 Hemicellulose

Hemicellulose is semi-soluble (i.e. unable to completely dissolve in water) and makes up 20%-35% of the total plant biomass. Unlike cellulose, it is a family of highly branched chemically heterogeneous polymers of pentoses, hexoses and sugar acids (Linder *et al.*, 2003; Kabel *et al.*, 2007). Hemicellulose interacts with neighbouring cellulose microfibrils (and other non-cellulosic cell wall structural polymers) in a structurally dependent manner, according to the sticky network model (Cosgrove, 2000). In this model, hemicelluloses are either tethered to the cellulose microfibrils by formation of multiple hydrogen bonds between the two polymers after deposition or become trapped during crystallisation. The two main representatives of hemicelluloses are GAX and xyloglucans (Figure 1.4). GAX in grasses is comprised of a (1→4)-β-D-xylopyranose monomer backbone (Aspinall and Preiss, 1980; Liab *et al.*, 2000) that can be substituted with a number of different groups such as (1→2)-α-linked and (1→3)-α-linked L-arabinofuranosyl groups, *O*-acetyl, and (1→2)-α-D-glucuronic acid or 4-*O*-methyl-D-glucuronic acid groups to varying degrees (Borneman *et al.*, 1990). The α-L-arabinofuranosyl side groups may also form diferulic crosslinks with phenolic polymers such as ferulic acid by becoming esterified at the *O*-2 or *O*-5 positions. Xyloglucan is the main hemicellulose in the primary cell wall of many dicotyledonous plants and consists of a repeating pattern of four (1→4)-β-D-glucose monomers substituted with (1→6)-α-xylopyranose residues that are tightly hydrogen bonded to neighbouring cellulose microfibrils (Pauly *et al.*, 1999).

1.3.4 Pectin

Pectin is structurally the most intricate (i.e. complex and difficult to degrade) family of polysaccharides because it is involved in the control of the cell wall structure and expansion, cell-cell interactions and signalling, as well as in plant defence mechanisms (Jarvis, 1984). Pectin is mainly located in the primary cell wall, featuring in two distinct types of polysaccharides: homogalacturonan (HG) and rhamnogalacturonan I (RG-I). HG is a linear chain of (1→4)-α-D-galacturonic acid (GalA) residues that may be partially methylesterified or *O*-acetylated (Shibuya and Nakane, 1984; Carpita, 1989). Rhamnogalacturonan I (RG-I) is a more complex heteropolymer comprised of a backbone of repeating (1→2)-α-L-rhamnosyl-(1→4)-α-D-GalA subunits, that can be profoundly substituted with complex side chains consisting of arabinofuranosyl units

and arabinogalactans (Saulnier and Thibault, 1999). Arabinogalactans are by nature heterogeneous but may consist of a backbone of (1→3)-β-linked D-galactose residues substituted with (1→6)-β-D-galactose and/or (1→3)-α-L-arabinofuranose. Furthermore, similar to GAX, the side chains of RG-I may be esterified with ferulic and *p*-coumaric acids (Nakamura *et al.*, 2002b; Coenen *et al.*, 2007). It is believed that pectic polysaccharides are covalently cross-linked to one another via their backbones. Thus it is possible that both HG and RG-I are very tightly associated with other plant cell wall polysaccharides including hemicellulose and cellulose (Nakamura *et al.*, 2002a).

1.3.5 Lignin

Lignin is a crucial secondary cell wall component as it plays an important role in maintaining cell structure integrity, water transport and protecting against plant pathogens (Grabber, 2005). Lignin is not susceptible to enzymatic degradation by the rumen microbiota, and hence its presence in the cell wall is thought to be partly responsible for the recalcitrance of ruminant forages to breakdown in the rumen (Armstrong and Gilbert, 1985; McAllister *et al.*, 1994). In dicotyledonous plants, lignins contain mainly guaiacyl and syringyl subunits and a small quantity of *p*-hydroxyphenyl units (Campbell and Sederoff, 1996). An important attribute of grass lignins is the incorporation of diferulic acids during development by means of ether bonds linking the ferulic acid hydroxyl groups forming diferulic bonds (Scalbert *et al.*, 1985; Kondo *et al.*, 1990).

1.4 Lignocellulose degradation

For complete degradation of lignocellulosic material, the actions of three types of polysaccharide degrading enzymes (polysaccharidases) are required to work in concert. These enzymes are: (i) *O*-Glycosidehydrolases (GH), which hydrolyse glycosidic bonds between carbohydrate monomers or between carbohydrate and non-carbohydrate moieties and this activity takes place via general acid catalysis leading to an overall retention or inversion at the site of catalytic activity (Henrissat, 1991; Henrissat and Davies, 1997; Henrissat, 1998), (ii) Carbohydrate Esterases (CE), which hydrolyse the ester linkages of ferulic acid or acetate side chains and (iii) Polysaccharide Lyases (PL), which hydrolyse polysaccharide chains by β-elimination. Those lignocellulose-

degrading enzymes important for rumen forage degradation and their substrates are summarised in Table 1.2 and Figure 1.5.

Table 1.2. Summary of the hemicellulolytic and cellulolytic enzymes important for lignocellulose degradation.

Enzyme	Substrate
Endo- β -1,4-xylanase	(1 \rightarrow 4)- β -D-xylan
Exo- β -1,4-xylosidase (β -xylosidase)	(1 \rightarrow 4)- β -D-xylooligomers, xylobiose
α -L-arabinofuranosidase	α -L-arabinofuranosyl (1 \rightarrow 2) or (1 \rightarrow 3) xylooligomers
α -Glucuronidase	4- <i>O</i> -methyl-D-glucuronic acid (substituted to xylan)
Acetyl xylan esterase	<i>O</i> -acetyl xylan
Ferulic/ <i>p</i> -coumaric acid esterase	Ferulic/ <i>p</i> -coumaric acid esterified arabinofuranosyl xylooligomers
α -Galactosidase	(1 \rightarrow 6)- α -D-galactose
Endo- β -1,4-glucanase	(1 \rightarrow 4)- β -D-glucan/cellulose
Cellodextrinase Cellobiohydrolase	(1 \rightarrow 4)- β -D-glucose/cellobiose
β -Glucosidase	(1 \rightarrow 4)- β -D-cellodextrins and cellobiose

Descriptions and classifications from the CAZy database (Henrissat. *et al.*, 2011).

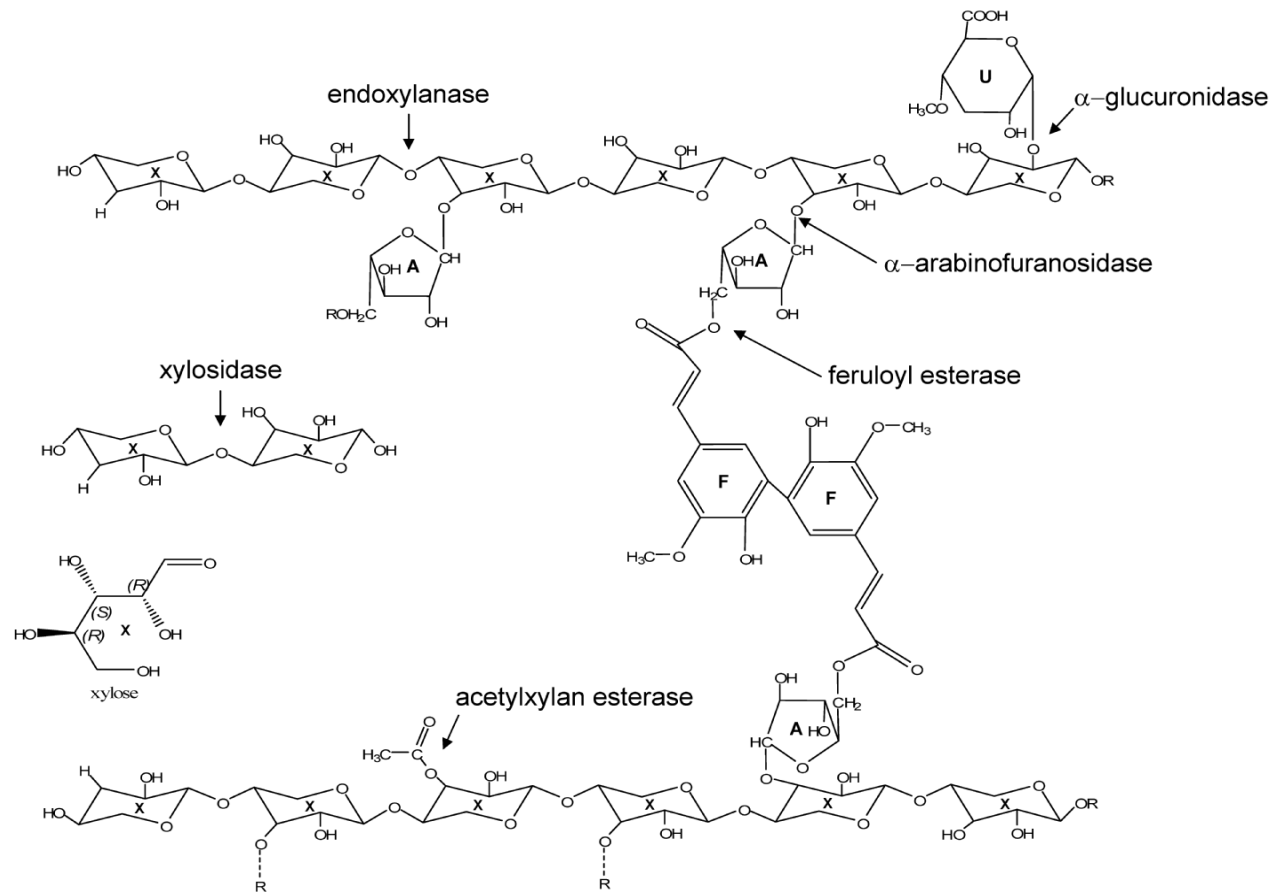


Figure 1.5. Schematic diagram illustrating the major linkages in the forage plant cell wall and the hemicellulolytic and cellulolytic enzymes important for lignocellulose degradation by ruminal microorganisms. A, arabinose; F, ferulic acid; U, 4-O-methylglucuronic acid; X, xylose; R, continuation of glycan chain. The major enzymes cleaving the plant cell wall include: endoxylanase, xylosidase, acetylxylan esterase, α -arabinofuranosidase, feruloyl esterase, α -glucuronidase. Modified image from Chesson and Forsberg, 1997.

1.4.1 Hemicellulose degradation

Due to the complex nature and diverse chemical structure, hydrolysis of hemicelluloses requires a variety of enzymes. Endoxylanases ((1→4)-β-D-xylanxylanohydrolases) cleave the internal linkages within the xylan backbone forming xylobiose and xylooligosaccharides. These end products are further degraded into xylobiose and/or xylose by β-D-xylosidases ((1→4)-β-D-xylan-xylohydrolases). The diverse range of substituents attached to hemicellulose are hydrolysed from the xylan backbone by the activity of α-D-glucuronidases, α-L-arabinofuranosidases, acetylxylan esterases, ferulic acid and *p*-coumaric acid esterases (Mackenzie *et al.*, 1987). Ferulic acid esterases may be vital in ruminant fibre degradation as they are thought to increase accessibility of hemicellulose to other fibre degrading enzymes by hydrolysing the ester bond between ferulic acids and hemicellulose (Black *et al.*, 1996; Janecek *et al.*, 2003).

1.4.2 Cellulose degradation

Three types of enzymes catalyse cellulose hydrolysis to glucose: (i) Endo-β-1,4-glucanases cleave (1→4)-β-D-glycosidic bonds at internal positions along the cellulose chain generating cellodextrins of variable chain lengths, (ii) Exoglucanases such as cellodextrinases and cellobiohydrolases act in a progressive manner on the reducing or non-reducing ends of cellulose to liberate glucose and cellobiose. Exoglucanases are also capable of degrading microcrystalline cellulose by stripping cellulose microfibrils from the microcrystalline structure (Teeri, 1997), and (iii) β-glucosidases complete the degradation process by hydrolysing soluble cellobiose and cellodextrins to glucose.

1.4.3 Pectin degradation

Pectin degradation is facilitated by three classes of enzymes: (i) Depolymerases that act directly upon the pectin backbone, and are subdivided into various lyases (acting by β-elimination). Pectate lyases cleave (1→4)-α-linkages between galacturonic acids and may preferentially attack regions of unmethylated polygalacturonate (pectate), although they can act on regions with low levels of methylation (Tardy *et al.*, 1997). (ii) Hydrolases involved in pectin degradation include endopolygalacturonases, exopolygalacturonases, exo-poly-α-galacturonases and rhamnogalacturonases. Polygalacturonans are cleaved at random internal positions, and at the non-reducing

chain ends in polygalacturonans. Rhamnogalacturonases hydrolyse bonds between D-galacturonic acid and L-rhamnose in rhamnogalacturonan I (Sakai *et al.*, 1993). (iii) Esterases that hydrolyse acetyl, methyl or ferulic acids substituents from the main chains of the (1→4)- α -galacturonic acids. Acetylerases remove C-2 and C-3 attached acetyl groups, and methylesterases act on methoxylated groups releasing methanol. Acetylerase activity is enhanced when esterified pectins are pre-treated with methylesterase, suggesting there is an optimal sequence of esterase activities for pectin degradation (Shevchik and Hugouvieux-Cotte-Pattat, 2003).

1.5 Rumen Microbial Genomics

1.5.1 The fibrolytic rumen bacteria

To date our understanding of the various enzymatic processes that drive rumen forage degradation have been derived from molecular and biochemical analyses of single genes and/or gene products produced by microbes that have been grown in pure culture. Only a select few organisms are able to both efficiently breakdown and utilise the non-cellulosic components of the plant cell walls such as pectin and hemicellulose (Stewart *et al.*, 1997). Based on current knowledge, the predominant fibrolytic bacteria within the rumen of animals fed a high forage diet are summarised in Table 1.3 (Table adapted from Weimer, 1998). Genome sequence information concerning rumen microorganisms is scarce, as complete genomes are available for only four fibrolytic bacteria (Table 1.4). Many interesting discoveries have been made from the various fibrolytic bacterial genomes, among these are a large number of genes that code for proteins that facilitate breakdown of plant cell wall polysaccharides.

Table 1.3. Summary of the major lignocellulose degrading bacteria of the bovine rumen.

Phylum	Species	Predominant Substrate	Main fermentation products	Reference		
Firmicutes (Gram-positive)	<i>Butyrivibrio fibrisolvens</i>	X, P, S	butyrate/formate/lactate	(Kopečný <i>et al.</i> , 2003)		
	<i>Butyrivibrio hungatei</i>			(Kopečný <i>et al.</i> , 2003)		
	<i>Butyrivibrio proteoclasticus</i>			(Moon <i>et al.</i> , 2008)		
	<i>Pseudobutyrvibrio ruminis</i>			(Schoep and Gregg, 2007)		
	<i>Pseudobutyrvibrio xylanovorans</i>			(Kopečný <i>et al.</i> , 2003)		
	<i>Lachnospira</i> sp.			P	(Cotta and Forster, 2006)	
	<i>Ruminococcus flavefaciens</i>			C, X	formate/succinate	(Ayers, 1958)
	<i>Ruminococcus albus</i>			C, X	formate/succinate	(Patterson <i>et al.</i> , 1975)
	<i>Eubacterium cellulosolvens</i>			C, X	lactate/formate	(Anderson and Blair, 1996)
	<i>Clostridium</i> sp.					(Kelly <i>et al.</i> , 1987)
Bacteroidetes (Gram-negative)	<i>Prevotella ruminicola</i>	X, S, P	propionate/succinate	(Purushe <i>et al.</i> , 2010)		
	<i>Prevotella bryantii</i>	X, S, P	propionate/succinate	(Purushe <i>et al.</i> , 2010)		
Fibrobacter (Gram-negative)	<i>Fibrobacter succinogenes</i>	C	formate/succinate	(Suen <i>et al.</i> , 2011)		
	<i>Fibrobacter intestinalis</i>	C	formate/succinate	(Qi <i>et al.</i> , 2005)		

X, xylan, S, starch, C, cellulose and P, pectin.

Table 1.4. Publicly available genome sequences from fibrolytic rumen bacteria.

Organism	Contigs	Size (bp)	G+C%	Family	Sequencing Project*	Country	NCBI Number (NC)	Reference
<i>Butyrivibrio proteoclasticus</i> B316	1	3,555,059	40	Lachnospiraceae	AgResearch	NZ	014387	(Kelly <i>et al.</i> , 2010)
	1	361,399	39				014389	
	1	302,355	40				014388	
	1	186,328	38				014390	
<i>Prevotella ruminicola</i> 23	1	3,619,559	47	Prevotellaceae	NACGFRB	USA	014033	(Purushe <i>et al.</i> , 2010)
<i>Fibrobacter succinogenes</i> S85	1	3,842,635	48	Fibrobacteraceae	JGI/NACGFRB	USA	013410	(Suen <i>et al.</i> , 2011)
<i>Ruminococcus albus</i> 7	1	3,685,408	44	Ruminococcaceae	JGI	USA	014833	Unpublished
	1	420,706	38				014824	
	1	352,646	44				014825	
	1	15,907	36				014826	
	1	7,420	42				014827	
<i>Prevotella bryantii</i> B14	98	3,592,947	39	Prevotellaceae	NACGFRB	USA	ADWO00000000	(Purushe <i>et al.</i> , 2010)
<i>Ruminococcus albus</i> 8	245	4,373,730	46	Ruminococcaceae	NACGFRB	USA	ADKM00000000	Unpublished
<i>Ruminococcus flavefaciens</i> FD-1	119	4,573,608	45	Ruminococcaceae	NACGFRB	USA	ACOK00000000	(Miller <i>et al.</i> , 2009)
<i>Eubacterium cellulosolvens</i> 6	107	3,260,436	48	Lachnospiraceae	JGI	USA	AEOA00000000	Unpublished

*NACGFRB, North American Consortium for the Genomics of Fibrolytic Rumen Bacteria. JGI, Joint Genome Institute. www.ncbi.nlm.nih.gov

1.5.2 *Butyrivibrio* and *Pseudobutyrvibrio*

The genera *Butyrivibrio* and *Pseudobutyrvibrio* belong to the genetically diverse *Lachnospiraceae* family and previously were all called *Butyrivibrio fibrisolvens*. However they have been divided into two genera and several species (Orpin *et al.*, 1985; Willems *et al.*, 1996; Forster *et al.*, 1997; Kopecny *et al.*, 2001; Kopecny *et al.*, 2003; Kim *et al.*, 2011). Bacteria of the *Butyrivibrio* and *Pseudobutyrvibrio* assemblage are thought to be the main butyrate producers in the rumen and are considered amongst the most effective hemicellulose degraders (Diez-Gonzalez *et al.*, 1999; Paillard *et al.*, 2007a). Xylans of different chemical and physical properties from a range of forages can be degraded by *Butyrivibrio* strains (Coen and Dehority, 1970; Miron and Benghedalia, 1993; Miron *et al.*, 1994; Hespell and Cotta, 1995). *Butyrivibrio* strains are capable of growing on a range of simple sugars and plant polysaccharides such as pectins, mannans, starch and hemicelluloses. While they can produce a number of cellulose degrading enzymes, not all strains are able to utilise cellulose *in vitro* (Attwood and Reilly, 1995). Recent studies have shown that the bulk of the *Butyrivibrio* population in the rumen actually adheres to the solids fraction in rumen contents (Whitford *et al.*, 1998; Tajima *et al.*, 1999; Koike *et al.*, 2003b), however the mechanisms responsible for fibre adhesion are currently not understood. Another key feature of *Butyrivibrio* species is their high proteolytic activity (Cotta and Hespell, 1986; Attwood and Reilly, 1995; Sales *et al.*, 2000).

1.5.3 *Butyrivibrio proteoclasticus* B316

Butyrivibrio proteoclasticus B316 was isolated from rumen contents of a New Zealand cow grazing a ryegrass-clover pasture (Attwood *et al.*, 1996). It was originally described as *Clostridium proteoclasticum*, however further taxonomic studies indicated that its reclassification into the genus *Butyrivibrio* (within the *Clostridial* subcluster XIVa) was more appropriate, as shown in Figure 1.6 (Attwood *et al.*, 1996; Moon *et al.*, 2008). The *B. proteoclasticus* B316 genome is 4.4 Mb in size and encodes 3813 coding sequences (CDSs) that are distributed across four replicons, consisting of the main chromosome (BPc1), two megaplastids (pCY360 and pCY186) and a secondary chromosome, or chromid (BPc2). *B. proteoclasticus* has a Gram-positive cell wall structure even though cultures stain Gram-negative. Although cells have a single sub-polar flagellum, *B. proteoclasticus* is not motile (Kelly *et al.*, 2010).

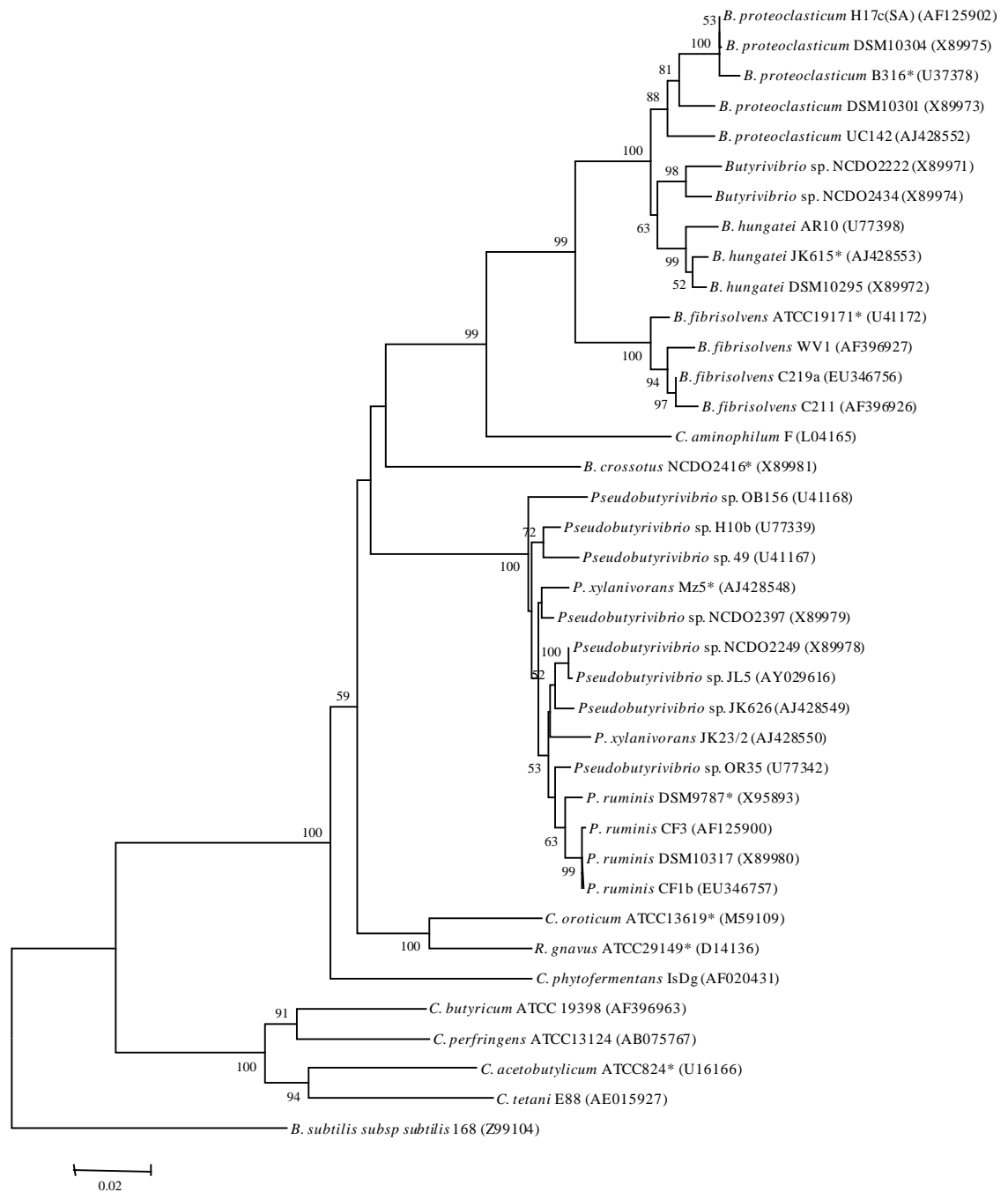


Figure 1.6. 16S rRNA gene phylogeny using near full-length sequences from Clostridial Cluster I and subcluster XIVa strains, with *Bacillus subtilis* sub sp. *subtilis* as an outgroup. Percentage bootstrap values were calculated from 10,000 replicates and are shown at nodes if >50 %. GenBank accession numbers are shown in parentheses. Bar, 0.02 nucleotide substitutions per site. * indicates type strain. Figure adapted from Moon *et al.*, 2008.

B. proteoclasticus cells are slightly curved rods that form short chains or pairs but will occasionally also form long chains (Figure 1.7). *B. proteoclasticus* is prevalent in the rumen of animals grazing fresh pasture. Estimates of its populations via competitive PCR or real time PCR are ~2% and 9% respectively (Reilly and Attwood, 1998; Paillard *et al.*, 2007b). *B. proteoclasticus* produces butyrate, is highly xylanolytic and proteolytic and can also convert linoleic acid into stearic acid and thus may have an important role in biohydrogenation in the rumen (Attwood *et al.*, 1996).

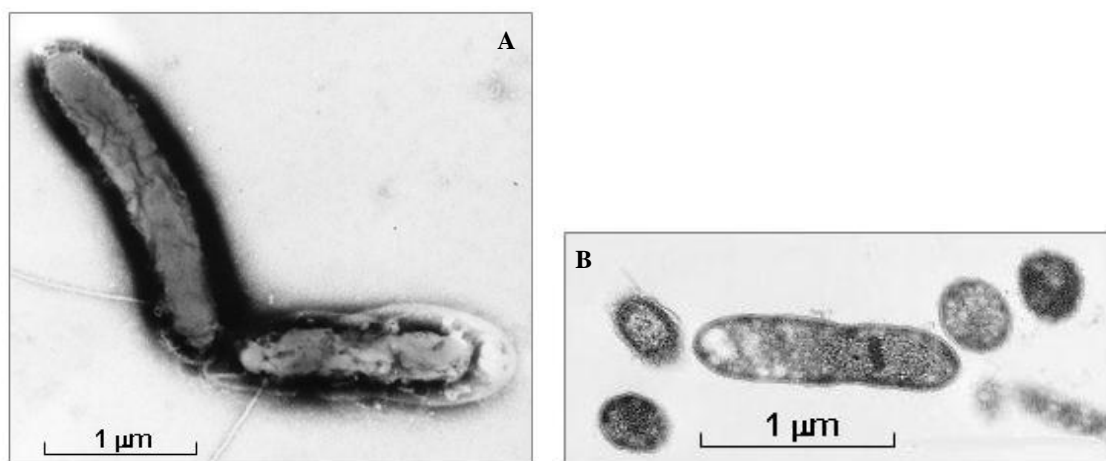


Figure 1.7. Ultrastructure of *B. proteoclasticus* B316 cells. Electron micrographs of negatively stained *B. proteoclasticus* cells showing the sub-polar flagellum (A), and thin-sectioned cells showing the gram-positive cell wall structure (B). Figure from Attwood *et al.*, 1996.

1.6 Project Background

The AgResearch Research and Capability Fund (ARC; “Development of a functional metagenomics capability for the analysis of fibre-degrading rumen microbiota”, contract number ARC0719) project was aimed at developing culture-independent methods to analyse the rumen microbiota, focusing on the fibre-adherent microbes. Three types of data were generated from this work: (i) 16S rRNA gene sequence that provided insight into the microbial community composition of the rumen, (ii) shotgun sequencing data that was used to gain an understanding of the genes present and the associated functions of these genes in the rumen community, and (iii) functional metagenomics was used to identify bioactivities involved in lignocellulose degradation (Moon *et al.*, 2010a; Moon *et al.*, 2010b). From the ARC project, plant-adherent rumen bacteria were obtained through fractionation of rumen contents, which is the process of separating bacteria that are tightly adhered to the plant matter from planktonic and loosely adherent

bacteria within the rumen contents. In the ARC project, two pasture-grazing cows (722 and 723) from AgResearch Grasslands, Palmerston North were sampled for rumen contents. The extraction of the plant-adherent (AD-adherent) fraction of rumen contents from pasture grazed cows was carried out using a modified protocol described by Larue *et al.*, (2005). Briefly, the rumen content was separated into solid and liquid fractions and the adherent bacteria were released by incubation of solid digesta in detergent. The extracted DNA from adherent bacteria was amplified using 16S primers (V3 region) and analysis of the resulting sequences revealed that in both animals the most prevalent phylum was Firmicutes, class Clostridia, order Clostridiales and the most prominent family was Lachnospiraceae.

The parallel New Economy Research Fund (NERF; “Accessing the uncultured rumen microbiome”, contract number C10X0803) project aims to explore the uncultivated microbes of the rumen investigating the development of cultivation methods (Noel. S *et al.*, Unpublished) and culture-independent methods (metagenomics). Samantha Noel’s PhD cultivation experiments used dilution theory to produce over a thousand cultured isolates from the adherent fraction of rumen contents collected from pasture-fed dairy cows over a 15 month period. From the cultured isolates, 16S rRNA gene sequences were obtained via Sanger sequencing for 141 isolates and incorporated into the ARC 16S dataset. The co-occurrence of these 16S rRNA gene sequences with the sequences from the ARC metagenomic project were determined by combining all these sequences and constructing a combined phylogenetic tree. Twenty sequences representing cultured isolates which grouped with the largest clusters of organisms were selected as the initial candidate list for genome sequencing.

1.7 Project Aims

The overall purpose of the ARC and NERF research programmes is to improve our understanding of the polysaccharolytic capability of the rumen microbial ecosystem in pasture grazed cattle in New Zealand. A better understanding of the metabolic processes performed by the fibrolytic rumen microbiota is expected to generate new opportunities to enhance forage degradation, manipulate rumen function, and improve the productivity of the forage fed ruminants. Specifically the aims of this MSc project were to:

1. Characterise a set of bacterial isolates obtained from the fibre-adherent rumen microbial fraction of dairy cows grazing on grass pastures (NERF programme), in order to identify strains that adhere to and efficiently degrade plant polysaccharides.
2. Sequence and annotate the genome of one bacterial isolate, identified in the first aim.
3. Identify the relevant genes encoding the polysaccharide-degrading enzymes of the organism and any surface structures involved in its attachment to the substrate.
4. Compare the genome of the bacterial isolate to genomes of other fibrolytic rumen bacteria to gain insight into the genome organisation and modes of fibre degradation.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

The bacterial strains used in this work are listed in Table 2.1.1.

Table 2.1. Bacterial strains.

Species	Strain	Culture collection accession number*	Source
<i>Butyrivibrio fibrisolvens</i>	D1 ^T	ATCC 19171	Rod Mackie, University of Illinois at Urbana-Champaign, IL, USA
<i>Butyrivibrio hungatei</i>	JK615	DSM 14810	Jan Kopečný, Institute of Animal Physiology and Genetics, Prague, Czech Republic
<i>Butyrivibrio proteoclasticus</i>	B316 ^T	DSM 14932	AgResearch, Grassland Research Centre, Palmerston North, New Zealand
<i>Escherichia coli</i>	DH5a	DSM 6897	Paul Rainey, University of Auckland, School of biological Sciences, Auckland, New Zealand
<i>Fibrobacter succinogenes</i>	HM2	ATCC 43856	M.P. Bryant, University of Illinois at Urbana-Champaign, IL, USA
<i>Pseudobutyrvibrio xylanivorans</i>	Mz5 ^T	DSM 14809	Jan Kopečný, Institute of Animal Physiology and Genetics, Prague, Czech Republic
<i>Ruminococcus flavefaciens</i>	FD1	ATCC 19208	H. Flint, Rowett Inst. Scotland
<i>Streptococcus bovis</i>	2B	ATCC 33317	A. Klieve, DPI, Brisbane, Australia

* ATCC: American Type Culture Collection, USA; DSM: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany.

2.1.2 Buffers and reagents

dNTP Solution

Deoxynucleotide triphosphates (dNTPs), dATP, dCTP, dGTP, and dTTP, were supplied separately, each at a concentration of 100 mM, (Invitrogen, Carlsbad, CA, USA). The dNTPs were mixed in equimolar concentrations and diluted in dH₂O to give 20 mM stock solutions. Stock solutions were stored at -20°C until required.

EC Buffer

EC buffer is composed of 100 mM EDTA, 1 M NaCl, 35 mM N-Lauryl-sarcosine (Sigma-Aldrich, St. Louis, MO, USA) and 6 mM Trizma base (Invitrogen, Carlsbad, CA, USA) dissolved in dH₂O. The buffer is adjusted to pH 7.6 with NaOH, then autoclaved at 121°C for 20 min.

EDTA-Sarkosyl Solution

EDTA-Sarkosyl solution is composed of 0.5 M ethylenediaminetetraacetic acid (EDTA) and 35 mM N-Lauryl-sarcosine dissolved in dH₂O. The pH of the reagent is adjusted to pH 8.0 with NaOH, then autoclaved at 121°C for 20 min.

Glycerol Solution (40%)

Glycerol solution is composed of 4.4 M glycerol (VWR International Ltd., Lutterworth, Leicestershire, U.K.) in water. The solution was mixed, boiled for 5 min, cooled to room temperature under O₂-free CO₂, autoclaved at 121°C for 20 min and left to equilibrate for approximately 72 hr in an anaerobic chamber (Kenters *et al.*, 2010).

Mineral Solution

Mineral Solution is composed of 45 mM (NH₄)₂SO₄, 10 mM (CaCl₂·2H₂O), 10 mM (MgSO₄·7H₂O), 200 mM NaCl and 45 mM KH₂PO₄ (all supplied by VWR International Ltd., Lutterworth, Leicestershire, U.K.). All components were dissolved in dH₂O and the solution autoclaved at 121°C for 20 min (Kenters *et al.*, 2010).

Modified Karnovsky's Fixative

Modified Karnovsky's Fixative was made by heating paraformaldehyde (2 g per 100 ml) in dH₂O to 60-70°C, and adding 1 M NaOH dropwise until the solution became clear. The solution was cooled in an ice bath and Buffer Salts (2.51 g Na₂HPO₄·12H₂O and 0.41 g KH₂PO₄) and Glutaraldehyde Solution (12 ml of 25% glutaraldehyde in dH₂O) were added. The fixative was stored at -20°C (Neinhuis and Edelmann, 1996).

Phosphate-Buffered Saline (PBS, 10×) Solution

10× PBS solution is composed of 0.01 g CaCl₂ and 0.01 g MgCl₂ dissolved in 100 ml dH₂O. The resulting solution was added to a solution containing 8.00 g NaCl, 0.20 g KCl, 1.44 g Na₂HPO₄ and 0.24 g KH₂PO₄ dissolved in 1000 ml dH₂O. The combined solution was adjusted to pH 7.4, mixed well and autoclaved at 121°C for 20 min. The solution was used as a 10× stock (Neinhuis and Edelmann, 1996).

Paraformaldehyde (PFA) Solution (4%)

PFA Solution is composed of 12 g PFA that was dissolved into 195 ml dH₂O and heated to 60°C on a hotplate in a fume cupboard and 2 M NaOH was added until the solution became transparent. The solution was removed from the hotplate and 99 ml of 3x PBS solution was added and allowed to cool to room temperature. The pH was adjusted to

7.2 and the solution was filter sterilised. The PFA solution must be used within 24 hr or stored at -20°C in small aliquots (Neinhuis and Edelmann, 1996).

Saline-EDTA Solution

Saline-EDTA Solution is composed of 10 mM EDTA (VWR International Ltd., Lutterworth, Leicestershire, UK) and 150 mM NaCl dissolved in dH₂O. The pH of the solution was adjusted to 8.0, then autoclaved at 121°C for 20 min.

Selenite-Tungstate Solution

Selenite-Tungstate solution is composed of 12.5 mM NaOH, 0.01 mM Na₂SeO₃·5H₂O and 0.01 mM Na₂WO₄·2H₂O dissolved in dH₂O. The mixture was autoclaved at 121°C for 20 min (Tschech and Pfennig, 1984).

Sodium Dodecyl Sulphate (SDS) Solution (20% w/v)

SDS solution is composed of 200 g of SDS dissolved in 1.0 L of dH₂O. The pH of the solution was adjusted to 7.2, then filter sterilised into a sterile Schott bottle.

Substrate Solutions

For carbon source utilisation experiments, 20% (w/v) stocks were made for 31 sugar substrates. Soluble sugars were made up in 20 ml of dH₂O in N₂ filled Schott bottles by adding 4 g of each sugar substrate, dissolving, then autoclaving at 121°C for 20 min. The sugar substrates used were as follows: amygdalin, arabinose, cellobiose, cellulose, dextrin, esculin, fructose, galactose, glucose, glycerol, glycogen, myo-inositol, inulin, lactose, maltose, mannitol, mannose, melezitose, melibiose, pectin, raffinose, rhamnose, ribose, rutin, salicin, sorbitol, starch, sucrose, trehalose, xylan and xylose. The solutions were stored in the dark at room temperature until use (Leedle and Hespell, 1980).

Tris-Acetate-EDTA (TAE) Buffer (50× Stock solution)

TAE buffer (50×) is composed of 950 mM acetic acid (VWR International Ltd., Lutterworth, Leicestershire, UK), 50 mM EDTA and 2 M Trizma base (Invitrogen, Carlsbad, CA, USA) dissolved in dH₂O. The pH was adjusted to 8.0 before autoclaving at 121°C for 20 min. A working solution (1×) was made by diluting the stock solution 1:50 in dH₂O.

Tris-Borate-EDTA (TBE) Buffer (5× stock solution)

TBE buffer (5×) is composed of 445 mM B(OH)₃ (VWR International Ltd., Lutterworth, Leicestershire, U.K), 10 mM EDTA and 445 mM Trizma base dissolved in dH₂O and autoclaved at 121°C for 20 min. A working solution (0.5×) was made by diluting the stock solution 1:10 in dH₂O.

Tris-EDTA (TE) Buffers

Three types of TE buffer were used: TE buffer 10/0.1 (10 mM Trizma base and 0.1 mM EDTA), TE buffer 10/1 (10 mM Trizma base and 1 mM EDTA) and TE buffer 10/100 (10 mM Trizma base and 0.1 M EDTA). The buffer components were dissolved in dH₂O and the pH was adjusted to 8.0 before autoclaving at 121°C for 20 min.

Tris-EDTA-Sucrose (TES) Buffer

TES buffer is composed of 1 mM EDTA, 250 mM sucrose and 10 mM Trizma base dissolved in dH₂O. The pH was adjusted to 7.5 before autoclaving at 121°C for 20 min.

Volatile Fatty Acid (VFA) Solution

VFA Solution is composed of 114 mM butyric acid, 28 mM *iso*-butyric acid, 24 mM *iso*-valeric acid, 24 mM *n*-valeric acid (all supplied by Sigma-Aldrich), 24 mM *D,L*-2-methyl butyric acid, 202 mM propionic acid (both supplied by MERCK, Darmstadt, Germany) and 710 mM acetic acid in dH₂O. The pH was adjusted to 7.5 with 1 M NaOH and the mixture was stored at -20°C (Attwood *et al.*, 1998).

Wash Solution for Pulse-Field Gel-Electrophoresis

Wash Solution for PFGE was composed of 2 M NaCl and 20 mM Trizma base dissolved in dH₂O. The pH was adjusted to 7.6 with 6 M HCl then the solution was autoclaved at 121°C for 20 min.

2.1.3 Media and media components

Rumen Fluid

Rumen fluid was collected from rumen-cannulated cattle that had been fasted for 14-18 hr. The rumen contents were squeezed through 2 layers of cheesecloth and the liquid was collected and centrifuged twice at 20,000×g, with the pelleted microbial and plant material being discarded each time. The resulting liquid was stored frozen at -20°C until required (Kenters *et al.*, 2010).

Clarified Rumen Fluid

Frozen rumen fluid was thawed and centrifuged at 20,000×g at 4°C for 15 min. The decanted liquid was bubbled with N₂ for 10 min, then autoclaved under N₂ at 121°C for 15 min (to inactivate viruses, etc) in serum vials closed with a butyl rubber stopper and crimp-sealed. The autoclaved liquid was allowed to cool and with stirring, 1.63 g of MgCl₂·6H₂O and 1.18 g of CaCl₂·2H₂O per 100 ml of rumen fluid was added. A heavy precipitate formed which was removed by centrifugation at 30,000×g at 4°C for 60 min. The resulting supernatant was collected and represents the Clarified Rumen Fluid (Kenters *et al.*, 2010).

Cellobiose–Xylose–Arabinose–Lactate–Casamino acids–Peptone–Yeast extract–Rumen Fluid–Vitamins (GenRFV)

GenRFV is composed of the components listed in Table 2.2 added to 100 ml of Clarified Rumen Fluid.

Table 2.2. GenRFV components.

Chemical	Volume
D-glucose	0.36 g
D-cellobiose	0.34 g
D-xylose	0.30 g
L-arabinose	0.30 g
Na L-lactate syrup	0.88 ml
Casamino acids	2 g
Bacto-Peptone	2 g
Yeast extract	2 g

The mixture was stirred and N₂ gas was bubbled through the mixture for 15 min. The mixture was transferred to a N₂-flushed, sterile, serum vial through a 0.2 µm pore size filter using a syringe and needle. For each 100 ml of GenRFV, 2ml of Vitamin Solution was added via syringe and needle. To achieve the final desired substrate and growth factor concentrations, 0.5 ml of GenRFV was added per 9.5 ml of medium. The final concentrations of each GenRFV component in the media were as follows: 5% (v/v) rumen fluid, 0.5 mM cellobiose, 1 mM each of glucose, xylose, and arabinose, 5 mM lactate, 4 mM each of Ca²⁺ and Mg²⁺, and 1 g each of casamino acids, Bacto-Peptone, and yeast extract per L (Kenters *et al.*, 2010).

No Substrate-Rumen Fluid-Vitamins (NoSubRFV)

NoSubRFV was prepared by adding 2 g of yeast extract to 100 ml of Clarified Rumen Fluid and bubbling with N₂ gas for 15 min. The mixture was transferred to a N₂-flushed, sterile, serum vial through a 0.2 µm pore size filter using a syringe and needle. Two ml of Vitamin Solution was added per 100 ml of mixture. To achieve final concentrations of 5% (v/v) rumen fluid, 4 mM each of Ca²⁺ and Mg²⁺, and 1 g of yeast extract per litre, 0.5 ml of NoSubRFV was added to 9.5 ml of medium (Kenters *et al.*, 2010).

Trace Element Solution

The components listed in Table 2.3 were added and the solution was mixed well. The nitrilotriacetic acid was dissolved in dH₂O and the pH was adjusted to 6.5 using 1 M KOH before adding to the remaining solution. The mixture was autoclaved at 121°C for 20 min and stored at 4°C (Widdel *et al.*, 1983).

Table 2.3. Trace Element Solution components.

Chemical	Concentration
KAl(SO ₄) ₂ ·12H ₂ O	42 µM
B(OH) ₃	162 µM
CaCl ₂ ·2H ₂ O	680 µM
CoSO ₄ ·7H ₂ O	640 µM
CuSO ₄ ·5H ₂ O	40 µM
FeSO ₄ ·7H ₂ O	360 µM
MnSO ₄ ·2H ₂ O	2.24 mM
MgSO ₄ ·7H ₂ O	12 mM
NaCl	17 mM
NiCl ₂ ·6H ₂ O	100 µM
Nitrilotriacetic acid	8 mM
Na ₂ MoO ₄ ·2H ₂ O	40 µM
Na ₂ SeO ₃ ·5H ₂ O	1 µM
ZnSO ₄ ·7H ₂ O	625 µM

Vitamin Solution

To make Vitamin Solution, dH₂O was boiled and allowed to cool to room temperature under O₂-free CO₂. All the components listed in Table 2.4 were added and O₂-free CO₂ was allowed to bubble through the solution for a further 20 min before the solution was filter sterilised into Hungate tubes in 10 ml aliquots. The tubes were wrapped in aluminum foil to protect against light and stored either at -20°C, or kept at 4°C as working stocks (Kenters *et al.*, 2010).

Table 2.4. Vitamin Solution components.

Chemical	Concentration
Biotin	8 μ M
D-Ca-pantothenate	23 μ M
Folic acid	4.5 μ M
Lipoic acid	24 μ M
Nicotinic acid	40 μ M
p-Aminobenzoic acid	36 μ M
Pyridoxine-HCl	50 μ M
Riboflavin	13 μ M
Thiamine-HCl dihydrate	15 μ M
Vitamin B ₁₂	75 nM

Basal Medium plus Yeast Extract (BY) Medium

BY Medium was prepared from the ingredients listed in Table 2.5 as follows: All components, except L-Cysteine-HCl and Vitamin Solution, were dissolved in 1 L of dH₂O, and boiled to remove dissolved O₂. The Vitamin Solution, NoSubRFV or GenRFV containing the Clarified Rumen Fluid were injected into each Hungate tube just prior to use. Resazurin was used to indicate that the solution was anoxic as it changed colour from red to colourless. The boiled colourless solution was allowed to cool on ice to approximately room temperature, while gassing with O₂-free CO₂. L-Cysteine-HCl was added and the resulting solution was distributed in 9.5 ml aliquots into CO₂-flushed 10 ml Hungate tubes. The tubes were sealed and autoclaved at 121°C for 20 min (Leedle and Hespell, 1980).

Table 2.5. BY Medium components.

Medium component	Volume/1 L
NaCl	1 g
KH ₂ PO ₄	0.5 g
(NH ₄) ₂ SO ₄	250 mg
CaCl ₂ ·2H ₂ O	130 mg
MgSO ₄ ·7H ₂ O	200 mg
K ₂ HPO ₄	1 g
Clarified Rumen Fluid	300 ml
dH ₂ O	360 ml
NaHCO ₃	5 g
Resazurin (0.1%, w/v)	10 drops
L-cysteine-HCl	500 mg
Bacto Yeast extract	1 g
Trace Element Solution	1 ml

RM02 Medium

RM02 medium was prepared as for BY Medium, but using the components listed in Table 2.6 (Kenters *et al.*, 2010).

Table 2.6. RM02 Medium components.

Medium component	Volume/ L
dH ₂ O	950 ml
(NH ₄) ₂ SO ₄	0.6 g
K ₂ HPO ₄	1.4 g
L-Cysteine-HCl	0.5 g
KCl	1.5 g
Resazurin (0.1% solution)	4 drops
NaHCO ₃	4.2 g
Selenite-Tungstate Solution	1 ml
Trace Element Solution	1 ml
Vitamin Solution	1%

2.1.4 Enzymes, Buffers and Reaction Conditions

DNA Degradase

DNA Degradase (Zymo Research Corporation, Irvine, CA, USA) for degradation of DNA to deoxynucleotide monophosphates was supplied in a 10 U/ μ L stock solution that was prepared by dissolving 3 μ l of DNA Degradase in 7.5 μ l of 10 \times DNA Degradase Reaction Buffer and stored at -20°C.

Lysozyme

Lysozyme (Boehringer Mannheim GmbH, Mannheim, Germany) was prepared as a 25 mg/ml stock solution by dissolving 125 mg lysozyme in 5 ml TE (10/1) Buffer and stored at -20°C.

Proteinase K

Proteinase K (Roche Diagnostics, Basel, Switzerland) was used from a 10 mg/ml stock solution that was made by dissolving 50 mg of Proteinase K in 5 ml TE (10/1) Buffer and stored at -20°C.

Restriction Endonucleases

All Restriction Endonucleases (REs), their reaction buffers and bovine serum albumin (BSA) were supplied by New England Biolabs, Beverley, MA, USA.

Ribonuclease A (RNaseA)

RNaseA (Sigma-Aldrich, St. Louis, MO, USA) was prepared as a 10 mg/ml stock solution as follows: 50 mg of RNaseA was combined with 5 ml of 10 mM Tris-HCl buffer. The solution was boiled for 15 min to degrade any contaminating DNase, and then allowed to cool to room temperature and stored at -20°C.

2.1.5 General Laboratory Equipment

The centrifuges used in this work are listed in Table 2.7. Centrifuge tubes were supplied in a variety of different capacities as listed in Table 2.8.

Table 2.7. Centrifuge specifications and suppliers.

Brand	Vessels spun	Max RCF (×g)	Temp Control	Supplier
MiniSpin+	0.6 ml and 1.5 ml Eppendorf tubes	14,100	No	Eppendorf (Hamburg, Germany)
Biofuge Fresco	1.5 ml Eppendorf tubes	16,000	Yes	Heraeus (Hanau, Germany)
IEC Centra	15 ml, 50 ml Falcon tubes	3,500	Yes	Thermo Fisher Scientific, Inc. (Waltham, MA, USA)
Sorvall Evolution RC	50 ml Oakridge tubes, *250 ml tubes	48,000 (SS34) 27,500 (GSA)	Yes	Thermo Fisher Scientific, Inc. (Waltham, MA, USA)

Table 2.8. Centrifuge tubes and suppliers.

Brand	Capacity	Supplier
Eppendorf tubes	0.6 ml, 1.5 ml	Eppendorf (Hamburg, Germany)
Falcon tubes	15 ml, 50 ml	Becton, Dickinson & Co. (Sparks, MD, USA)
Oakridge tubes	50 ml	Thermo Fisher Scientific, Inc. (Waltham, MA, USA)
Wide-mouth centrifuge bottles	250 ml	Nalgene (Rochester, NY, USA)

The Pulse-Field Gel-Electrophoresis (PFGE) unit used was a CHEF-DR® III system powered by a Powerpac Basic (Bio-Rad, Hercules, CA, USA). Electroporation was carried out using a Gene Pulser™ system (Bio-Rad, Hercules, CA, USA). Gas chromatography (GC) was carried out using an Aerograph 660 (Wilkins Instruments & Research, Inc., Walnut Creek, CA, USA). The gel tanks used for conventional agarose gel electrophoresis were a wide Mini-sub® cell GT and for PFGE, a CHEF electrophoresis cell was used (both supplied by Bio-Rad, Hercules, CA, USA). DNA gels (both conventional agarose and pulsed-field gels) were digitally photographed and analysed using a Gel Logic 200 Imaging System (Eastman Kodak Company, Rochester, NY, USA). An Olympus Vanox AHB3 (Olympus America Inc., Center Valley, PA,

USA) was used for light microscopy. The Polymerase Chain Reactions (PCRs) were carried out in a Px2 Thermal Cycler (ThermoHybaid, Sedanstr, Germany). The pH meter used to measure and adjust the pH of solutions and buffers was a PHM62 pH meter (Radiometer, Copenhagen, Sweden). Nitrogen and carbon dioxide gases were supplied at food grade while hydrogen was supplied either as a pure gas or an 80:20 (v/v) mixture with CO₂ (BOC, Auckland, NZ).

Gel Migration Size Standards

The Gel Migration Standards used in this thesis are listed in Table 2.9.

Table 2.9. Gel Migration Size Standards.

Standard	Range	Use*	Supplier
1 Kb+	100 bp – 12 Kb	GE	Invitrogen (Carlsbad, CA, USA)
Lambda (λ) ladder	48.5 Kb – 727.5 Kb	PFGE	New England Biolabs (Beverly, MA, USA)
Mid range marker I	15 Kb – 242.5 Kb	PFGE	New England Biolabs (Beverly, MA, USA)
Low range marker	2.03 Kb – 194 Kb	PFGE	New England Biolabs (Beverly, MA, USA)

* GE: gel electrophoresis, PFGE: Pulse-field gel-electrophoresis.

2.1.6 Software

The software used in this thesis is listed in Table 2.10.

Table 2.10. Software used.

Software	Application	Source	Reference(s)
Artemis (v 7.0)	Genome sequence viewing and analysis	http://www.sanger.ac.uk/Software/Artemis/	(Rutherford <i>et al.</i> , 2000)
ClustalW (v 2)	Global sequence alignment	ftp://ftp.ebi.ac.uk/pub/software/clustalw2	(Thompson <i>et al.</i> , 1997)
DNAPlotter	Circular and linear interactive genome visualization tool	http://www.sanger.ac.uk/resources/software/dnaplotter/	(Carver <i>et al.</i> , 2009)
EMBOSS	Identification of DNA secondary structures	http://www.interactive-biosoftware.com/embosswin.html	(Rice <i>et al.</i> , 2000)
GAMOLA	Automated genome sequence annotation	Supplied by the programmer Eric Altermann	(Altermann and Klaenhammer, 2003)
MEGA4	Sequence alignment and phylogenetic analysis tool	http://www.megasoftware.net/	(Kumar <i>et al.</i> , 2008)
MUMmer 3.0	Large-scale DNA and protein sequence alignment tool	http://mummer.sourceforge.net/	(Delcher <i>et al.</i> , 2003)
Newbler (GS De Novo Assembler)	Genome assembly	http://www.454.com/	http://www.454.com/
Staden package (v 1.6.0)	DNA sequence assembly management	http://sourceforge.net/	(Staden <i>et al.</i> , 2000)
Vector NTI advance package (v 9)	DNA sequence display and analysis	http://www.invitrogen.com/	(Lu and Moriyama, 2004)

2.2 Methods

2.2.1 Bacterial isolate selection

The largest 16S rRNA gene sequence clusters (97% sequence identity) from the AgResearch Capability Fund project containing representative rumen bacterial cultures from the NERF project, were highlighted as the initial candidate cultures for this study. Twenty bacterial isolates, covering five different cluster groups were selected for this study. Among these, 9 isolates were from the *B. fibrisolvens* and *B. hungatei* cluster (Group 383), 9 from two *Pseudobutyrvibrio xylanivorans* cluster (7 from Group 247 and 2 from Group 245), and single isolates from the *Selenomonas ruminantium* cluster (Group 212) and the *Lachnospiraceae* cluster (Group 121).

2.2.2 Growth conditions

All *Butyrvibrio* and *Pseudobutyrvibrio* species, *Selenomonas ruminantium* and *B. proteoclasticus* B316, were cultured anaerobically in RM02 medium at 39°C unless otherwise specified. Where possible, cultures were transferred only for a limited number of sub-cultures before fresh cultures were revived from frozen stocks (-85°C), to avoid *in-vitro* culture-biased evolution (Papadopoulos *et al.*, 1999). Culture purity was verified via wet mounts and Gram staining. Wet mounts were made by dropping ~25 µl of the culture onto a slide using a sterile 1 ml syringe and needle. A coverslip was laid across the culture and the resulting slide was subsequently examined using a light microscope under phase contrast illumination. Gram stains were carried out as follows: two drops of culture were applied to the surface of a glass slide and left to air dry next to a Bunsen burner. Once dry, the slide was passed through the flame several times to heat-fix the cells. The slide was stained in succession with Crystal Violet Solution (10% crystal violet (w/v) in ethanol) for 1 min, Iodine Solution (0.3% (w/v) iodine and 0.7% (w/v) potassium iodide in dH₂O) for 1 min, and in acetone until the slide was decolourised. The slide was counter-stained with Safranin Solution (2.5% safranin (w/v) in ethanol). After each stage of treatment the slides were washed gently with water. After the final staining and wash the slide was blotted dry with tissue paper and examined using a light microscope.

2.2.3 Bacterial growth curves

Fresh overnight bacterial cultures (0.5 ml inoculum) were used to inoculate Hungate tubes containing 9.5 ml of fresh, pre-warmed, RM02 medium, with cellobiose as the substrate and uninoculated RM02 medium as a blank. Optical density (OD) readings were taken using an Ultrospec 1100 Pro spectrophotometer (GE Healthcare, Buckinghamshire, England) at a wave length of 600 nm as previously described (Lambrecht, 1966; Beard *et al.*, 1995). OD₆₀₀ readings were measured at regular 2 hr intervals until stationary phase, and the initial reading taken after inoculation subtracted to calculate the change in OD.

2.2.4 16S rRNA gene sequencing

Strains were grown as described in Section 2.2.2 using four 16S universal primers (fD1 forward 5'-gagtttgatcMtggtcag-3' (Justé *et al.*, 2008), rD1 reverse 5'-aaggaggtgatccaRccg-3' (Weisburg *et al.*, 1991; Escalante *et al.*, 2001), F1 forward 5'-actctacgggaggcagca-3' (Mrázek and Kopečný, 2001) and R3 reverse 5'-ccagggtatctaactctg-3') (PP and Rees, 1996), DNA was extracted (Section 2.2.13) and PCR products prepared for 16S rRNA sequencing at the Allan Wilson Centre Genome Sequencing Service at Massey University, Palmerston North (Section 2.2.14). The ABI trace files were processed using the Staden package. The ClustalW (Thompson *et al.*, 1997) program was used to align the 16S rRNA sequences and phylogenetic analysis was performed using Molecular Evolutionary Genetics Analysis (MEGA4) software version 4.0 (Tamura *et al.*, 2007).

2.2.5 Carbon source utilisation

To test bacterial carbon source utilisation, fresh overnight cultures were used to inoculate Hungate tubes containing 31 different sugar substrates including seven insoluble polymer sugars (glycogen, pectin, inulin, cellulose, dextrin, starch and xylan). OD₆₀₀ readings were measured initially after inoculation, and after 24 hr and 72 hr of incubation for soluble substrates (Moon *et al.*, 2008), while growth on insoluble substrates was assessed by microscopic examination and measurement of increased protein after growth.

2.2.6 Culture Storage

Cultures were inoculated and grown overnight in 10 ml Hungate tubes containing 5 ml RM02 medium, 0.5 ml of culture inoculum, 0.5 ml of GenRFV. The cultures were supplemented with 3 ml of Glycerol Solution (40% v/v) then stored at -85°C.

2.2.7 Percentage Guanine plus Cysteine (%G+C) Content

%G+C content was determined for six bacterial strains as described by Ramsahoye. Genomic DNA of each bacterial strain (3 µg) was digested to deoxynucleotide monophosphates using DNA Degradase, and quantified using a NanoDrop® ND-1000 (NanoDrop Technologies, Inc., Wilmington, DE, USA). Samples were analysed by reversed-phase high-performance liquid chromatography (HPLC) as follows: Luna 5 µl C8 100A, 150 × 4.6 mm (OOF-4040-EO) column temp 6°C, flow rate 1 min, Solvent A = 50 mM (NH₄)₂HPO₄, pH 4.1, Solvent B = 100% acetonitrile, isocratic solutions (97% Solvent A and 3% Solvent B), injection volume 50 µl, run time was 21 min between injections (Ramsahoye, 2002).

2.2.8 Volatile Fatty Acid (VFA) Analysis

To determine the VFA profile of bacterial strains, cultures were grown in RM02 medium supplemented with 0.5 ml GenRFV solution. After growth, cultures were centrifuged at 10,000×g for 15 min at 4°C and the supernant was filtered through 0.45 µm polycarbonate filter (Nucleopore Corp., Pleasanton, CA, USA) into autosampler vials. The filtered samples were used immediately for chromatographic analysis (Ehrlich *et al.*, 1981) with the parameters described below: BioRad 2XHPX-87H column, column temperature was 45°C, solvent was 5 mM H₂SO₄ (isocratic), flow rate was 0.8 ml/min (isocratic), and injection volume was 50 µl loop. The standards used for column calibration were 5 mM glucose, 10 mM succinic acid, 10 mM lactic acid, 20 mM formic acid, 20 mM acetic acid, 15 mM propionic acid, 20 mM ethanol, and 10 mM *n*-butyric acid (Attwood *et al.*, 1998).

2.2.9 Adhesion Assays

Visualising Bacterial Adherence with DAPI staining

The Adhesion Assay was used to visualise bacterial adherence to the insoluble carbohydrates sisal string (Donaghys, Christchurch, NZ) and ryegrass Neutral Detergent

Fibre (NDF) using Light Microscopy. Four bacterial strains were used as controls: *Fibrobacter succinogenes* HM2, *Ruminococcus flavefaciens* FD1, *Streptococcus bovis* 2B and *B. proteoclasticus* B316. All bacterial isolates (AB2010, MB2003, MA3014, AB3002, MA3010 and MB2006) and controls were grown in both BY and RM02 media using the GenRFV carbohydrate mix, containing either five strands of finely separated sisal or three strands of 10 mm NDF ryegrass fibre. The cultures were incubated overnight at 37°C. After 24 and 48 hr incubation, a few fibres were taken from each of the cultures and placed in 1.5 ml Eppendorf tubes, and fixed by adding 300 µl PBS and 900 µl of 4% PFA Solution (1:3 ratio of sample: PFA Solution). The samples were fixed for approximately 3-4 hr at 4°C then rinsed with 1 ml PBS.

Staining

One ml of PBS containing 1 mg of 4', 6-diamidino-2-phenylindole (DAPI) was added to each tube and left to fix for approximately 10 min. Then samples were rinsed with 1 ml of PBS, three times in succession, inverting the fibres and mixing well. A single fibre was removed from each tube, placed on a glass microscope slide surrounded with PBS and a cover slip placed on top. The samples were observed under an Olympus Vanox AHB3 Light Microscope (Olympus America Inc., Center Valley, PA, USA) using the default DAPI settings to visualise the fluorescent images (Hamada and Fujita, 1983).

2.2.10 Scanning (SEM) and Transmission Electron Microscopy (TEM)

Preparation and fixation of NDF samples for SEM imaging

The NDF fraction was produced and prepared for experimental use as outlined by Maynard *et al* (Maynard *et al.*, 1979). Triplicate bacterial samples were grown in 10 ml BY medium (supplemented with 0.2 ml cellobiose 10% w/v) containing a single 10 mm ryegrass NDF strand (Van Soest *et al.*, 1991). Cultures were incubated at 37°C for 24 hr, the broths (along with the NDF strand) emptied into a sterile Petri dish. The NDF strands were separated and washed five times in 50 ml volumes of PBS with sterile forceps. Excess PBS was removed by flicking onto an adsorbent paper towel. The NDF samples from each bacterial strain were then fixed in 3 ml of Karnovsky's Fixative solution for 48 hr.

Biological specimen processing for SEM

Samples were fixed in TEM primary fixative (3% glutaraldehyde, 2% formaldehyde in 0.1 M Phosphate Buffer pH 7.2) for two days at room temperature. Three washes using PBS Buffer were carried out for 10-15 min each. Samples were dehydrated through a series of ethanol solutions: 25%, 50%, 75%, and 95% for 10-15 min each and 2× in 100% ethanol for approximately 1 hr. The samples were then Critical Point (CP) dried using liquid CO₂ as the CP fluid. The dried samples were mounted onto an aluminum specimen support stub using double-sided adhesive tape or conductive silver paint, sputter coated with gold, and observed via SEM (Neinhuis and Edelmann, 1996).

Negative Stain Method for Biological Specimen Processing for TEM

Triplicate bacterial samples were grown in 10 ml RM02 and BY medium (supplemented with 0.2 ml cellobiose 10% w/v) and incubated at 37°C for 24 hr. A drop of bacterial sample was placed onto a piece of parafilm followed by placement of a Formvar grid (shiny side down) on top of the drop of the sample. After 4 min, the liquid was drained from the grid with the edge of a piece of filter paper. The grid was placed on a drop of 2% phosphotungstic acid stain (or 2% (w/v) uranyl acetate in water for 4 min and drained with filter paper. The grid was allowed to dry, and observed under TEM (Bozzola and Russell, 1999).

2.2.11 Motility Assay

To examine *Butyrivibrio* sp. MB2003 motility, a Motility Agar Stab Test was carried out using *B. proteoclasticus* B316 as a negative control. The motility test used 5 ml of 1% agar poured in a 10 ml Hungate tube and allowed to set. Freshly grown cells were stab-inoculated into the agar using a straightened inoculating pick, about two thirds of the way into the agar, ensuring to make the stab as straight as possible, as it was the amount of growth away from the stab that was to be examined. The cultures were incubated at 37°C for 24 to 48 hr and observed under a light source. Non-motile bacteria only grow where they were inoculated, whereas motile bacteria produce a diffuse or cloudy growth pattern a distance from the stab area (Tittsler and Sandholzer, 1936).

2.2.12 Pulsed-field gel electrophoresis (PFGE)

DNA extraction for PFGE

To isolate genomic DNA for PFGE cultures were grown to approximately mid-exponential phase, and then heated to 70°C for 10 min to destroy heat-sensitive nucleases. Cells were transferred to 1.5 ml Eppendorf tubes and harvested by centrifugation at 5,000×g for 5 min. The supernatant was discarded and the harvested cells were subsequently resuspended in 1 ml Wash solution and harvested again by centrifugation at 5,000×g for 5 min. This step was repeated once more to give a total of two washes and cells were then resuspended in 150 µl of the Wash Solution. The cell suspension was mixed 1:1 (v/v) with molten 2% Pulsed-Field Certified Low Melt Agarose (Bio-Rad, Hercules, CA, USA) in 125 mM EDTA equilibrated at 50°C. The resulting cell suspension was dispensed into 10-well reusable plug moulds (Bio-Rad, Hercules, CA, USA) and allowed to solidify at room temperature. Once solidified, plugs were transferred to Universal Bottles (up to 4 plugs per Universal) containing 4 ml of EC buffer containing 1 mg/ml lysozyme. Plugs were incubated on an Orbitech XL flask shaker with shaking (Infors HT) for 18-24 hr at 37°C to allow cell lysis to occur. The EC buffer/lysozyme solution was replaced with 4 ml of 0.5 M EDTA-Sarkosyl solution containing 0.5 mg/ml of Proteinase K. Plugs were incubated on the flask shaker for 16-24 hr at 37°C. Fresh EDTA-Sarkosyl/Proteinase K solution was added and the plugs were again incubated on the flask shaker for 16-24 hr at 37°C. The EDTA-Sarkosyl/Proteinase K solution was replaced with TE (10/1) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) to eliminate residual Proteinase K activity, and incubated on the flask shaker for 2 hr at 37°C. Plugs were washed once in TE (10/100) and stored at 4°C in the same solution (Yeoman, 2009).

Restriction Endonuclease digestion of PFGE gel plugs

The appropriate gel plugs were cut to ~2.5 mm using a sterile scalpel, carefully transferred to a sterile Eppendorf tube containing 1 ml of TE 10/0.1 and allowed to equilibrate for 1 hr. The TE 10/0.1 buffer was removed, the plug immersed in an excess of the appropriate Restriction Endonuclease (RE) Buffer (and BSA where necessary), and the plug was allowed to equilibrate in this solution for 1 hr. The RE Buffer was removed and a fresh 100 µl of the same buffer was added along with 0.2 µl of the appropriate 1 U RE. The plug was incubated in this solution at the optimal temperature

of the RE for 16-24 hr. The RE buffer was removed and 500 μ l of TE 10/100 buffer was added. The plug was kept in this buffer at 4°C until PFGE was performed (Yeoman, 2009).

PFGE

PGFE was carried out as described by Gardiner (Gardiner, 1991) using a Bio-Rad CHEF-DR® III system (Bio-Rad, Hercules, CA, USA). Pulsed-field Certified Agarose (Bio-Rad, Hercules, CA, USA) was mixed with 0.5x TBE buffer to give a final concentration of 1%. The mixture was boiled to dissolve the agarose and then allowed to cool to 50°C in a waterbath. Gels were cast in either a 14× 13 cm casting stand with a 1.5 mm thick 10 well comb, or a 20.25× 14 cm casting stand with a 1.5 mm thick 15 well comb. The gel was allowed to cool to room temperature and solidify. The appropriate agarose plugs were carefully transferred to the wells of the solidified gel. An appropriate DNA size ladder was added to the gel, typically the bordering wells. The plugs were then embedded in the gel with molten 1% Pulsed-field Certified Agarose in 0.5× TBE. The gel was transferred to the Chef Electrophoresis cell containing 2 L of 0.5× TBE. The TBE within the electrophoresis cell was maintained at 14°C by circulation through a Bio-Rad Model 1000 mini chiller using a variable speed pump set at 95%. The DNA fragments were separated by applying 5.5 volts/cm at 120° angles. Pulse switch times and electrophoretic duration were dependent upon the expected size range of the resulting RE fragments. The gel was stained with 0.5 μ g/ml ethidium bromide for 40 min, then washed in water for 10 min and analysed by UV transillumination at λ =590 nm and photographed using a Gel Logic 200 system (Eastman Kodak Company, Rochester, NY, USA).

2.2.13 *Butyrivibrio* sp. MB2003 Genome Project

Strain Information and Growth Conditions

The strain chosen for whole genome sequencing was *Butyrivibrio* sp. MB2003 which was isolated from the rumen contents of grass-fed cattle (Noel. S *et al.*, Unpublished). The cultures were grown under anaerobic conditions at 39°C in RM02 medium broth supplemented with GenRFV.

DNA extraction

Genomic DNA was extracted using the method described by Saito and Miura (Saito and Miura, 1963). Briefly, 10 ml cultures were harvested at mid-exponential phase ($OD_{600} \approx 0.4$) by centrifugation at $5,000 \times g$ for 5 min. The cell pellet was resuspended in 500 μ l Saline-EDTA Solution and harvested by centrifugation at $5000 \times g$ for 5 min. The cell pellet was resuspended in 500 μ l of Saline-EDTA Solution containing 1 mg/ml lysozyme and 20 μ g/ml Ribonuclease A (RNaseA). This solution was incubated for >1 hr at 37°C to allow the lysozyme to facilitate cell lysis, and the RNaseA to subsequently degrade RNA. Sodium dodecyl sulfate (SDS, 27 μ l) solution, along with 11 μ l of Proteinase K solution was added to give a final concentration of 1% (w/v) and 200 μ g/ml respectively. The solution was incubated for 1.5 hr at 60°C to allow the Proteinase K to degrade cellular protein and inactivate nucleases. TE Buffer (10/1) (200 μ l) was added and the mixture was extracted with phenol:chloroform. One volume of room temperature, buffer-saturated phenol was added to the DNA extraction and mixed by several inversions before precipitated protein and other organic material was fractionated by centrifugation at $13,000 \times g$ for 5 min. The supernatant was carefully transferred to a fresh Eppendorf tube and mixed with an equal volume of phenol:chloroform:isoamyl alcohol mixture. The mixture was inverted several times before the phases were separated by centrifugation at $13,000 \times g$ for 5 min. The supernatant was transferred to a fresh eppendorf tube and this extraction step was repeated with an equal volume of chloroform to remove residual phenol. The supernatant was transferred to a fresh Eppendorf tube and 0.1 volumes 5 M ammonium acetate and 3 volumes of absolute ethanol were added and mixed. The mixture was left 16-24 hr at -20°C then centrifuged at $16,000 \times g$ for 30 min at 4°C . The supernatant was carefully removed using a pipette to avoid disturbing the DNA pellet. The pellet was washed in 200 μ l of 70% ethanol and centrifuged again at $16,000 \times g$ for 15 min at 4°C . The supernatant was carefully removed using a pipette and allowed to dry at 37°C for 30 min. The DNA pellet was then dissolved in an appropriate volume of either sterile dH_2O , TE (10/1) or 10 mM Tris-HCl, depending on the subsequent application. DNA concentration and purity was determined by spectrophotometry using the NanoDrop $\text{\textcircled{R}}$ ND-1000 (NanoDrop Technologies, Inc., Wilmington, DE, USA). Absorbance readings were determined at $\lambda = 230$ nm, 260 nm and 280 nm. DNA concentration was determined using the Beer-Lambert equation modified to use an extinction coefficient of 50 ng-

cm/ml for DNA and 40 ng-cm/ml for RNA (Petersen *et al.*, 2008). The purity of the samples was determined by analysis of the A_{260}/A_{280} and A_{260}/A_{230} ratios.

Genome Sequencing and Assembly

The genome sequence of *Butyrivibrio* sp. MB2003 was determined using a pyrosequencing approach (Macrogen Corporation, Korea). A genomic DNA (gDNA) sample of high quality, double-stranded, non-degraded and purified by column purification protocols with an A_{260}/A_{280} ratio of approximately 1.8 was sent for sequencing at Macrogen Corporation, Korea for ½ plate of titanium paired-end (454 GS FLX-Titanium) pyrosequencing (Kircher and Kelso, 2010). Sequence data was received as a Standard Flowgram Format (SFF) file. Three separate programs were used to determine the baseline information from the SFF file. The SFFtools (www.454.com) program was utilised to generate a multifasta file of all the reads within the SFF input file. To count the number of reads within the newly generated multifasta file the GREP Unix function was used. The INFOSEQ program from the EMBOSS (Rice *et al.*, 2000) package was utilised to analyse the length and distribution of the individual sequence reads. The read length and average %G+C content data for each read were processed and visualised using Microsoft Excel. Genome assembly of the SFF file was performed using the Newbler GS De novo Assembler software (www.454.com). The assembled contig fasta file was analysed using INFOSEQ to generate information regarding contig length and %G+C content, the results were summarised in Microsoft Excel spreadsheets.

Sequence Quality Control

Any contigs within the DNA sequence found to have a %G+C of greater than 43% were removed from the assembled data. The total length and the average %G+C content of the remaining contigs were processed in INFOSEQ and summarised using Microsoft Excel.

Sequence Assembly Management

The assembled sequence data was converted into an experimental (EXP) file format using the Pregap4 program of the Staden package. The EXP files were entered into the Gap4 program of the Staden package for further processing. Gap4 is a graphical interface program that enables visualisation and management of the assembled data.

Contig Scaffolding

Within Gap4, a list of contigs with their corresponding scaffolds (33 contigs arranged into 8 scaffolds) obtained from the Newbler scaffold file, was created and tabulated using Microsoft Excel. Basic Local Alignment Sequence Tool (BLAST) analysis of 800bp of sequence at each contig end was performed to identify super-scaffolds. The PERL script “get_contig_ends.pl” was used to extract 800 bp at both ends of each contig and outputted to a multi-FASTA file format. BLASTX compares translational products of the nucleotide query sequence to a protein database (<http://www.ncbi.nlm.nih.gov/>). The UNIX command below initiates a BLASTX of the modified multifasta file on the AgResearch IMPALA server:

```
/usr/local/blast/blastall -p blastx -d /data/databases/flatfile/illuminati_blastdata/nr -i ContigsOfInterest.fasta -o ContigsOfInterest.fasta.nr -v5 -b5 &
```

The BLAST hits were analysed and the results summarised within Microsoft Excel, the top BLAST hit and corresponding accession number were recorded. BLAST results indicating contaminant contigs were identified and eliminated from the Gap4 database. The BLASTX results also produced information leading to the identification of three potential extrachromosomal elements within the *Butyrivibrio* sp. MB2003 genome. The three contigs were circularised using standard PCRs.

2.2.14 Gap Closure

Primer Design

The primer design function within the Gap4 program was used to design primers for gap closure of the *Butyrivibrio* sp. MB2003 genome project. For regions difficult to sequence, such as ribosomal rRNA operons, repeat regions and regions with possible stable secondary structures, the contigs of interest were visualised using Artemis (Rutherford *et al.*, 2000). BLASTN and BLASTX results for the coding ORFs in that region were examined. The information from the BLAST searches was used to select appropriate regions to allow design of suitable primers. Primers synthesized at 50 nM scale by IDT (Integrated DNA Technologies; <http://eu.idtdna.com>) were supplied desalted and lysophilized. Primers were reconstituted in sterile dH₂O water at a concentration of 100 µM and stored at -20°C. PCR reactions and subsequent preparations of PCR products were carried out for all gap closing reactions as described in sections below.

Polymerase Chain Reactions

Polymerase Chain Reactions (PCR) were performed in 50 µl volumes using a standard PCR protocol. Constituents were added to dH₂O in a 0.2 ml thin wall PCR tube (Quality Scientific Plastics, Petaluma, CA, USA), in the order listed in Table 2.11, on ice. The solution was mixed using a P200 micropipetter (Thermo LabSystems, Waltham, MA, USA) and a sterile 200 µl pipette tip and subsequently transferred to a PX2 Thermal Cycler (Thermo). The reaction was cycled through a 2 min denaturation at 94°C, followed by 35 cycles of:

- 94°C for 30 sec
- 56°C for 30 sec
- 72°C for 2 min

The reaction underwent a 2 min elongation at 72°C and was cooled to 4°C until examined by gel-electrophoresis.

Table 2.11 PCR assembly of reagents.

PCR constituents	Amount per 1x Reaction	Final Concentration
10x PCR Buffer, minus Mg	5 µL	1X
10 mM dNTP mixture (each)	4 µL	0.2 mM each
50 mM MgCl ₂	1.5 µL	1.5 mM
Primer Mix (20 µM each 5' & 3')	0.5 µL each	0.2 µM each
Template DNA (diluted)	1 µL	(as required)
Platinum Taq DNA Polymerase	0.2 µL	1.0 unit
dH ₂ O (autoclaved)	40.3 µL	N/A
Total Volume	50 µL	

Long Range PCR

Long Range PCR was performed using the Eppendorf Triple Master system (Eppendorf, Hamburg, Germany). The reaction was prepared as described in Table 2.12.

The reaction was cycled through a 3 min denaturation at 93°C, followed by 35 cycles of:

- 93°C for 15 sec
- 56°C for 30 sec
- 68°C for 7 min

The reaction was cooled to 4°C until examined by gel-electrophoresis.

Table 2.12 Long Range PCR assembly of reagents.

LR PCR constituents	Amount per 1x Reaction	Final Concentration
10x HiFi PCR Buffer	5 µL	1X
20 mM dNTP mixture (each)	2 µL	0.2 mM each
50 mM MgSO ₄	2 µL	2.0 mM
Primer Mix (20 µM each 5' & 3')	1.0 µL each	0.4 µM each
Template DNA (diluted)	1 µL	(as required)
HiFi Platinum Taq (5 units/µL)	0.2 µL	1.0 unit
dH ₂ O (autoclaved)	37.8 µL	N/A
Total Volume	50 µL	

Agarose gel-electrophoresis

Agarose gel-electrophoresis was carried out using Bio-Rad equipment (Bio-Rad, Hercules, CA, USA). Agarose (Invitrogen, Carlsbad, CA, USA) or Low Melt Agarose (Progen, Heidelberg, Germany) was mixed with 1× TAE buffer to give a final concentration of 1.5% w/v agarose or 1% w/v Low Melt Agarose, unless otherwise stated. The mixture was boiled to melt the agarose and then allowed to cool to 50°C in a waterbath. Gels were cast in a 15× 10 cm casting stand with either a 1.5 mm thick 15 well comb, or a 1.5 mm thick 20 well comb. The gel was allowed to cool to room temperature and solidify. The comb was removed and the gel tray was transferred to a wide Mini-sub® cell GT (Bio-Rad) and immersed in 1× TAE buffer. Unless otherwise stated, 9 µl of each PCR-amplified product was dispensed on to a small square of Parafilm M (American National Can, Chicago, IL, USA), combined with 1 µl of 10× BlueJuice™ (Invitrogen, Carlsbad, CA, USA) and carefully transferred to a separate well. Also, 8 µl of the 1 Kb+ ladder (Invitrogen, Carlsbad, CA, USA) was added typically to bordering wells. The DNA fragments were separated using a Powerpac Basic applying 80 volts ($5 \frac{1}{3}$ volts/cm) for 60 min. The gel was stained with 0.5 µg/ml ethidium bromide for 40 min, then washed in water for 10 min and analysed by UV transillumination at $\lambda=590$ nm and photographed using a Gel Logic 200 system (Eastman Kodak Company, Rochester, NY, USA).

PCR clean up

DNA fragments resulting from PCR reactions were purified using a QIAquick® PCR purification kit (Qiagen, Hilden, Germany). Binding (PB) Buffer from the kit (5 volumes) were mixed with the product of each PCR reaction and the solution was applied to a QIAquick spin column and subjected to centrifugation for 60 sec at 13,000×g. During this process, DNA of 0.1 – 10 Kb binds to the column, while all other constituents of the reaction flow through. The flow through was discarded and 750 µl of Elution (PE) Buffer from the kit was applied to the column and centrifuged for 60 sec at 13,000×g. This washing step was repeated once more, discarding the flow-through each time. The column was centrifuged for an additional 1 min at 13,000×g to remove any residual PE buffer, and 30 µl of sterile dH₂O was added to the centre of the QIAquick column and incubated at room temperature for 5 min. The eluate was transferred to a fresh sterile Eppendorf tube by centrifugation at 13,000×g for 60 sec.

PCR Product Sequencing

All DNA sequencing reactions of PCR products were conducted by the Allan Wilson Centre Genome Sequencing Service at Massey University, Palmerston North. This service included fluorescent labeling of PCR products using the Big Dye™ Terminator (Version 3.1), a Ready Reaction Cycle Sequencing Kit, subsequent removal of unincorporated fluorescent dideoxy NTPs (ddNTPs) by cleanup and precipitation of products and capillary separation on an ABI3730 Genetic Analyzer (Applied Biosystems Inc., Carlsbad, California). Results were received as ABI tracefiles that were analysed using Trev (Ewing *et al.*, 1998; Bonfield *et al.*, 2002), a DNA trace visualization and editing tool from the Staden package.

Gap Closure

The *Butyrivibrio* sp. MB2003 ABI trace files received from sequencing of the PCR products were entered into the Gap4 database. The ABI files were pre-processed as summarised in Figure 2.1 and the process were repeated until gaps were closed.

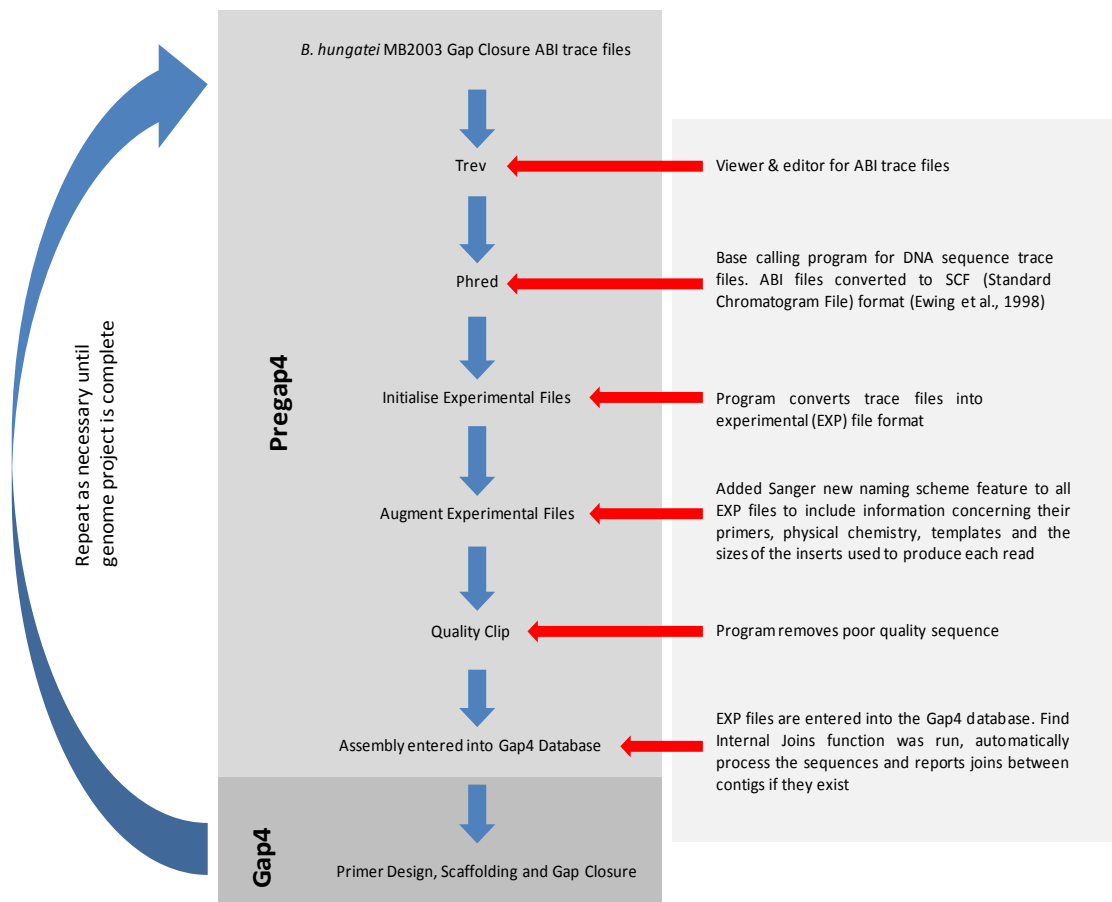


Figure 2.1. Outline of the pre-processing and assembly of the *Butyrivibrio* sp. MB2003 gap closure sequences.

2.2.15 Annotation of the *Butyrivibrio* sp. MB2003 genome

Gene Prediction

The GAMOLA (Global Annotation of Multiplexed On-site bLasted DNA-sequences) (Altermann and Klaenhammer, 2003) software was used to predict protein-coding Open Reading Frames (ORFs) and provide an automated annotation of the generated gene model. To determine ORFs present within each input DNA sequence GAMOLA utilises the gene finding programs: GLIMMER3.0 (Gene Locator and Interpolated Markov ModelER) (Delcher *et al.*, 1999), CRITICA (Coding Region Identification Tool Invoking Comparative Analysis) (Badger and Olsen, 1999), and PRODIGAL (Prokaryotic Dynamic Programming Genefinding Algorithm) (Hyatt *et al.*, 2010) to create an additive gene model. Preliminary annotation and functional characterisation of the ORFs in the MB2003 genome were made based on analysis of BLASTP and BLASTX results against the NCBI (National Center for Biotechnology Information) NR (non-redundant) database (Sayers *et al.*, 2011). The search results were presented in a report that places the results with the lowest “e-value” at the top along with their general description, along with the local alignment. The BLAST hit with lowest e-value for a particular sequence was determined as the initial “auto” annotation for that ORF. The HMMER3 program (Release 24) was utilised for Protein Family (PFam) analysis (Eddy, 1998). The InterPro database (2007/2008 release) was also incorporated into the analysis of PFams as an additional source of information regarding PFams, and was also utilised in the analysis of TIGRfams. TIGRfams (Version 9) were used to provide useful information for identifying functionally related proteins based on sequence similarity (Haft *et al.*, 2003). COGs (Clusters of Orthologous Groups of proteins) were used to represent proteins or clusters of genes that correspond to a conserved domain (Tatusov *et al.*, 2000). The Rfam database provides information on the non-coding RNA (ncRNA) elements (Gardner *et al.*, 2009). The INFERNAL (INFERENCE of RNA Alignment) package was utilised for annotation of Rfams (Version 9.1) (Eddy, 2002). The SignalP database was used to predict presence of signal peptides (including their respective cleavage sites) based on artificial neural networks and HMMs (Dyrløv Bendtsen *et al.*, 2004). The TMHMM database was used to predict and describe transmembrane helices for individual protein domains based on HMMs (Krogh *et al.*, 2001). Upon completion of the annotation process, the Artemis program (Rutherford *et*

al., 2000) was used to visualise, analyse and interpret the results produced by GAMOLA.

Manual Curation

After completion of GAMOLA analysis, manual annotation of the predicted ORFs was carried out to verify annotations or make alterations to the start and stop of each predicted Coding Sequence (CDS) if necessary. Each ORF was analysed individually via Artemis and annotated based on the best possible BLASTP hit for that particular sequence, the e-value of the corresponding protein hit, and the e-value of the corresponding Pfam result. Significant similarity to a characterised homolog was an alignment with at least 30% amino acid sequence identity and 50% sequence similarity across at least 70% of both ORFs. Another factor in determining correct annotation of a CDS was the associated TIGR domain classification. TIGRFams differ from Pfams in that they provide more accurate functional assignments to the genes they describe, whereas Pfams use names which represent only a small subset of the functions of the genes which they describe. Ribosomal RNA (rRNA) genes were identified based on their BLASTN analysis and were annotated manually. ORFs smaller than 50 amino acids that lacked a significant match to any previously described gene and gave no trusted hits to known HMMs were eliminated. Putative functions were assigned based upon the bioinformatic analyses and where possible identified by the appropriate Gene Ontologies and/or Enzyme Commission (EC) number(s). ORFs displaying no significant sequence similarity to any previously described gene were annotated as “hypothetical” proteins, whereas those displaying significant sequence similarity to a gene of unknown function were annotated as “conserved hypothetical” proteins. Any alterations to the beginning and/or end of the amino acid sequence for each ORF were made based on sequence data and protein comparisons. Manual annotation was first performed on the three extrachromosomal elements, the chromosome was annotated last. Gap closure continued during the annotation process and currently 7 rounds of gap closure have been completed for this work.

2.2.16 Analysis of sequence data

MUMmer Plot

The software MUMmer 3 (Delcher *et al.*, 2003) was used to generate a synteny plot of the *Butyrivibrio* sp. MB2003 chromosome versus the *B. proteoclasticus* B316 chromosome to allow genome wide comparisons to be made. The Gsview (Lang, 2002) program was used to visualise the generated MUMmer plot.

Circular Plot

DNAPlotter (Carver *et al.*, 2009) was used to display the sequence information of the four replicons of *Butyrivibrio* sp. MB2003 as circular plots. The program also illustrates the %G+C content and %GC skew information as part of the circular plot. The resulting images were saved as PostScript format files and visualised using the Gsview program.

Genes involved in Polysaccharide Breakdown

Comparative analysis of the total number of glycoside hydrolases, polysaccharide lyases, carbohydrate esterases and carbohydrate-binding proteins was performed for *Butyrivibrio* sp. MB2003 and *B. proteoclasticus* B316. The genes encoding polysaccharide degrading enzymes involved in the breakdown of plant cell walls were derived from GAMOLA output and the CAZy (Carbohydrate-Active enZYmes) database (Henrissat. *et al.*, 2011), and were tabulated using Microsoft Excel.

Chapter 3

Results

3.1 Bacterial isolate selection

Prior to the start of this project, work within a NERF program had isolated a range of rumen bacteria from the plant-adherent fraction of rumen contents from pasture-grazed dairy cows (Noel. S *et al.* Unpublished). The relative abundance of these organisms was assessed by comparing their partial 16S rRNA gene sequences (500bp, V1-V3 region) with sequences obtained from high-throughput pyrosequencing of 16S rRNA genes amplified from DNA extracted from rumen contents of pasture-fed dairy cows in a parallel ARC project. The sequences of the freshly isolated bacteria were analysed and clustered into groups within a phylogenetic tree. The phylogenetic tree was composed of 347 OTUs with 3,870 different representative 16S rRNA gene sequences from the V1-V3 region, of which 151 culturable isolates were available for study. The co-clustering of sequences from cultured organisms with large groups of sequences derived directly from rumen contents DNA was used as an indicator of their abundance in the rumen and by inference, their importance in the plant fibre degradation process.

Altogether, 20 isolates covering five large clusters were selected for study. The *B. fibrisolvens/hungatei* family (cluster group 383) contained 9 isolates. The *P. xylanivorans* cluster consisted of two separate groups (247 and 245); the 247 cluster group contained 7 isolates while the 245 cluster group contained 2 isolates. A single isolate associated with the *Selenomonas ruminantium* cluster (cluster group 212) and one isolate clustered with the *Lachnospiraceae* family AB2031 (cluster group 121) were chosen for study, giving a total of 20 candidate isolates for study (Figure 3.1 and Table 3.1).

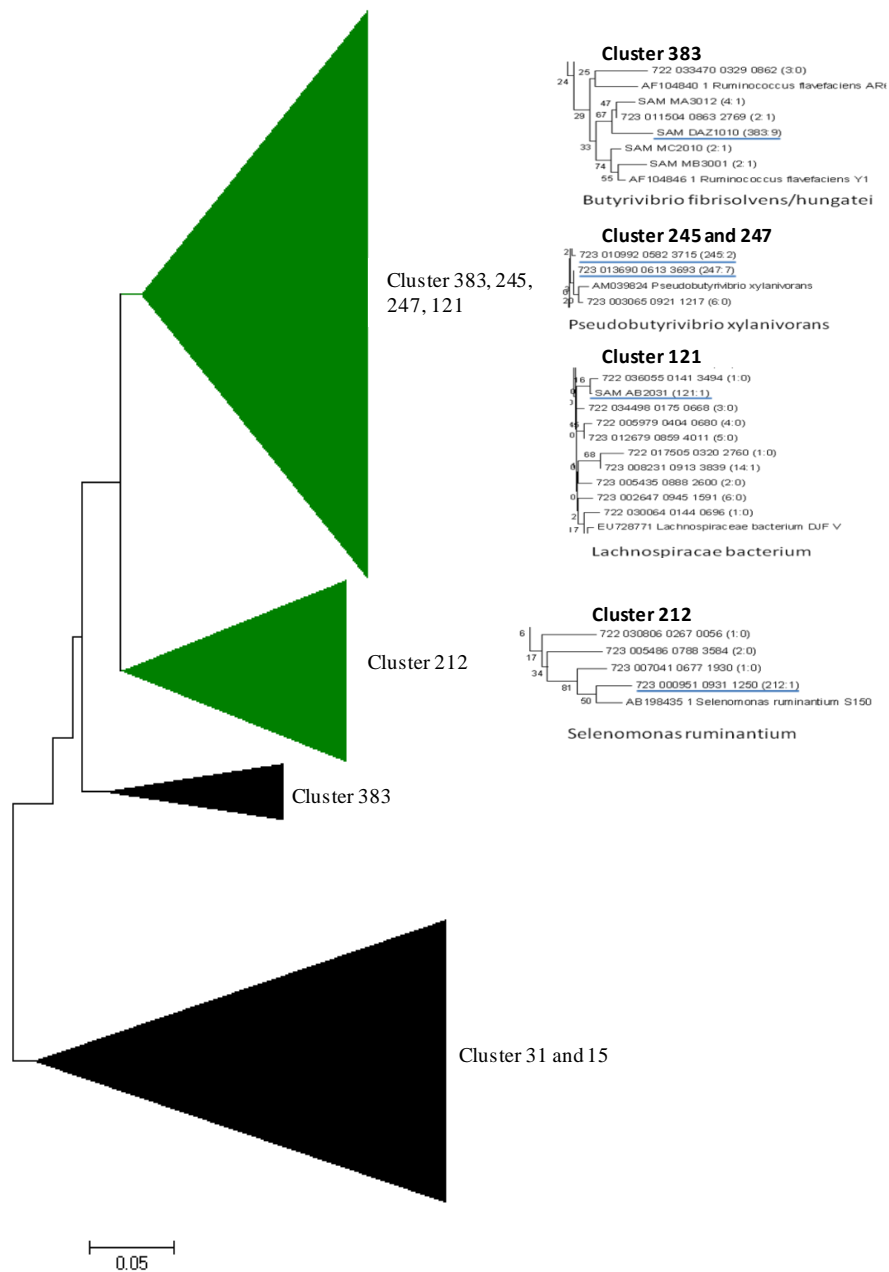


Figure 3.1. Phylogenetic tree based on the 16S rRNA gene sequences from rumen bacterial isolates cultivated in the NERF study, plant-adherent rumen microbial communities isolated in the ARC study, and a selection of reference rumen bacteria. The raw 16S rRNA pyrosequence dataset from the ARC study (V3 region) was pre-processed using the Ribosomal Database Project (RDP) Release 9 pyrosequencing pipeline (Cole *et al.*, 2009), then combined with 16S rRNA sequences (V1-V3 region) from bacterial isolates cultivated in the NERF study (Noel. S *et al.* Unpublished). In RDP, the sequences were aligned using the Infernal aligner (Nawrocki and Eddy, 2007) and 97% sequence identity clusters were generated using the Complete Linkage Clustering tool, resulting in 306 clusters. A representative sequence from each cluster was selected using the Dereplicate and FASTA sequence selection tools, and aligned with 16S rRNA gene sequences from 54 reference rumen bacterial isolates using the Infernal aligner. The resulting alignment consisted of a total of 360 sequences and 1807 positions. The alignment was imported into MEGA4 (Tamura *et al.*, 2007) for phylogenetic analysis. The phylogeny shown was determined using the Neighbour Joining method (Saitou and Nei, 1987), with evolutionary distances computed using the Kimura 2-parameter method (Kimura, 1980). The sum of branch lengths = 15.6573; the tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. For simplicity, the four major clades that comprise the tree are represented by triangles, and the locations of the most abundant clusters containing cultivated representatives are indicated to the right of these four clades. The triangles in green represent clades containing the most abundant clusters with cultivated representatives selected for study, the triangles in black represent clades containing other remaining clusters. The subtrees to the right underline the specific cluster groups from which candidate isolates were selected. The first number in brackets corresponds to the number of ARC sequences in the cluster, and the second is the number of cultivated NERF sequences in the cluster. The reference isolate description for each specific cluster group is given below each subtree. *Isolate codes: The first two letters of the isolate designation correspond to month (May or August), and to cow (cows A, B, C, D, E were located at Lye farm, Dairy NZ, Hamilton). The first number designates the dilution factor of the inoculum for the culture and the remaining 3 numbers indicate the culture number. The DAZ notation refers to a cow located at the AgResearch, Grasslands campus in Palmerston North used in preliminary experiments.

Table 3.1. Summary of occurrence of all 20 isolates within large clusters of 16S rRNA gene sequences derived from pasture-fed cows.

Clusters	Number of sequences in cluster	16S Reference sequence	Isolates*									Total
1	383	<i>Butyrivibrio fibrisolvens/hungatei</i>	AB2010	MB2006	DAZ1011	MA3006	MB2003	DAZ1010	MA2014	MB2014	MB3002	9
2	247	<i>Pseudobutyrvibrio xylanivorans</i>	DAZ1012	MA3010	DAZ2006	DAZ1009	AB2006	MA2017	DAZ1006			7
3	245	<i>Pseudobutyrvibrio xylanivorans</i>	MA2006	MA3014								2
4	212	<i>Selenomonas ruminantium</i>	AB3002									1
5	121	<i>Lachnospiraceae</i>	AB2031									1

Each isolate is grouped according to their closest relative deduced via 16S rRNA sequence. *Isolate codes: The first two letters of the isolate designation correspond to month (May or August), and to cow (cows A, B, C, D, E were located at Lye farm, Dairy NZ, Hamilton). The first number designates the dilution factor of the inoculum for the culture and the remaining 3 numbers indicate the culture number. The DAZ notation refers to a cow located at the AgResearch, Grasslands campus in Palmerston North used in preliminary experiments.

3.2 Characterisation of candidate isolates

Initial characterisation of the 20 original isolates was based on: co-occurrence in large clusters of 16S rRNA gene sequences retrieved from rumen contents, individual 16S rRNA gene sequences and PFGE patterns (Figure 3.2). From these characterisation results, 6 isolates were selected for further phenotypic characterisation.

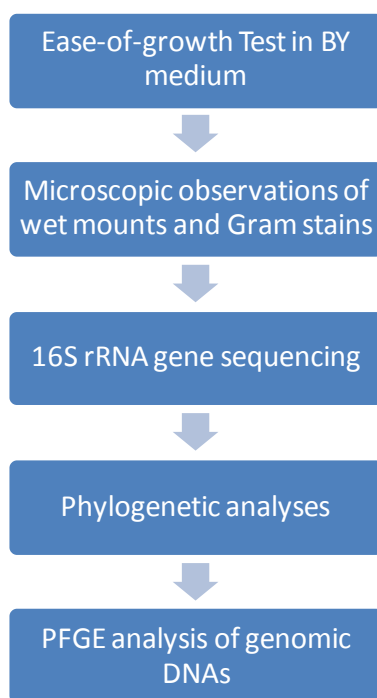


Figure 3.2. Outline of the initial phenotypic characterisation experiments of the 20 original isolates.

The 20 isolates selected for analysis were revived from frozen stocks by inoculating 0.5 ml of the thawed culture into BY medium and incubating overnight at 37°C. Thirteen cultures grew after overnight incubation and displayed similar optical densities. Isolates AB3002, DAZ1009, MA2017 and DAZ1011 formed sediment at the bottom of the tube after overnight growth but, when inverted displayed even turbidity. Isolate MB2003 displayed the best growth after overnight incubation and the culture displayed rope-like consistency once inverted. Isolates AB2031 and DAZ2006 did not grow overnight and were incubated for a longer period. Isolate AB2031 displayed a small amount of growth after two days of incubation while DAZ2006 did not grow in BY medium, after three repetitions of the ease-of-growth test. Thus isolates AB2031 and DAZ2006 were eliminated from further study based on their failure to meet the ease of growth requirement.

3.2.1 Microscopic observations of wet mounts and Gram stains

All of the isolates (except DAZ2006) were inoculated onto RM02 medium and incubated overnight at 37°C. Wet mounts were prepared for each isolate and observed using phase contrast microscopy (Table 3.2).

Table 3.2. Microscopic observations of isolates.

Isolate	Cell characteristics	Gram Stain
AB2006	medium length rod, slightly curved, mainly form pairs and few long chains	- ve
AB2010	small length rod, slightly curved, mainly single cells	- ve
AB2031	medium length rod, slightly curved, mainly form pairs and few clusters	- ve
AB3002	large length rod, curved, mainly single cells, few pairs	- ve
DAZ1006	medium length rod, slightly curved, mainly form pairs	- ve
DAZ1009	medium length rod, slightly curved, mainly single cells, few pairs and few clusters	- ve
DAZ1010	medium length rod, straight, mainly single cells, some pairs	- ve
DAZ1011	small length rod, slightly curved, mainly form pairs and few chains	- ve
DAZ1012	small length rod, slightly curved, predominantly single cells	- ve
MA2006	medium length rod, slightly curved, mainly form pairs and few large chains	- ve
MA2014	small length rod, slightly curved, mainly form clusters	+ ve
MA2017	small length rod, slightly curved, mainly form pairs	- ve
MA3006	medium length rod, slightly curved, mainly form pairs	- ve
MA3010	medium length rod, slightly curved, mainly form pairs and few large chains, cocci may also be present possibly contaminated	- ve
MA3014	small length rod, slightly curved, mainly form large chains and few clusters	- ve
MB2003	small length rod, curved, mainly forms pairs and few large chains and clusters	- ve
MB2006	small length rod, slightly curved, mainly form pairs and few chains	- ve
MB2014	small length rod, slightly curved, mainly form pairs	- ve
MB3002	medium length rod, slightly curved, mainly single cells, few pairs	- ve

The samples were very similar in their morphological characteristics and varied only in their size and tendency to form chains, pairs or to exist as single cells. All cultures stained Gram negative, except for MA2014 which stained Gram positive. Overnight cultures in RM02 medium were streaked onto RM02 agar plates and allowed to grow overnight, anaerobically. Individual, well separated colonies were restreaked and checked for purity via microscopic examination. Pure cultures of each isolate were stored at -85°C in 10 ml Hungate tubes containing 5 ml RM02 medium supplemented with 40% v/v glycerol.

3.2.2 16S rRNA gene sequencing

Genomic DNAs were extracted from each of the bacterial isolates and their 16S rRNA genes amplified using 16S rRNA primers (see 2.2.4) and sequenced by the Alan Wilson Centre at Massey University. All isolates produced single, clear PCR product bands when analysed by agarose gel electrophoresis, with concentrations of approximately 40-92 ng/μl. The purified PCR products were sent for 16S rRNA sequencing with four 16S rRNA gene primers (fD1, rD1, F1 and R3) which generate almost full length 16S rRNA gene sequences. The sequence analysis results are shown in Table 3.3. Seventeen of the bacterial isolate sequences matched closely with *Butyrivibrio* or *Pseudobutyrvibrio* 16S rRNA gene sequences. One isolate, AB3002, matched most closely with *Selenomonas ruminantium*. Isolates AB2006 and DAZ1012 produced poor quality sequences from two separate sequencing attempts. Searches with these sequences against both the RDP and BLASTN databases produced inconsistent results, thus isolates AB2006 and DAZ1012 were excluded as candidates for further study.

3.2.3 Phylogenetic analyses

The 16S rRNA sequences of the remaining 16 candidate isolates were aligned using ClustalW and their phylogenetic relationships determined using MEGA4 software and shown as a tree (Figure 3.3). The bottom portion of the tree shows that 9 isolates grouped with the *B. hungatei* species. The isolates MA2006, MA3014, MA3010, and DAZ1009 grouped with the *Pseudobutyrvibrio* sp. and two isolates (MA2017 and DAZ1006) grouped with the *P. ruminis*. Isolate AB3002 clustered close to *Selenomonas ruminantium*.

Table 3.3. RDP and BLASTN search results for 16S rRNA gene sequences of each bacterial isolate.

Isolate	RDP ^a	NCBI nr ^b
AB2006 (1)	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.701)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (100%)
AB2006 (2) (poor)	<i>bacterium</i> YE59 (0.596)	Uncultured bacterium clone SJTU_D_12_11 (100%)
AB2010	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.966)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (99%)
AB3002	<i>Selenomonas ruminantium</i> ; S206; (0.881)	<i>Selenomonas ruminantium</i> strain: S206 (98%)
DAZ1006	<i>Pseudobutyrvibrio ruminis</i> ; pC-XS7; (0.999)	<i>Pseudobutyrvibrio ruminis</i> strain pC-XS7 (100%)
DAZ1009	<i>Pseudobutyrvibrio ruminis</i> ; pC-XS7 (0.949)	<i>Pseudobutyrvibrio ruminis</i> strain pC-XS7 (98%)
DAZ1010	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.968)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (99%)
DAZ1011	<i>Butyrivibrio fibrisolvens</i> ; NCDO 2398 (0.969)	<i>Butyrivibrio fibrisolvens</i> (strain NCDO 2398) (99%)
DAZ1012 (1) (poor)	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.880)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (100%)
DAZ1012 (2) (poor)	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.754)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (98%)
MA2006	<i>Butyrivibrio fibrisolvens</i> ; 1.230 (0.950)	<i>Pseudobutyrvibrio ruminis</i> strain pC-XS7 (98%)
MA2014	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.966)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (99%)
MA2017	<i>Pseudobutyrvibrio ruminis</i> ; pC-XS7 (0.964)	<i>Pseudobutyrvibrio ruminis</i> strain pC-XS7 (99%)
MA3006	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.968)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (99%)
MA3010	<i>Pseudobutyrvibrio ruminis</i> ; pC-XS7 (0.974)	<i>Pseudobutyrvibrio ruminis</i> strain pC-XS7 (99%)
MA3014	<i>Butyrivibrio fibrisolvens</i> ; 1.230 (0.957)	<i>Butyrivibrio fibrisolvens</i> (100%)
MB2003	<i>Butyrivibrio fibrisolvens</i> ; NCDO 2398 (0.895)	<i>Butyrivibrio fibrisolvens</i> (strain NCDO 2398) (99%)
MB2006	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.966)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (100%)
MB2014	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.964)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (99%)
MB3002	<i>Butyrivibrio fibrisolvens</i> ; M55 (0.960)	<i>Butyrivibrio fibrisolvens</i> isolate M55 (100%)

^aThe RDP database was searched using the SEQ MATCH algorithm. Numbers in brackets indicate percentage similarity values. ^bThe NCBI nr (<http://blast.ncbi.nlm.nih.gov/Blast>) database was searched using the BLASTN algorithm. Numbers in brackets indicate % identity.

3.2.4 PFGE analysis of genomic DNAs

Pulse-Field Gel Electrophoresis (PFGE) was carried out on restriction enzyme digests of DNAs isolated from 18 isolates (all except DAZ2006 and AB2031) to examine genome differences among the bacterial isolates. Genomic DNAs digested with *Apa*I (recognition site GGGCC|C) showed that all isolates have different restriction digestion profiles or patterns. Ten samples that had the clearest bands and produced the highest quality PFGE profiles were selected for further study. These 10 isolates were: MA3010, MA3014, DAZ1011, MB2003, MB2014, DAZ1010, AB2010, MB3002, MB2006 and MA2014 (Figure 3.4). The remaining 8 samples produced poor PFGE profiles and were eliminated as candidates for further study after multiple attempts. PFGE plugs of the 10 best samples were run uncut, on a 1% agarose gel to determine if there was evidence of extra chromosomal elements such as plasmids, megaplasmids and secondary chromosomes (Figure 3.5). All samples except MB2014 produced a band near the top of the gel which represents uncut chromosomal DNA. In addition, MB2003 produced two clear bands below the chromosomal band which, possibly represent megaplasmids. MA3010 also produced two additional bands, while MA3014 produced one band. DAZ1011 also produced a single band however the DNA preparation was partially degraded and produced a smeared appearance in the gel.

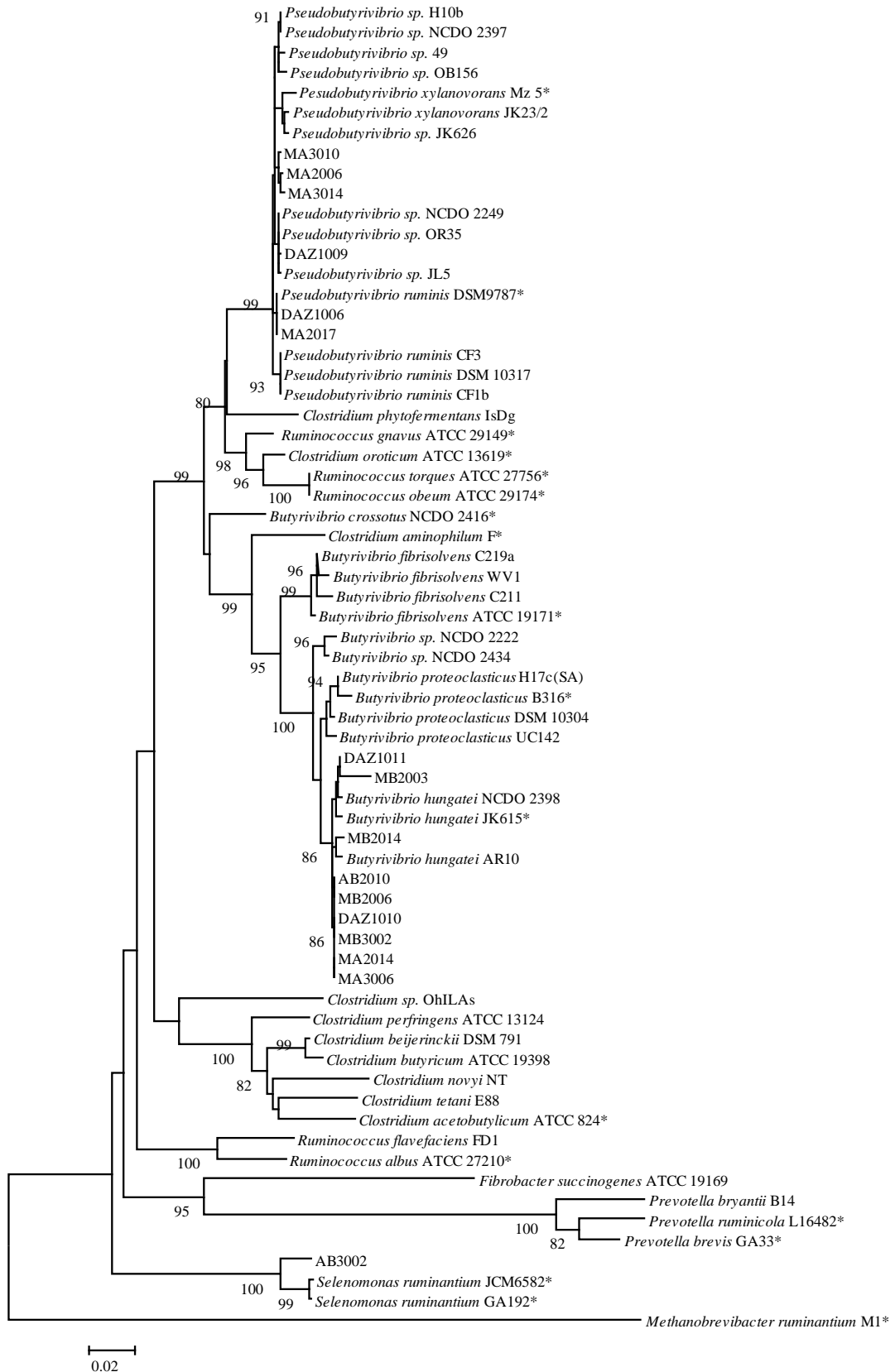


Figure 3.3. Phylogenetic tree based on 16S rRNA sequence data showing the 16 remaining isolates. *Indicates type strain. 16S rRNA gene phylogeny was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) with distances calculated using the Kimura 2-parameter method (Kimura, 1980) and implemented in MEGA4 (Tamura *et al.*, 2007). *M. ruminantium* M1 was used as an outgroup. Percentage bootstrap values were calculated from 10,000 replicates (Felsenstein, 1985) and are shown at nodes if >80%.

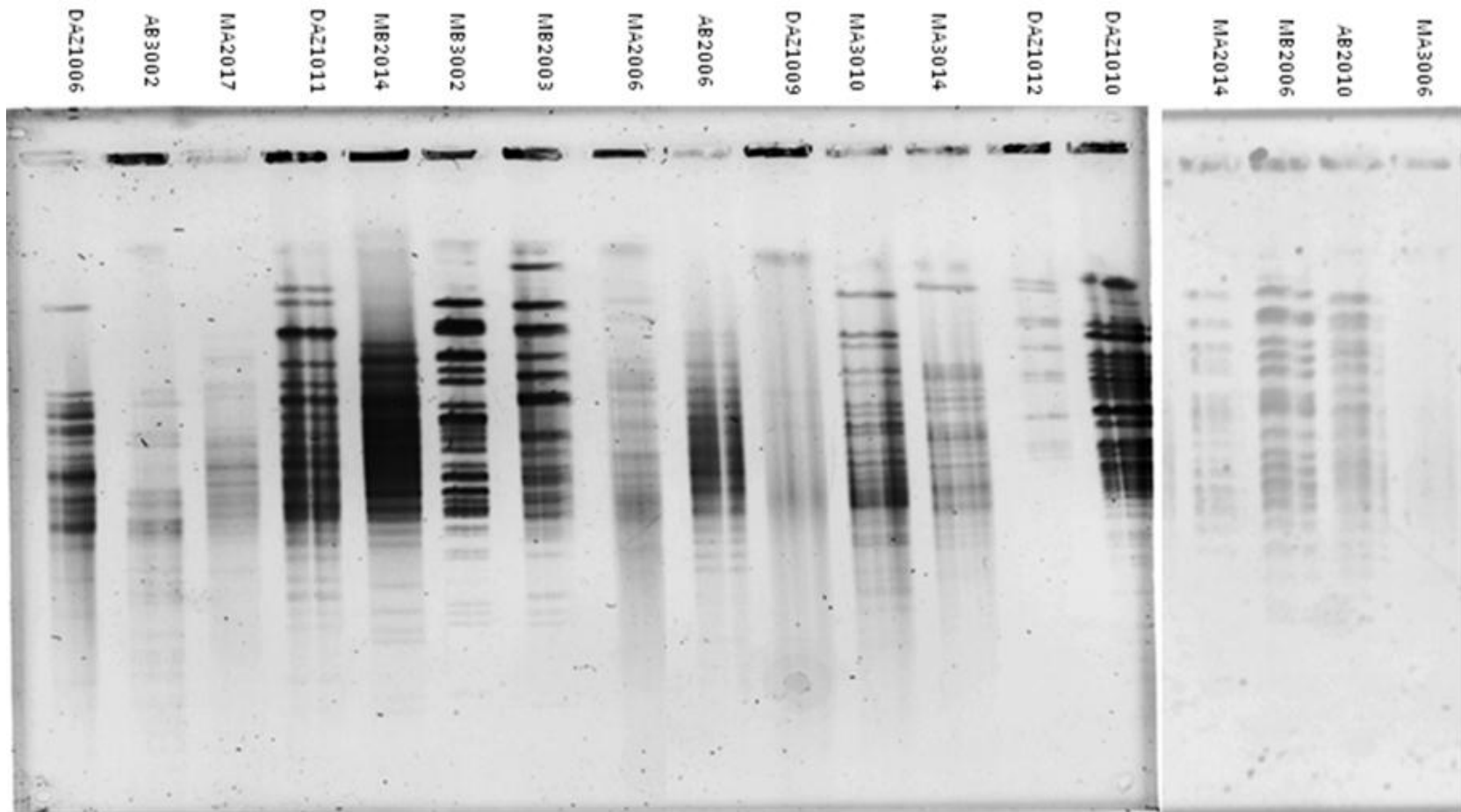


Figure 3.4. 1% Agarose PFGE gels of 18 candidate isolates. DNAs were digested with *ApaI*. The gel was run for 20 hr with the pulse time ramped from 1 to 30 sec.

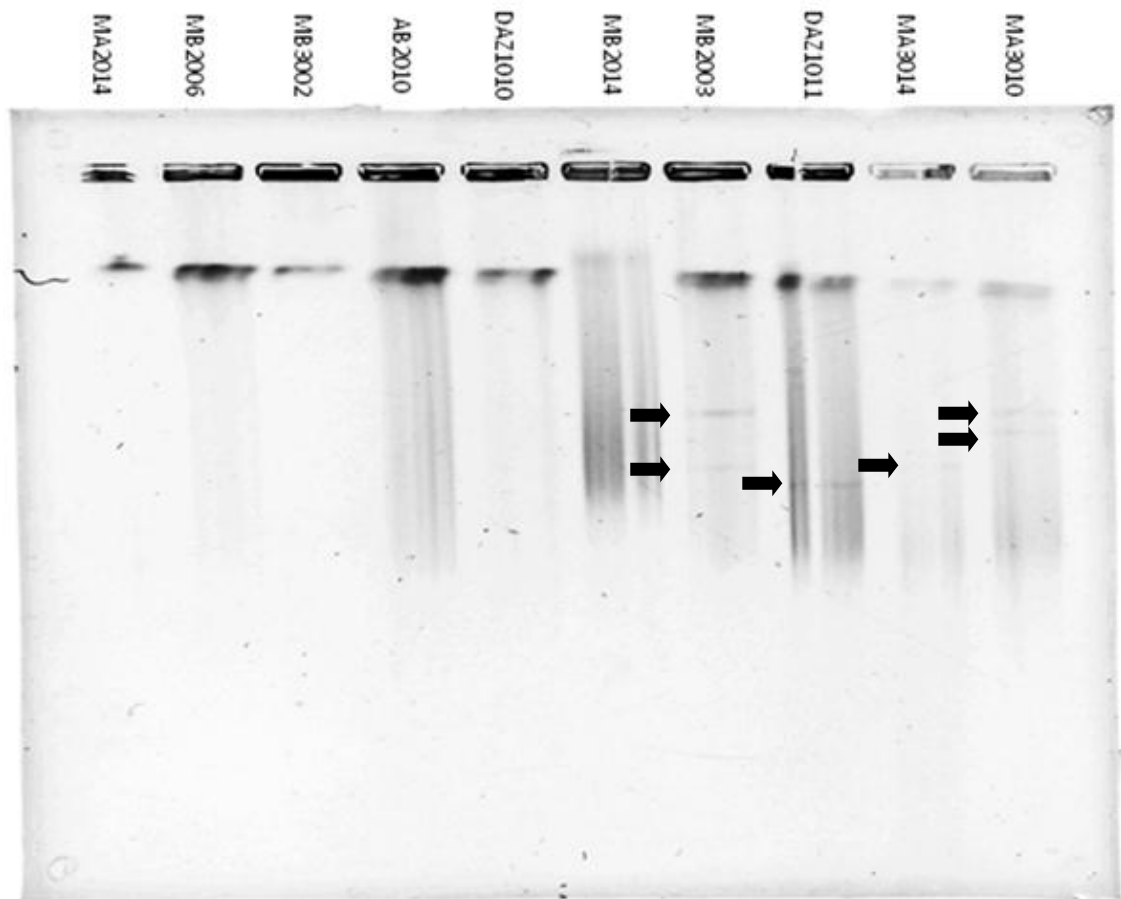


Figure 3.5. 1% Agarose PFGE gel of uncut genomic DNA from 10 selected isolates. The gel was run for 20 hr with the pulse time ramped from 1 to 30 sec. Arrows represent putative megaplasmids.

3.3 Final strain selection and phenotypic characterisation

From the original 20 strains, the list of candidate isolates was narrowed down to 6 strains based on; ease of growth, co-occurrence in large clusters of 16S rRNA gene sequences retrieved from rumen contents, individual 16S rRNA gene sequences and PFGE patterns. The isolates selected for further phenotypic characterisation were MB2006, AB2010, MB2003, MA3014, MA3010 and AB3002. Phenotypic characterisation was divided into four main experiments that provided sufficient information to choose two final candidate isolates for whole-genome sequencing.

3.3.1 Carbon source utilisation

Experiments were performed to determine carbon-source utilisation for each of the bacterial isolates (Table 3.4). Overall cultures displayed optimal growth on cellobiose and sucrose, and all cultures were able to utilise arabinose, dextrin, glucose, and pectin to achieve some level of growth. Substrates that all cultures could not utilise include: myo-inositol, mannitol and sorbitol.

Table 3.4. Carbon source utilisation of the six candidate isolates.

Substrate	<i>Butyrivibrio hungatei</i>			<i>Pseudobutyrvibrio xylanivorans</i>		<i>Selenomonas ruminantium</i>
	AB2010	MB2003	MB2006	MA3010	MA3014	AB3002
Amygdalin	++	+	-	++	++	++
Arabinose	++	++	+	++	+	+
Cellobiose	++	++	++	++	++	++
Dextrin	++*	++*	++*	++*	++*	++*
Fructose	++	-	++	++	++	++
Galactose	+	++	-	++	++	++
Glucose	++	++	++	++	++	++
Glycerol	+	-	-	+	-	+
Glycogen	-*	-*	-*	-*	-*	-*
myo-Inositol	-	-	-	-	-	-
Inulin	+*	++*	++*	++*	++*	++*
Lactose	-	++	-	-	+	++
Maltose	+	++	+	++	++	++
Mannitol	-	-	-	-	-	-
Mannose	-	-	+	+	+	++
Melezitose	+	-	-	+	-	+
Melibiose	+	-	-	++	++	++
Pectin	++*	++*	++*	++*	++*	++*
Raffinose	+	-	-	++	++	++
Rhamnose	+	-	-	+	-	+
Ribose	+	-	-	+	-	+
Salicin	++	++	++	++	++	+
Sorbitol	-	-	-	-	-	-
Starch	++*	+*	+*	+*	++*	++*
Sucrose	++	++	++	++	++	++
Trehalose	++	-	++	-	-	++
Xylan	+	++*	+*	+*	++	++*
Xylose	++	++	++	++	++	+

Positive controls were individual strains grown in RM02 with the GenRFV supplemental mixture. Growth: ++, 0.5-1.0 ΔOD_{600} reading, +, 0.2-0.5 ΔOD_{600} reading and -, 0-0.2 ΔOD_{600} reading. *Insoluble substrates.

3.3.2 Insoluble substrate utilisation

Several of the substrates commonly utilised by rumen bacteria are insoluble which prevents estimation of their use by measurement of change in optical density. To assess utilisation of the insoluble substrates glycogen, pectin, inulin, cellulose, dextrin, starch and xylan, cultures containing these substrates were assessed for growth using visual and microscopic examination (Table 3.5). Isolates MB2003, MA3014 and AB3002 were able to grow well on xylan, MB2003, MA3014, MA3010, MB2006 and AB3002 were all able to grow on all insoluble substrates except glycogen. AB2010 did not grow in glycogen and displayed poor growth in all other insoluble substrates.

Table 3.5. Wet mount observations of insoluble carbon source utilisation of the six candidate isolates.

Substrate	Isolates observations*					
	AB2010	MB2003	MA3014	AB3002	MA3010	MB2006
Glycogen	DC, -, P	DC, -, P	DC, -, P	DC, -, P	DC, -, P	DC, -, P
Pectin	+, P	++, P	++, P	++, P	++, P	++, P
Inulin	+, P	++, P	+, P	++, P	++, P	++, P
Cellulose	+, P	+, P	+, P	+, P	+, P	+, P
Dextrin	+, P	++, P	++, P	++, P	+, P	+, P
Starch	+, P	++, P	++, P	++, P	++, P	++, P
Xylan	+, P	++, P	++, P	++, P	+, P	+, P

*DC, dead cells, (cells appeared shrivelled and disfigured, and showed no movement) -, no growth, +, poor growth, ++, good growth and P, pure culture.

3.3.3 Growth curves for six candidate isolates on optimal substrate (cellobiose)

The carbon-source utilisation experiments indicated that all the bacterial strains grew well on cellobiose. Therefore cellobiose was used as substrate to examine growth of the isolates over a 24 hr period. All six cultures followed very similar growth patterns. The exponential phase of growth for all cultures was between 4 and 8 hr of growth. Maximum cell density for all cultures was reached at 8 to 10 hr. Isolates MB2003, MB2006 and AB3002 displayed the highest maximum ODs among the six cultures. The stationary phase was from 10 to 24 hr (Figure 3.6).

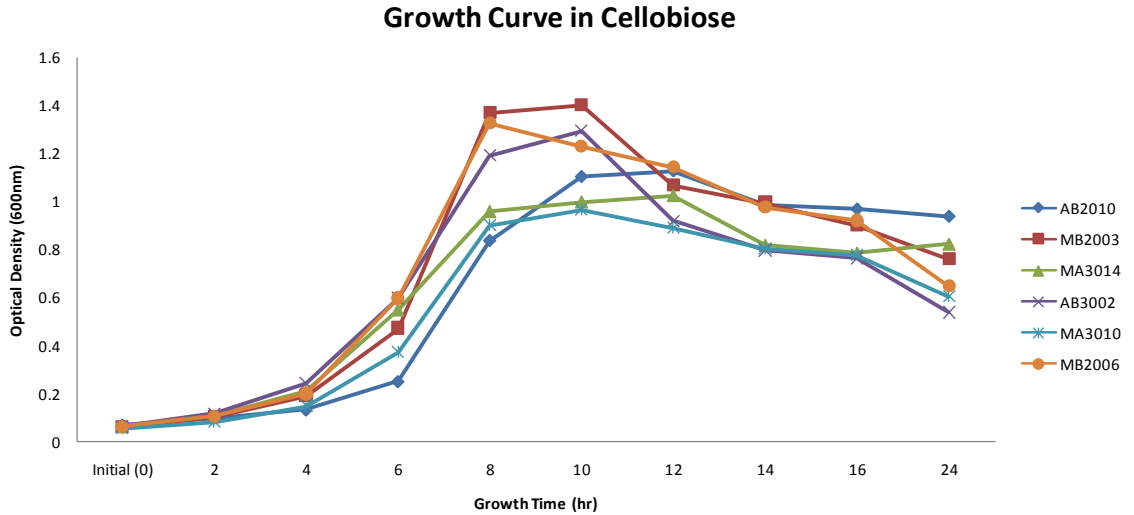


Figure 3.6. Bacterial growth curves of candidate isolates grown on cellobiose.

3.3.4 VFA analysis

VFA analysis was carried out to determine the fermentation end products of each of the isolates and to help confirm their taxonomic placement (Table 3.6). The *B. hungatei* isolates (MB2003, MB2010 and MB2006) all produced large amounts of formic acid (approximately 15-20 mM), and acetic and *n*-butyric acids both at approximately 5 mM. *P. xylanivorans* isolates (MA3014 and MA3010) produced similar amounts of *n*-butyric acid, less than 5 mM of formic acid, and substantial lactic acid. The characteristic differentiation between *Butyrivibrio* sp. and *Pseudobutyrvibrio* sp. is that *Pseudobutyrvibrio* sp. produce large amounts of lactic acid (approximately 10 mM) as opposed to large amounts of formic acid. The *Selenomonas ruminantium* produced approximately 15 mM of lactic acid and 2 mM of succinic acid, which differ from other two groups and is characteristic for *Selenomonas ruminantium* species (Martin and Dean, 1989).

Table 3.6. VFA analysis for the six candidate isolates. VFA concentrations are expressed as the average of two duplicate cultures.

Isolates	VFA concentration (mM)					
	Formic	Acetic	Propionic	n-Butyric	Succinic	Lactic
AB2010	20.9	4.0	0.0	7.0	0.0	0.4
MB2003	16.4	3.6	0.1	4.7	0.0	0.1
MB2006	17.2	3.3	0.0	5.0	0.0	0.5
MA3014	4.3	0.0	0.1	5.4	0.0	8.2
MA3010	2.4	0.0	0.0	5.3	0.0	9.8
AB3002	0.0	0.0	0.0	0.0	2.1	14.6

3.3.5 Moles %G+C content

The moles %G+C content of the DNA of each of the isolates was determined to help assign them to a taxonomic group. The %G+C content for five of the six isolates corresponded well to their taxonomic position assigned by 16S rRNA gene sequencing. The *B. hungatei* isolates (MB2003, MB2010 and MB2006) all produced moles % G+C contents of 38-39% which are typical of *Butyrivibrio* species. Isolate MA3014 gave 38% G+C content that is typical of *Pseudobutyrvibrio* species. The AB3002 isolate gave a moles %G+C content of 51% that is consistent with that of *Selenomonas ruminantium* (Table 3.7). However no data could be generated for MA3010 due to sample error.

Table 3.7. Moles %G+C content for the six candidate strains.

Isolate	Moles %G+C
AB2010	39 ±0.04
MB2003	38±0.06
MB2006	39±0.06
MA3014	38±0.01
MA3010	N/A
AB3002	51±0.02

3.3.6 Adherence

One of the main prerequisites for plant cell wall degradation is that the bacteria physically adhere to the plant material. Two methods of microscopy were used to visualise the adherence of the isolates MB2003 and MA3014 on plant material; observation of bacteria (stained with DAPI under a light microscope) grown with neutral detergent fraction (NDF) or sisal string, and Scanning Electron Microscopy of bacteria adherent to NDF. *B. hungatei* MB2003 displayed good growth in BY media in the presence of NDF (Figure 3.7 A and B) and showed adherence to the ends of the NDF material. There were signs that the NDF plant material had been degraded.

Light microscopy and SEM images of *P. xylanivorans* MA3014 grown in BY media in the presence of NDF (Figure 3.7 C and D, and Figure 3.9) showed that MA3014 adheres to the ends of NDF material, where the disintegration of the structure of the NDF edges was much greater than observed with MB2003. The images indicate degradation of NDF plant material as the MA3014 bacteria have colonised and formed a coat around the isolated piece of NDF resulting in breakdown of the plant material. Figure 3.7 E and F displays *B. proteoclasticus* B316 cultures adhered to the ends of the NDF plant material. The NDF structures appear more intact compared to the MB2003 and MA3014 cultures.

SEM was carried out as outlined in Methods Section 2.2.10. Figure 3.8 A and B show the adherence of MB2003 at the ends of the NDF material. A higher magnification (Figure 3.8 C and D) clearly illustrate observed adherence. The images also show the medium length, and slightly curved rod shape of MB2003 cells.

Compared to the MB2003 cultures, MA3014 appears to grow better in BY media, and attach more prominently at the ends of the NDF fibres (Figure 3.9 A and B). A large amount of material associated with the MA3014 cells can be clearly seen as a stringy dense matrix (Fig 3.9 C and D). The MA3014 cells appear shorter in length, less curved and thicker than the MB2003 cells. No cells were observed attached to the NDF plant material in negative controls.

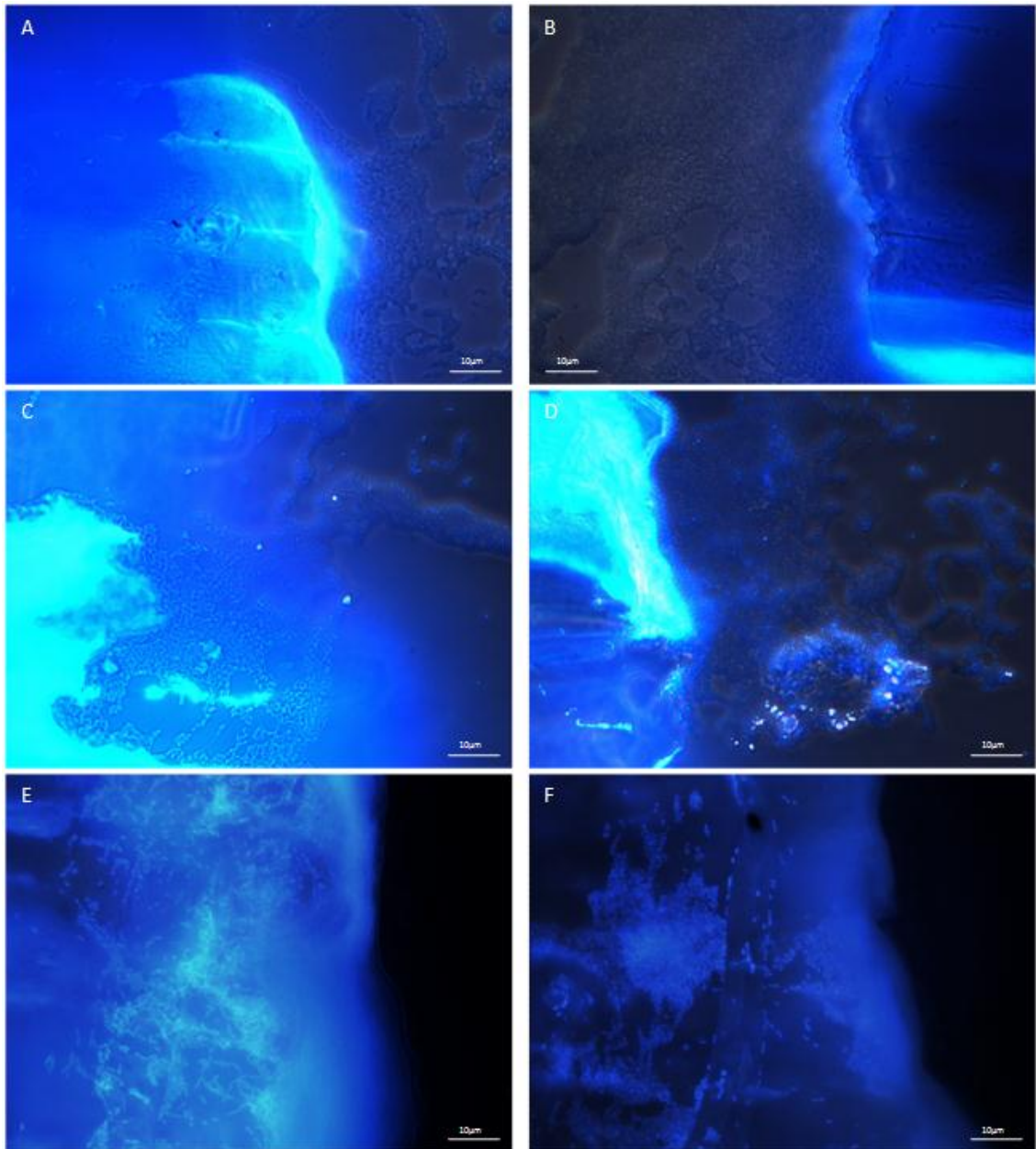


Figure 3.7. Fluorescence microscopy images of DAPI-stained MB2003 (A, B), MA3014 (C, D) and *B. proteoclasticus* B316 (E, F) cultures grown on BY media containing NDF plant material (100× magnification).

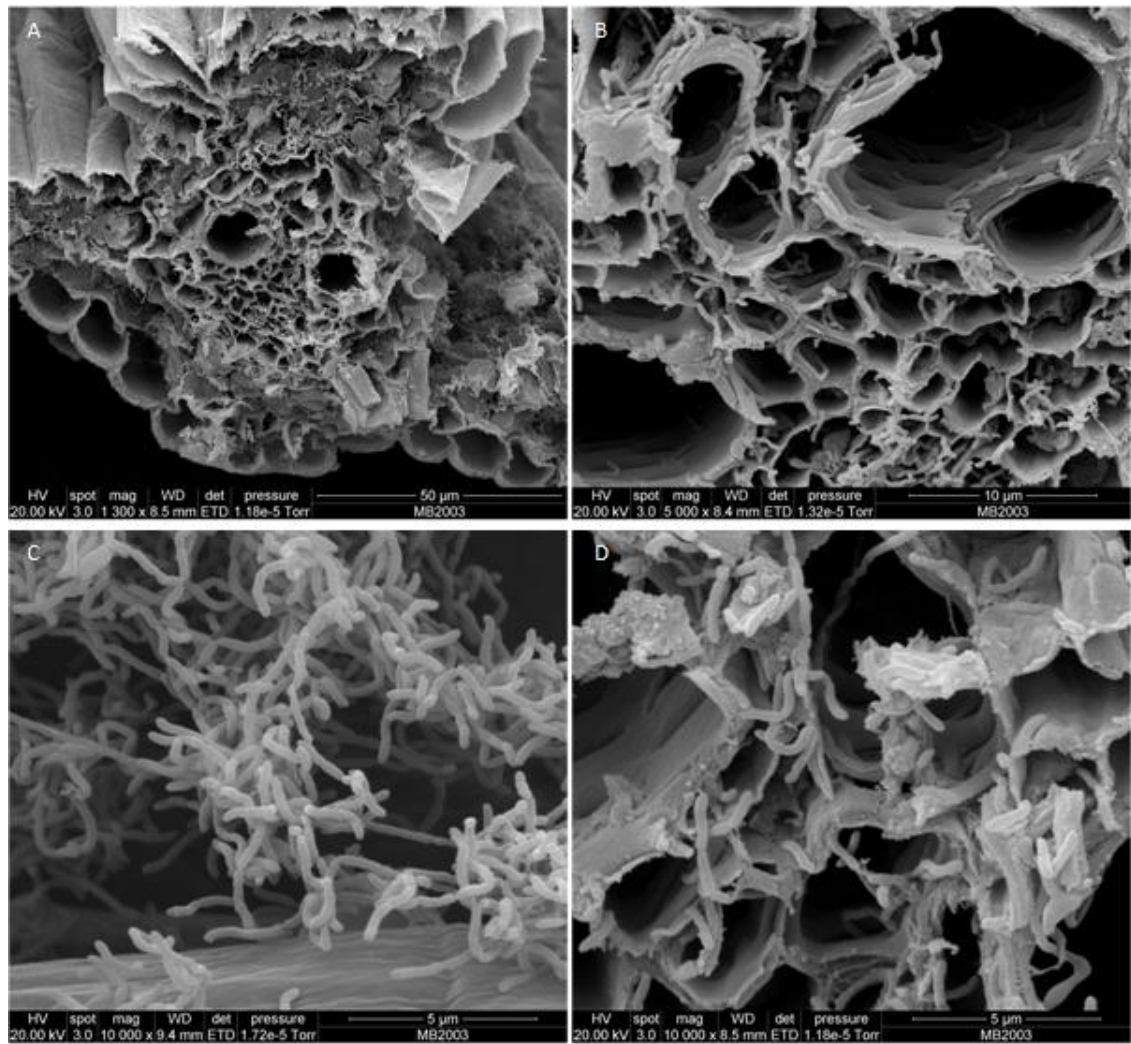


Figure 3.8. Scanning Electron Microscopy images of MB2003 cultures grown on BY media on NDF plant material. A, 1,300× magnification, B, 5,000× magnification, C and D, 10,000× magnification.

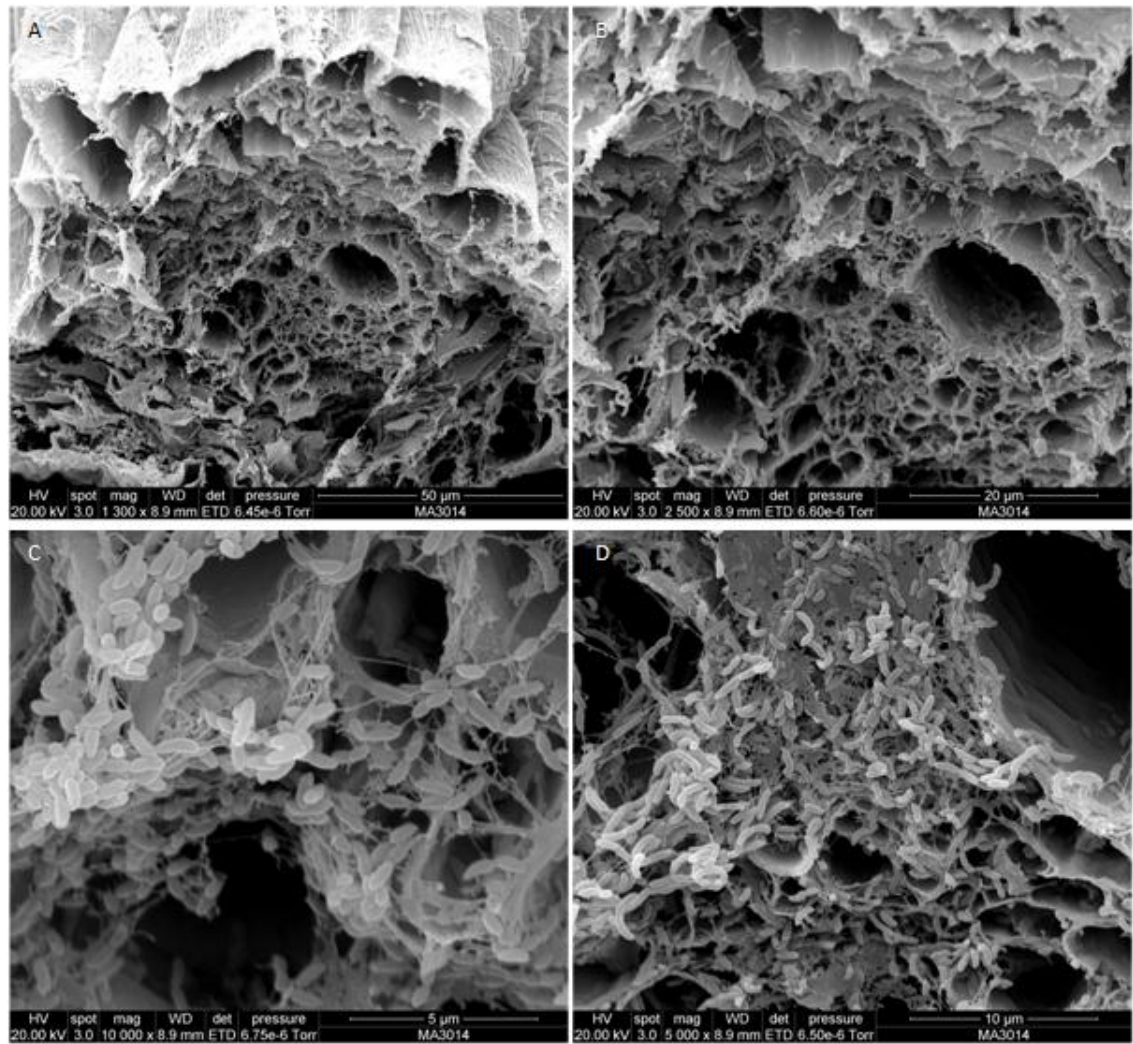


Figure 3.9. Scanning Electron Microscopy images of MA3014 cultures grown on BY media containing NDF plant material. A, 1,300× magnification, B, 2,500× magnification, C , 10,000× magnification and D, 5,000× magnification.

3.3.7 Taxonomic assignment

Two bacterial strains (MB2003 and MA3014) were selected for genome sequencing based on their phenotypic properties. These strains were identified as *Butyrivibrio hungatei* MB2003 and *Pseudobutyrvibrio xylanivorans* MA3014 by comparison of their properties with the taxonomic descriptions reported by Kopecny *et al.*, (2003).

3.4 Whole-genome sequencing of MB2003

Based on the phenotypic characterisation results, MA3014 and MB2003 were selected as the final genome sequencing candidates. Genomic DNA was extracted from both MA3014 and MB2003 as described in Methods Section 2.2.13. DNA purity and quality were checked and organism identities were confirmed by re-sequencing their 16S rRNA genes. High quality gDNA (1 mg) was shipped to Macrogen sequencing facility (Korea) and pyrosequenced using 454 Titanium paired-end sequencing technology. Initial examination of the sequence assembly information indicated that the MB2003 genome was smaller and contained fewer contigs (31 contigs in MB2003 compared to 116 in MA3014) arranged into fewer scaffolds compared to MA3014 (7 scaffolds in MB2003 and 13 in MA3014). Due to time constraints, MB2003 was chosen for further sequence assembly, gap closure and annotation.

3.4.1 Sequence statistics

The sequence statistics for the ½ plate of 454 Titanium paired-end sequencing obtained for MB2003 via Macrogen's whole-genome sequencing service are shown in Table 3.8.

Table 3.8. MB2003 sequence statistics.

Sequence characteristic	bp/%G+C
Average read length	377
Min read length	40
Max read length	891
Total number of reads	548373
Total base pairs sequenced	206681869
Ave. %G+C of the reads	39

The read length frequency graph is shown in Figure 3.10. The highest read length frequency was for read lengths of ~500 bp in length. The assembly of the sequence reads (sff file) was carried out in a one-step assembly using Newbler software (GC de novo Assembler). The initial assembly of the MB2003 sequence data resulted in 937 contigs (Table 3.9) which was an unusually high number of contigs. The contigs had an average %G+C content of 55%, while the experimentally determined %G+C content for MB2003 was 38.2%. The assembly results appeared to be inconsistent with a genome of approximately 4 Mb with a 38.2% G+C content, so we investigated the composition of the contigs further.

Frequency vs. Read Length

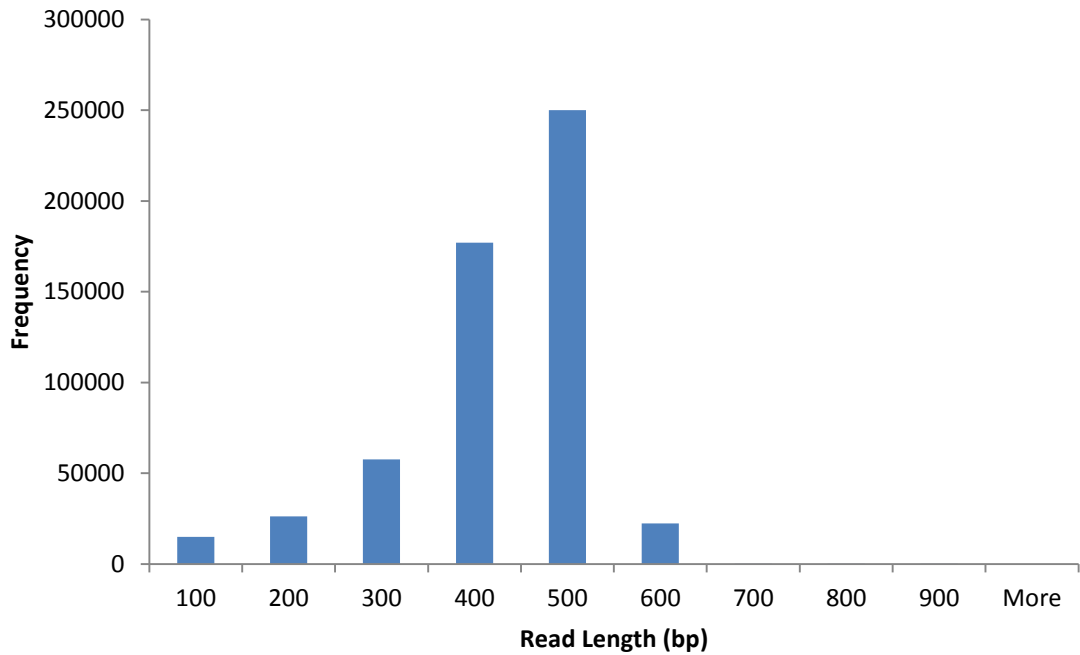


Figure 3.10. Bar chart of read length frequency for input files for ½ plate of 454 Titanium paired-end sequencing.

Table 3.9. MB2003 contig statistics.

Contig characteristic	Contig Length (bp)
Sum of contigs	4096761
Min	500
Max	368100
Average %G+C	55
Contig Number	937

A distribution of contig length showed that the vast majority of the 937 contigs were approximately 1000 bp in size (Figure 3.11). Furthermore, MacroGen informed us that there may have been a cross contamination of *Brucella* DNA from their sequencing facility into our samples, which may explain the large amount of small contigs and the unusually high %G+C content of 55%.

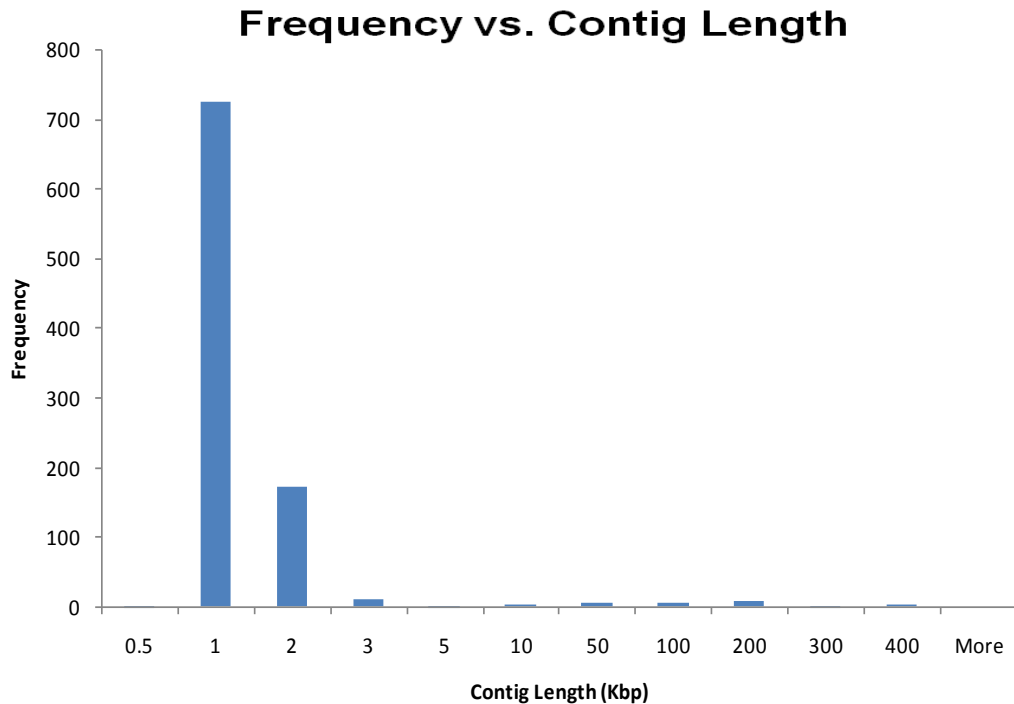


Figure 3.11. Distribution of contig length.

To identify contaminant contigs from the sequence assembly, a three-part analysis procedure was applied to all 937 contigs. Firstly, an arbitrary cut-off of 43% or above %G+C content was made for each contig. This identified 904 contigs (56% average %G+C content and an average size of 814 bp in length) as potential *Brucella* contig contaminants. *Brucella* species have %G+C contents ranging from 56 to 58% (Morgan and Corbel, 1976). Secondly, scaffold information obtained from the paired-end sequencing data was used to identify any contigs present in scaffolds. Contigs present in scaffold information are likely to represent the dominant MB2003 assembly. However, four contigs (contig02327, contig02328, contig01185 and 01186) that subsequently were identified as *Brucella* contamination were present in scaffold 121 and 77, respectively (BLASTX analysis see below). Thirdly, a BLASTX analysis against the NR database was carried out on all contigs to confirm the presence of *Brucella* sequence. Based on the combined results of the three approaches described above, 904 contigs identified as *Brucella* contaminated contigs were removed from the raw assembly. This contamination accounted for 734,762 bp in total which represents 17.9% of the original MB2003 sequencing data.

The arrangement of the 31 remaining MB2003 contigs into their appropriate scaffolds was determined by further analysis of the paired-end data information. Contigs were paired with their corresponding scaffolds which resulted in 4 major scaffolds and 3 single scaffolds containing all 31 contigs. These 7 scaffolds were arranged into the predicted order by carrying out BLAST searches of the ends of each of the contigs (800 bp) using the *B. proteoclasticus* B316 genome sequence as the scaffolding template.

By comparing the locus tags from BLAST analyses, the contigs at the ends of the scaffolds could be determined and thus the initial hypothetical order of scaffold arrangement was generated (Table 3.11). The top BLAST results for the ends of contig 00005f and 00005r were for bpr_III136 and III139 respective locus tags that are present in the *B. proteoclasticus* B316 chromid. The top BLAST hits for the ends of contig 00009f and contig 02243f were for plasmid replication proteins RepB. Thus contigs 00005, 00009 and 02243 appear to represent extrachromosomal elements. Combinations of PCRs were carried out for contigs at the end of each scaffold in order to determine the orientation of each scaffold to each other and also to close sequence gaps. PCR products from the reactions were sequenced and the new sequence added to the contig end. New primers were then designed near the new end of the contigs and the gap closure procedure was repeated.

After seven rounds of gap closure, the MB2003 draft genome contained a main chromosome with 7 gaps remaining, three of which correspond to rRNA regions (Figure 3.11), plus a 91 kb chromid (Bhu II), a small 6.3 kb plasmid (pNP6) and a 144 kb megaplasmid (pNP144). The three extrachromosomal replicons have all been closed and their sequence confirmed using PCRs. Once the genome is closed, an *in silico* restriction enzyme map will be carried out for MB2003 and assembly will be verified by PFGE.

Table 3.10. Arrangement of MB2003 contigs into scaffolds.

Scaffold Number	Contig Name	Contig Size (bp)	% G+C
scaffold001	contig00005	90577	37.6
scaffold002	contig00007	345280	39.9
scaffold002	contig00008	196580	41.9
scaffold003	contig00009	6344	35.7
scaffold115	contig02243	143881	36.9
scaffold116	contig02282	8953	38.9
scaffold116	contig02283	7091	42.7
scaffold116	contig02284	59856	41.1
scaffold116	contig02285	222644	40.3
scaffold116	contig02286	28584	40.8
scaffold116	contig02287	233430	39.5
scaffold116	contig02288	52481	39.7
scaffold116	contig02289	53306	40
scaffold116	contig02290	30640	40.1
scaffold121	contig02327	1686	53.1
scaffold121	contig02328	1136	50.6
scaffold121	contig02329	124873	40.8
scaffold121	contig02330	184603	39.8
scaffold122	contig02337	139170	39.1
scaffold122	contig02338	1097	40.7
scaffold122	contig02339	9032	39.3
scaffold122	contig02340	182228	38.1
scaffold122	contig02341	24938	38.6
scaffold122	contig02342	71625	38.8
scaffold122	contig02343	130003	39.2
scaffold122	contig02344	86794	39.3
scaffold122	contig02345	35293	40.4
scaffold122	contig02346	23544	39.8
scaffold122	contig02347	366544	39.7
scaffold122	contig02348	368100	39.8
scaffold122	contig02349	125618	40.2

Table 3.11. Summary of BLASTX search results of contig ends (800bp) for MB2003 genome. **Red** indicates contaminant contigs.

Contig Name	Accession	Locus_tag	Top blast Hit	Organism	% ID	E-value Score
contig00005f	YP_003832404.1	bpr_III136	GNAT family acetyltransferase/NUDIX domain-containing protein	<i>Butyrivibrio proteoclasticus</i> B316	100	3.00E-63
contig00005r	YP_003832407.1	bpr_III139	TetR family transcriptional regulator	<i>Butyrivibrioproteoclasticus</i> B316	93	1.00E-96
contig00007r	YP_003829901.1	bpr_I0574	Acetyltransferase	<i>Butyrivibrioproteoclasticus</i> B316	70	4.00E-46
contig00008f	ZP_02073755.1		hypothetical protein CLOL250_00503	<i>Clostridium</i> sp. L2-50	64	5.00E-23
contig00009f	YP_002286760.1		plasmid recombination protein	<i>Butyrivibrio fibrisolvens</i> OB157	41	4.00E-04
contig00009r	AAA23031.1		Replicase	<i>Campylobacter hyointestinalis</i>	34	2.00E-13
contig01185r	ZP_05955688.1		LOW QUALITY PROTEIN: outer membrane autotransporter barrel domain-containing protein	<i>Brucella pinnipedialis</i> B2/94	79	0
contig01186f	ZP_05189756.1		intimin/invasin family protein	<i>Brucella abortus</i> bv. 4 str. 292	100	e-147
contig02243f	YP_003832274.1	bpr_III001	replication initiation protein RepB1	<i>Butyrivibrio proteoclasticus</i> B316	55	1.00E-44
contig02243r	ZP_04455084.1		hypothetical protein GCWU000342_01100	<i>Shuttleworthia satelles</i> DSM 14600	48	6.00E-34
contig02282r	CBK94202.1		Helix-turn-helix.	<i>Eubacterium rectale</i> M104/1	47	7.00E-22
contig02290f	YP_003830557.1	bpr_I1236	hypothetical protein bpr_I1236	<i>Butyrivibrio proteoclasticus</i> B316	55	3.00E-51
contig02327r	ZP_03294132.1		hypothetical protein CLOHIR_02084	<i>Clostridium hiranonis</i> DSM 13275	90	3.00E-53
contig02330f	ZP_03801676.1		hypothetical protein COPCOM_03977	<i>Coprococcus comes</i> ATCC 27758	73	1.00E-04
contig02337r	YP_003830558.1	bpr_I1237	chemotaxis protein CheW	<i>Butyrivibrio proteoclasticus</i> B316	97	e-101
contig02349f	no hit					

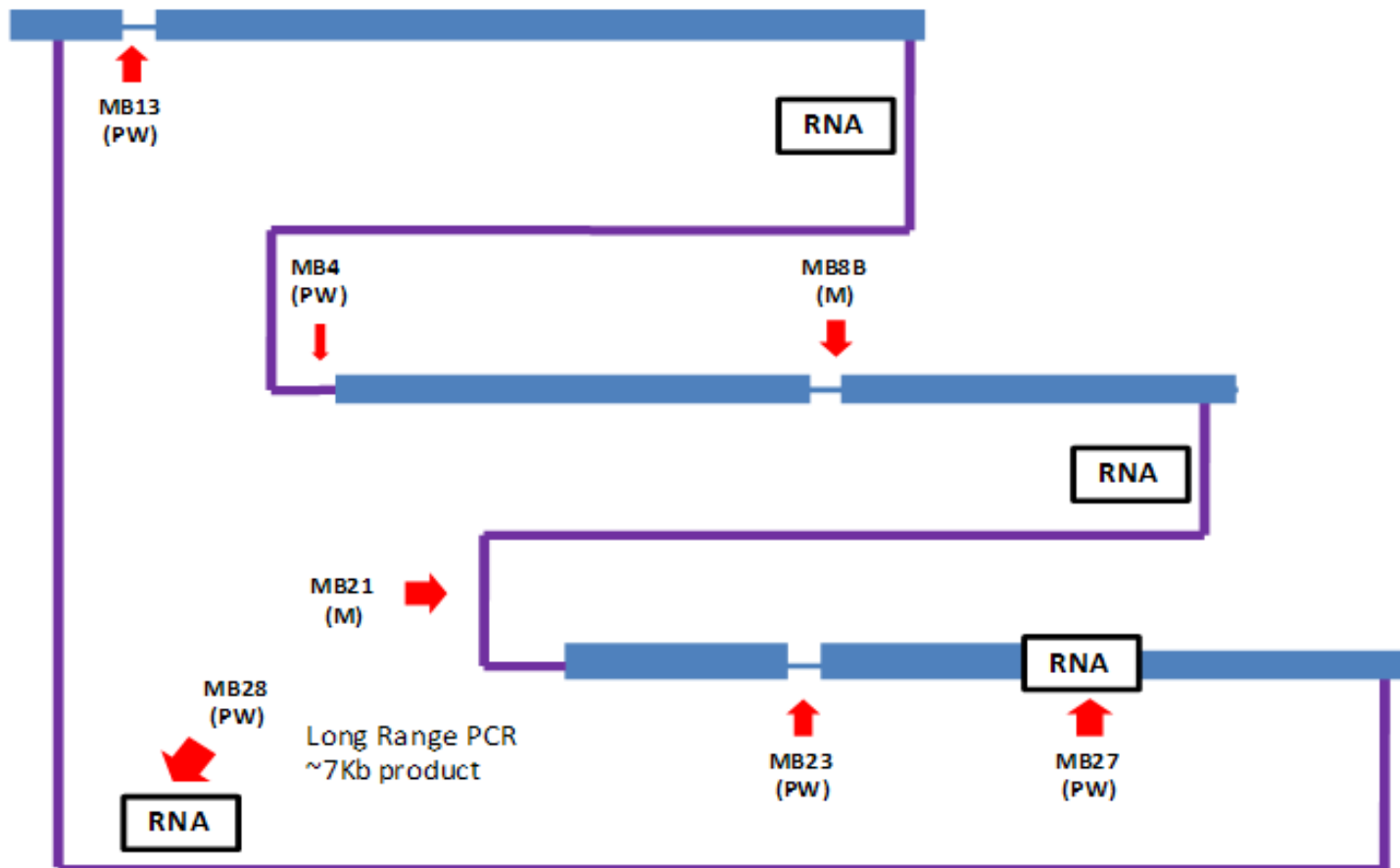


Figure 3.12. Diagrammatic representation of the MB2003 draft genome. Seven rounds of gap closure were completed for the MB2003 chromosome resulting in 1 super-scaffold representing 7 contigs and 7 gaps remaining to close. The total chromosome size is 3,129,625bp and total chromosome %G+C content is 39.86% (based on GAMOLA analysis). The blue boxes represent the contigs of the MB2003 chromosome. The purple lines show where the contigs are predicted to be attached to each other (not to scale). The RNA boxes indicate the positions of rRNA regions. The red arrows and associated MB labels indicate the positions of the gaps that are still to be closed, where PW indicates primer walking and M indicates that new primers need to be designed for these gap regions.

3.5 Comparison of the MB2003 genome to *B. proteoclasticus* B316

3.5.1 MUMmer plot

A MUMmer plot was used to align the incomplete draft MB2003 chromosome with the B316 complete chromosome. MUMmer is a whole genome alignment program that compares the genome of one organism with another by generating an alignment of their genomes in the form of a dot plot diagram. The plots reveal regions of conservation between the two genomes under comparison.

In a MUMmer plot the reference genome is plotted along the x-axis (in this case *B. proteoclasticus* B316) and the query genome is on the y-axis (MB2003 draft genome). Wherever the two sequences match a coloured dot is plotted on the chart, the forward matches are displayed in red and the reverse matches are displayed in blue. A straight line of dots starting from bottom left to the top right means that the two regions are exactly identical. The B316 vs MB2003 genome plot shows the presence of dots (blue and red combined) on a diagonal indicating that the two genomes contain many similar sequences and thus their genomes share many similar genes.

The plot also showed an X-shaped pattern known as an X-alignment (Figure 3.13). This X-shape is symmetrical about the putative origin of replication of the MB2003 and *B. proteoclasticus* B316 genomes and indicates that matching sequences tend to occur at the same distance from the origin/terminus of replication but not necessarily on the same side of the origin.

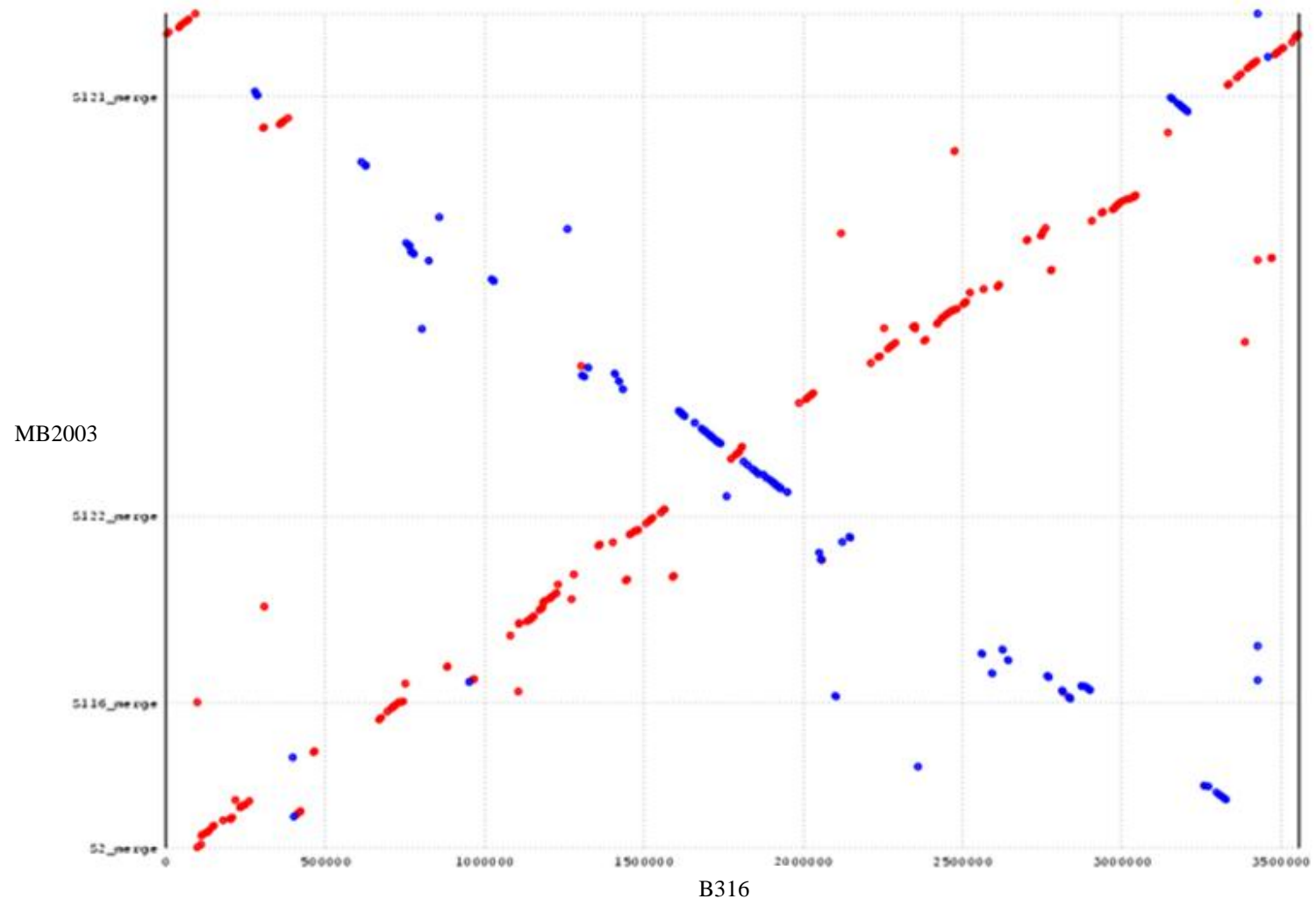


Figure 3.13. MUMmer plot of versus *B. proteoclasticus* B316 chromosomes.

3.6 Multi-replicon genome architecture

The draft *B. hungatei* MB2003 genome is currently in four replicons: the chromosome, chromid (Bhu II), megaplasmid (pNP144) and a small plasmid (pNP6). A comparison with *B. proteoclasticus* B316 and the draft MA3014 genome are shown in Table 3.12. The %G+C content for all four MB2003 replicons is similar, and is similar to the %G+C content of the *B. proteoclasticus* B316 replicons and the draft MA3014 genome. In total the MB2003 genome is more than 1 Mbp smaller than *B. proteoclasticus* B316, chromosome 400,000 bp smaller, the chromid is approximately 200,000 bp smaller and the megaplasms and plasmid are both smaller in size.

Table 3.12. Structure of *Butyrivibrio* and *Pseudobutyrvibrio* genomes.

Genome	Status	Organization	Contig No.	Size (bp)	GC %
<i>Butyrivibrio proteoclasticus</i> B316	Complete	Chromosome	1	3,554,804	40
		Chromid	1	302,358	40
		Megaplasmid	1	361,399	39
		Megaplasmid	1	186,325	38
				4,404,886	
<i>Butyrivibrio hungatei</i> MB2003	Draft	Chromosome	7	3,137,803	40
		Chromid	1	91,836	38
		Megaplasmid	1	143,881	37
		Plasmid	1	6,344	36
				3,374,113	
<i>Pseudobutyrvibrio</i> sp. MA3014	Draft	Nd	116	3,458,205	39

3.7 *Butyrivibrio hungatei* MB2003 genome annotation

Genome annotation was carried out as described in Methods Section 2.2.15. The draft *B. hungatei* MB2003 genome was run through GAMOLA, where autoannotation was carried out for all 2,982 genes in the four replicons.

3.7.1 MB2003 Chromid Bhu II

A chromid is an extrachromosomal element containing essential genes (Harrison *et al.*, 2010). *B. proteoclasticus* B316 was found to contain a single chromid but in general these elements are rare in Gram positive bacteria. The MB2003 chromid is the third largest replicon in this bacterium at 91,836 bp and 38% G+C content (Figure 3.14 and Table 3.13). The total coding percentage for the megaplasmid is 88% with 88 predicted genes in total. The chromid has the only copies of genes involved in fatty acid biosynthesis that are theoretically essential to the growth and survival of MB2003. The MB2003 chromid also contains four genes involved in carbohydrate degradation including a β -galactosidase (Bga42A), a β -glucosidase (Bgl3A), a feruloyl esterase (Est1A) and a polysaccharide deacetylase (Est4A). A full list of genes from the chromid is shown in Appendix I (Table A1).

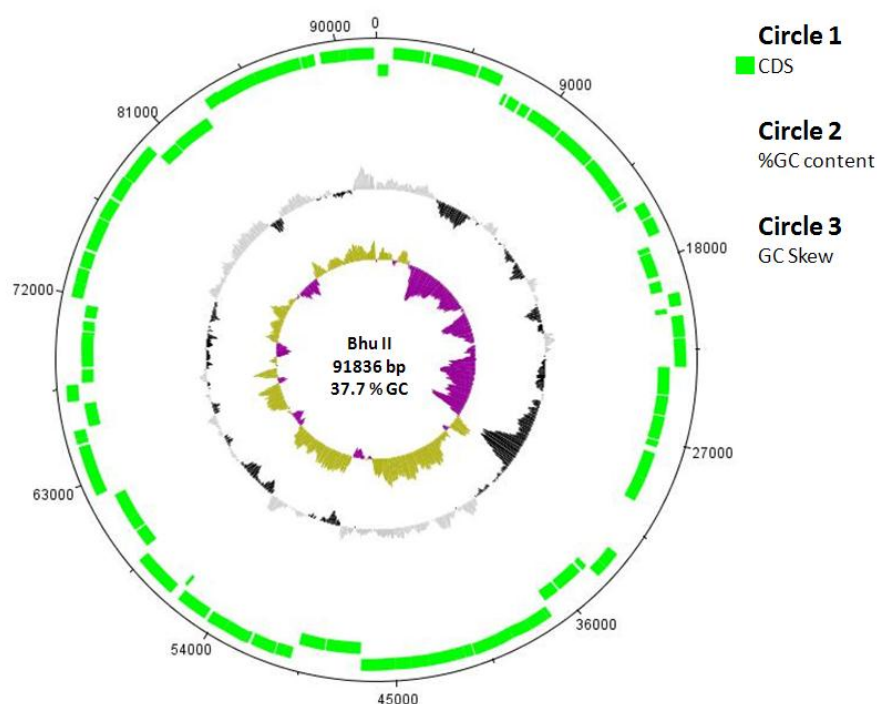


Figure 3.14. Circular plot of the MB2003 Bhu II chromid. The green colour coding of the genomic features in circle 1 represent Coding Sequence (CDS), %G+C content is shown in circle 2 and GC skew in circle 3.

Table 3.13. Sequence information of the MB2003 Bhu II chromid.

Total coding (bps)	80909
Total (bps)	91836
Coding %	88
Total number of genes	88

3.7.2 MB2003 Plasmid pNP6

The MB2003 pNP6 plasmid is the smallest of the four replicons at 6,344 bp and has a G+C content of 36% (Figure 3.15 and Table 3.14). The plasmid encodes only five genes, all of which appear to be involved in plasmid replication. pNP6 does not encode any genes involved in polysaccharide degradation. A full list of genes from the plasmid is shown in Appendix I (Table A2).

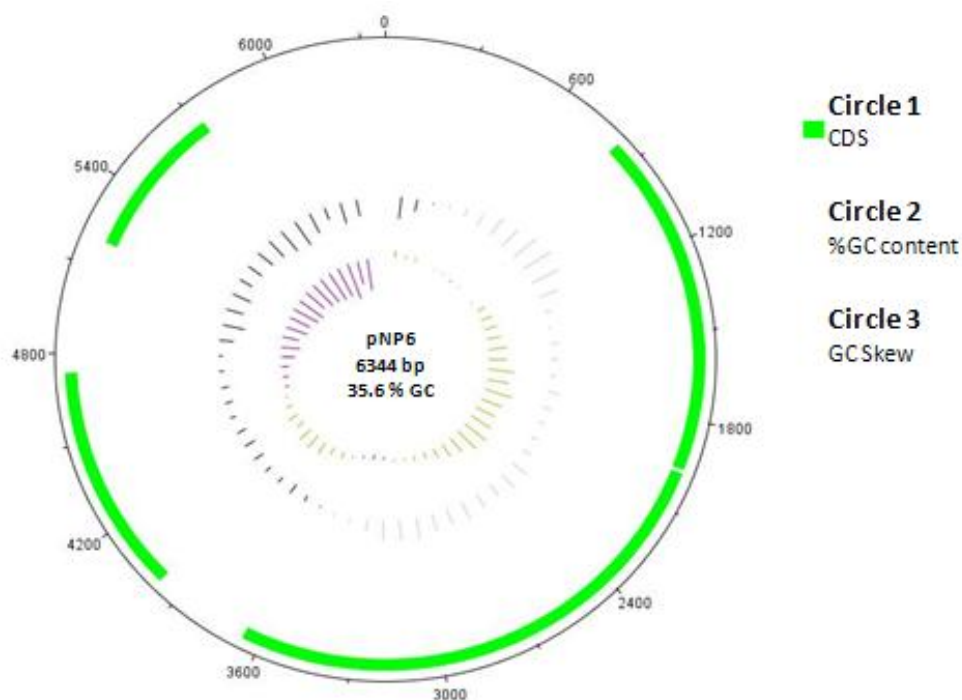


Figure 3.15. Circular plot of the MB2003 pNP6 plasmid. The green colour coding of the genomic features in circle 1 represent Coding Sequence (CDS), %G+C content is shown in circle 2 and GC skew in circle 3.

Table 3.14. Sequence information of the MB2003 pNP6 plasmid.

Total coding (bps)	4135
Total (bps)	6344
Coding %	65
Total number of genes	5

3.7.3 MB2003 Megaplasmid pNP144

The pNP144 megaplasmid is the second largest replicon at 143,881 bp and has a %G+C content of 37% (Figure 3.16 and Table 3.15). The total coding percentage for the megaplasmid is 80% with 154 predicted genes in total. Approximately 120 pNP144 genes encode hypothetical proteins and have no assigned function. None of the pNP144 genes are predicted to be involved in lignocellulose degradation. A full list of genes from the megaplasmid is shown in Appendix I (Table A3).

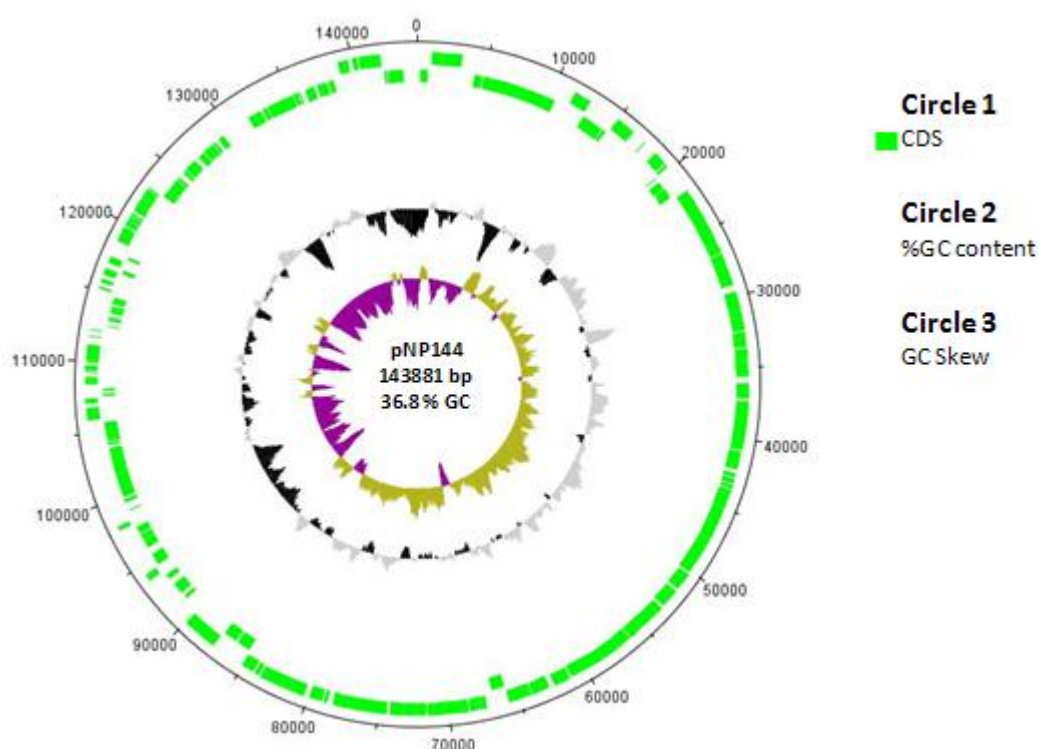


Figure 3.16. Circular plot of the MB2003 pNP144 megaplasmid. The green colour coding of the genomic features in circle 1 represent Coding Sequence (CDS), %G+C content is shown in circle 2 and GC skew in circle 3.

Table 3.15. Sequence information of the MB2003 pNP144 megaplasmid.

Total coding (bps)	116441
Total (bps)	144530
Coding %	80
Total number of genes	154

3.7.4 MB2003 Chromosome

The chromosome is the largest of the four replicons in the MB2003 genome with a provisional size of 3,137,803 bp and a %G+C content of 40%. The draft chromosome encodes a total of 2,735 genes and it is currently made up of 7 contigs arranged into 1 scaffold (Figure 3.17 and Table 3.16). All genes on the chromosome have been manually annotated and a full list of genes is shown in Appendix I (Table A4).

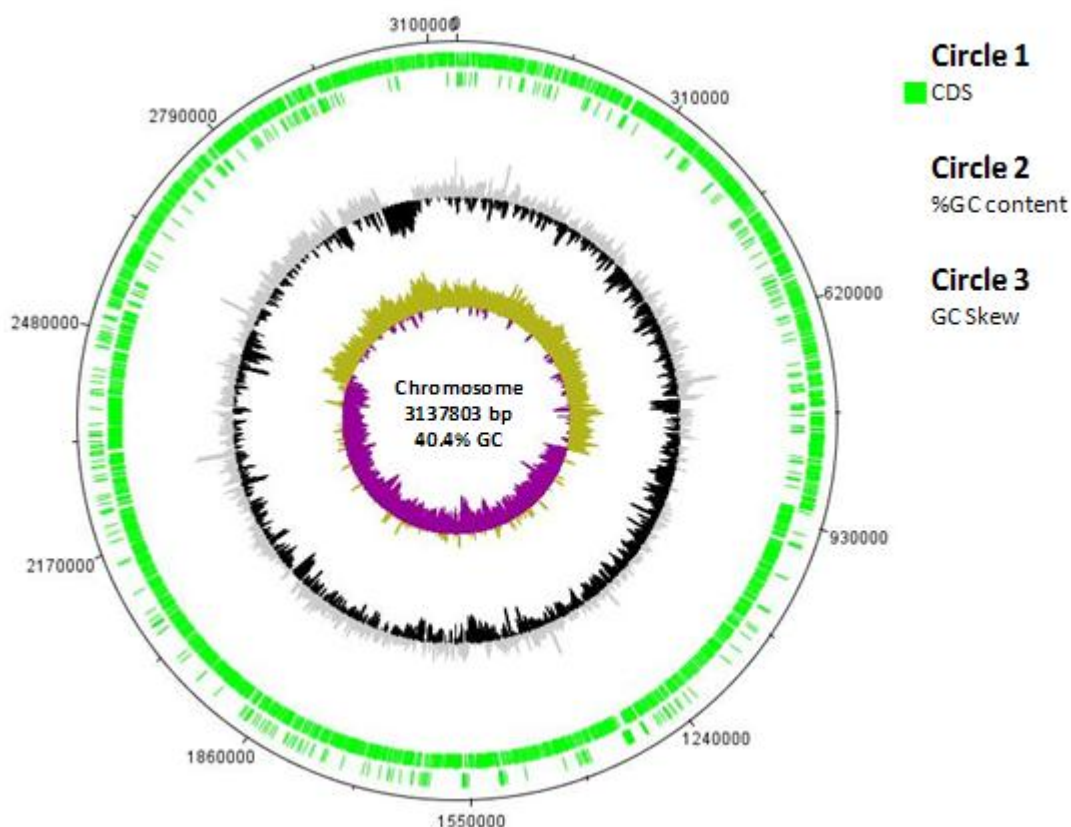


Figure 3.17. Circular plot of the draft MB2003 chromosome. The green colour coding of the genomic features in circle 1 represent Coding Sequence (CDS), %G+C content is shown in circle 2 and GC skew in circle 3.

Table 3.16. Sequence information of the MB2003 chromosome.

Total coding (bps)	2750638
Total (bps)	3137803
Coding %	88
Total number of genes	2735

3.8 Genes involved in polysaccharide breakdown

The main aim of sequencing the MB2003 genome was to identify genes involved in the breakdown of plant cell walls. Therefore the Carbohydrate-Active enZYmes (CAZy) database was used to identify glycoside hydrolases, glycosyl transferases, polysaccharide lyases, carbohydrate esterases and carbohydrate-binding protein families within the MB2003 glycobiome. The genes involved in polysaccharide breakdown are summarised in Table 3.17. A total of 90 genes were identified that encode either secreted or intracellular proteins that are involved in the breakdown of polysaccharides.

Table 3.17. MB2003 genes predicted to encode secreted or intracellular proteins involved in polysaccharide breakdown.

Proteins	Predicted Catalytic Domain*			
	GH	CE	PL	CBP
Predicted Secreted	16	2	-	1
Predicted Intracellular	60	10	1	-
Total	76	12	1	1

*GH, glycoside hydrolase, CE, carbohydrate esterase, PL, polysaccharide lyase and CBP, carbohydrate binding protein.

A comparison of the secreted and intracellular polysaccharide degrading genes of MB2003 with those from *B. proteoclasticus* B316 was made and is summarised in Table 3.18 and Table 3.19. Out of the 19 secreted genes in MB2003 involved in polysaccharide degradation, only two, lysozyme Lyc25B and feruloyl esterase Est1A, do not have a corresponding match with a gene in B316. These two secreted proteins are encoded by the MB2003 chromid (Bhu II).

The intracellular proteins involved in polysaccharide degradation predicted from the MB2003 genome were compared with homologues in the *B. proteoclasticus* B316 genome (Table 3.19). The majority (59) of MB2003 genes involved in polysaccharide breakdown excluding glycosyl transferases, had corresponding homologues in *B. proteoclasticus* B316. Three of the genes encoding intracellular proteins were found in the Bhu II chromid: a β -glucosidase (Bgl3A), a β -galactosidase (Bga42A) and a polysaccharide deacetylase (Est4A).

Table 3.18. Secreted polysaccharide degrading genes in MB2003 compared with their homologues in *B. proteoclasticus* B316.

MB2003 gene name	Size (aa)	Catalytic domains	Binding domains	Homologue in B316	% Match id/sim
β -N-acetylhexosaminidase Bhx3A	427	GH3		Bhx3B	48 /64
β -glucosidase Bgl3D	982	GH3	C-terminal TMH	Bgl3D	83 /89
endo-1,4- β -glucanase Cel5C	543	GH5	CBM2a	Cel5C	69 /81
endo-1,4- β -glucanase/xylanase Cel5A	417	GH5		Cel5A	55 /67
endo-1,4- β -xylanase Xyn10B	425	GH10		Xyn10E	68 /77
endo-1,4- β -xylanase Xyn10A	451	GH10		Xyn10A	80 /87
α -amylase Amy13B	536	GH13		Amy13B	58 /71
chitinase Chi18A	567	GH18		Chi18A	72 /82
lysozyme Lyc25A	362	GH25		Lyc25A	71 /84
lysozyme Lyc25B	515	GH25		Lyc25D	70 /83
lysozyme Lyc25C	561	GH25		Lyc25C	66 /79
lysozyme Lyc25D	242	GH25			
glycoside hydrolase family 30 Gh30A	575	GH30		Gh30A	69 /80
xylosidase/arabinofuranosidase Xsa43A	543	GH43	CBM6	Xsa43A	70 /79
arabinogalactan endo-1,4- β -galactosidase Agn53A	439	GH53		Agn53B	60 /74 *
4- α -glucanotransferase Mal77A	506	GH77		Mal77A	80 /88
feruloyl esterase Est1A	351	CE1		Est1C	69 /79
polysaccharide deacetylase Est4C	280	CE4		Est4C	60 /72
carbohydrate binding protein	983		CBM2a (x1), CBM6 (x5)		

*, not full length, % id, percentage identified amino acids, % sim, percentage similar amino acids with corresponding BLAST scores. Genes in **blue** are encoded in the chromid (Bhu II).

Table 3.19. Intracellular polysaccharide degrading genes in MB2003 compared with their homologues in *B. proteoclasticus* B316.

MB2003 gene name	Size (aa)	Catalytic domains	Binding domains	Homologue in B316	% Match id/sim
mannose 6-phosphate isomerase/glycoside hydrolase family 1 protein	765	GH1, PMI typeI			
β -galactosidase Bga2A	1034	GH2		Bga2A	76 /86
β -galactosidase Bga2C	714	GH2			
β -galactosidase Bga2B	825	GH2		Bga2B	70 /83
glycoside hydrolase family 2 Gh2B	641	GH2		Gh2B	77 /88
glycoside hydrolase family 2 Gh2A	912	GH2		Gh2F	48 /63
glycoside hydrolase family GH2C	776	GH2			
β -glucosidase Bgl3A	803	GH3		Bgl3C	57 /71
β -glucosidase Bgl3B	808	GH3		Bgl3E	83 /90
β -glucosidase Bgl3C	671	GH3			
β -xylosidase Xyl3A	707	GH3		Xyl3A	82 /87
reducing end xylose-releasing exo-oligoxylanase Xyn8A	383	GH8		Xyn8A	76 /83
cellodextrinase Cel9B	552	GH9	CelD	Cel9B	72 /83
1,4- α -glucan branching enzyme GlgB2	824	GH13	CBM48 (x2)	GlgB2	76 /87
1,4- α -glucan branching enzyme GlgB1	663	GH13	CBM48	GlgB1	88 /94
α -amylase Amy13C	434	GH13		Amy13C	88/94
α -amylase Amy13A	697	GH13		Amy13E	70 /82
α -amylase Amy13D	511	GH13		Amy13G	73 /83
glycogen debranching enzyme GlgX1	726	GH13	CBM48	GlgX1	85 /91
glycogen debranching enzyme GlgX2	648	GH13		GlgX2	63 /76
sucrose phosphorylase Suc13P	553	GH13		Suc13P	88 /92
lysozyme Lyc25E	1213	GH25	Big2 (x2)	Lyc25E	54 /66
α -galactosidase Aga27A	577	GH27		Aga27A	65 /77
glycoside hydrolase family 27 Gh27A	442	GH27		Gh27A	73 /85
polygalacturonase Pgl28A	531	GH28			
polygalacturonase Pgl28B	519	GH28		Pgl28B	76 /87
α -L-fucosidase Fuc29A	475	GH29		Fuc29A	49 /67
glycoside hydrolase family 31 Gh31A	756	GH31		Gh31A	58 /74
glycoside hydrolase family 31 Gh31C	674	GH31		Gh31C	81 /89
glycoside hydrolase family 31 Gh31B	635	GH31		Gh31D	83 /91
sucrose-6-phosphate hydrolase Scr32A	493	GH32		Scr32A	89 /94
β -galactosidase Bga35A	622	GH35		Bga35A	84 /90
β -galactosidase Bga35B	735	GH35		Bga35B	64 /76
α -galactosidase Aga36A	782	GH36		Aga36A	72 /82
α -galactosidase Aga36B	620	GH36		Aga36B	54 /71
α -galactosidase Aga36C	730	GH36		Aga36C	73 /85
α -mannosidase Man38A	1053	GH38		Man38A	55 /68
β -galactosidase Bga42A	673	GH42			
xylosidase/arabinofuranosidase Xsa43B	301	GH43		Xsa43I	45 /63
xylosidase/arabinofuranosidase Xsa43C	302	GH43			
xylosidase/arabinofuranosidase Xsa43D	517	GH43		Xsa43D	55 /69

xylosidase/arabinofuranosidase Xsa43E	352	GH43		
xylosidase/arabinofuranosidase Xsa43G	312	GH43	Xsa43G	85 /91
xylosidase/arabinofuranosidase and esterase Xsa43F	925	GH43, CE10	Xsa43H	72 /80
α -L-arabinofuranosidase Arf51C	630	GH51	CBM9	
α -L-arabinofuranosidase Arf51A	502	GH51	CBM_X	Arf51A 56 /72
α -L-arabinofuranosidase Arf51B	504	GH51		Arf51B 87 /92
α -D-glucuronidase Agu67A	662	GH67		Agu67A 70 /79
unsaturated glucuronyl hydrolase Ugl88A	385	GH88		Ugl88A 64 /76
unsaturated glucuronyl hydrolase Ugl88B	383	GH88		Ugl88B 43 /60
cellobiose phosphorylase Cbp94A	814	GH94		Cbp94A 89 /94
glycoside hydrolase family 95 Gh95A	734	GH95		Gh95A 43 /58
unsaturated rhamnogalacturonyl hydrolase Gh105A	349	GH105		Gh105A 93 /97
unsaturated rhamnogalacturonyl hydrolase Gh105B	363	GH105		Gh105C 77 /85
				lacto-N-biose phosphorylase
lacto-N-biose phosphorylase	722	GH112		78 /87
				se
α -glucuronidase Gh115A	947	GH115		Gh115A 72 /85
xylosidase Xyl120A	861	GH120		Gh120A 72 /81
xylosidase Xyl120B	664	GH120		Gh120B 57 /70
glycoside hydrolase	340	GH		NC
glycoside hydrolase	389	GH		NC
feruloyl esterase Est1B	243	CE1		Est1E 53 /71
acetyl-xylan esterase Est2A	372	CE2		Est2A 68 /79
polysaccharide deacetylase Est4A	207	CE4		
polysaccharide deacetylase Est4D	274	CE4		Est4D 70 /86
polysaccharide deacetylase Est4B	244	CE4		Est4E 57 /77
N-acetylglucosamine-6-phosphate deacetylase NagA	371	CE9		NagA 50 /68
acetyl-xylan esterase Est10A	275	CE10		Est10A 66 /79
carbohydrate esterase family 12 Est12A	584	CE12		Est12A 52 /66
carbohydrate esterase family 12 Est12B	244	CE12		
polysaccharide lyase Pl11A	746	PL11		
glycogen phosphorylase GlgP1	769	GT35		GlgP 92 /96
glycogen phosphorylase GlgP2	824	GT35		GlgP2 87 /93
acetyl-xylan esterase	626			acetyl-xylan esterase 58 /73

*, not full length, % id, percentage identified amino acids, % sim, percentage similar amino acids with corresponding BLAST scores, NC, not classified. Genes in blue are encoded in the chromid (Bhu II).

3.9 Xylan degradation in MB2003

MB2003 is able to grow on xylan and the genome encodes numerous enzymes predicted to be involved in this process. These include endo-1, 4- β -xylanases (GH5 and GH10), xylosidases (GH3, GH8, GH30, GH43 and GH120), arabinofuranosidases (GH43 and GH51), α -glucuronidases (GH67 and GH115), feruloyl esterases (CE1) and acetyl xylan esterases (CE2, CE4 and CE10) (Figure 3.18). As yet, no functional data is available for any of these enzymes, however comparisons can be made with the *B. proteoclasticus* B316 genome. Microarray analysis of *B. proteoclasticus* B316 (Kong, 2007) showed that two clusters of genes were significantly upregulated in cells grown on wheat arabinoxylan. These gene clusters were designated polysaccharide utilisation loci (PUL). Figure 3.19 shows that a very similar gene arrangement occurs in MB2003.

However, the MB2003 genes form a single PUL, containing genes predicted to have xylanase, xylosidase, xylosidase/arabinofuranosidase, α -glucuronidase and acetyl-xylan esterase activity. In *B. proteoclasticus* B316 a gene encoding a secreted feruloyl esterase is also present (Est1C). The equivalent gene is missing from this location in MB2003 but a close match (Est1A) is found on the chromid (Bhu II). Est1A is the only feruloyl esterase predicted to be secreted by MB2003. Also located in the MB2003 PUL is a gene encoding xylose isomerase (ORF530) which catalyses the first step in intracellular xylose metabolism via the pentose phosphate pathway.

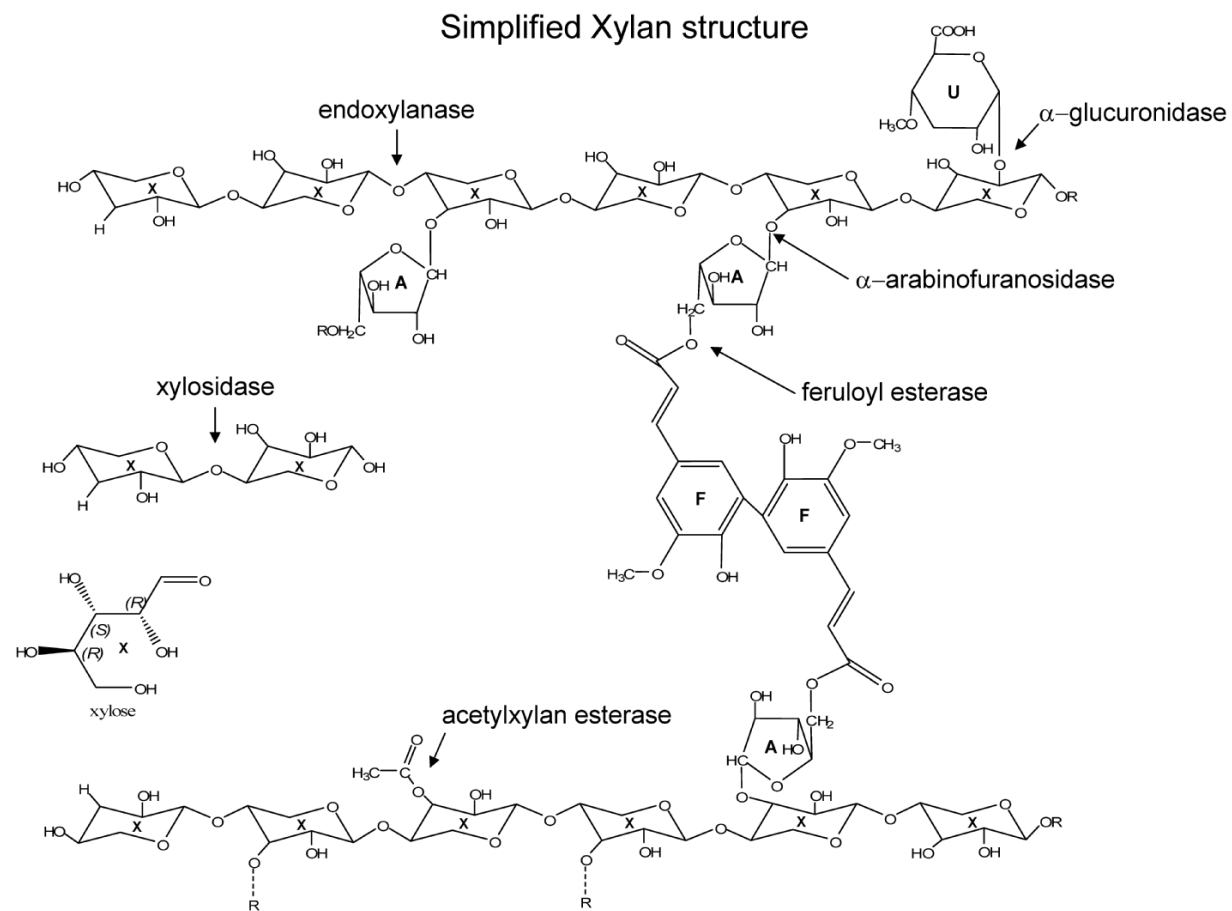
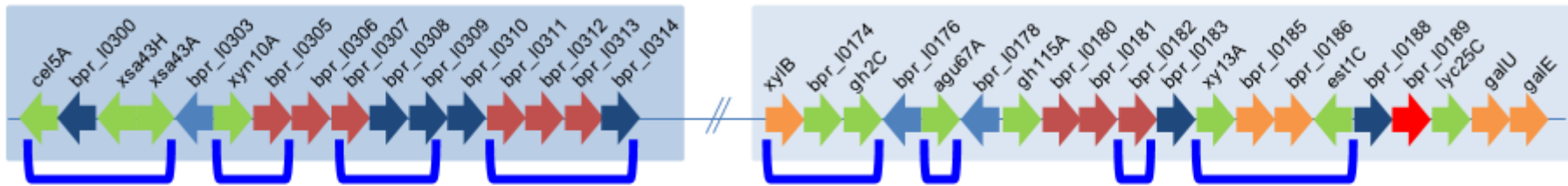


Figure 3.18. Diagrammatic representation of xylan, showing linkages attacked by hemicellulose-degrading enzymes. A, arabinose; F, ferulic acid; U, 4-O-methylglucuronic acid; X, xylose; R, continuation of glycan chain. Modified image from Chesson and Forsberg (1997).

Butyrivibrio proteoclasticus B316



Butyrivibrio sp. MB2003

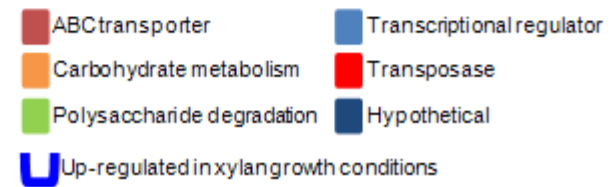


Figure 3.19. Comparison of PUL in *B. hungatei* MB2003 and *B. proteoclasticus* B316.

3.10 Motility

Upon completion of annotation and analysis of the MB2003 genome, several genes were uncovered that were thought to encode flagella. To investigate the possibility of the presence of functional flagella in MB2003, a simple motility assay was carried out. Soft agar medium was stabbed with cultures and observed after overnight inoculation at 37°C. The motility assay revealed that *B. hungatei* MB2003 and *B. proteoclasticus* B316 (described previously as non-motile (Attwood et al., 1996)) were non-motile whereas the *P. xylanovorans* MA3014 displayed a motile phenotype (Figure 3.20 A).

The performance of a bacterium in a motility assay does not however imply the absence of a flagellum, as it is known that *B. proteoclasticus* B316 contains a flagellum but is non-motile. Thus an analysis of the region encoding the flagella genes and the order of these genes for all three cultures was carried out. The chromosomal region starting at *flgE* and ending with *motB* genes revealed that the size of the *flgE* gene is the same for MB2003 and B316 but much smaller in MA3014, and that two extra genes were present in MB2003 (Ion channel and *flbD*) and B316 (*est1B* and *flbD*) but were absent in the MA3014 genome (Figure 3.20 B). The *flbD* genes are the same size in both MB2003 and B316 and the *est1B* and Ion channel genes are also approximately the same size in their respective genomes. The *motA* and *motB* genes are present in all three chromosomes and are approximately the same in size and do not appear to differ from one another. The MA3014 flagellar biosynthetic loci appear uninterrupted whereas in *B. hungatei* MB2003 and *B. proteoclasticus* B316 they appear to have insertions in flagella genes that might result in impaired motility.

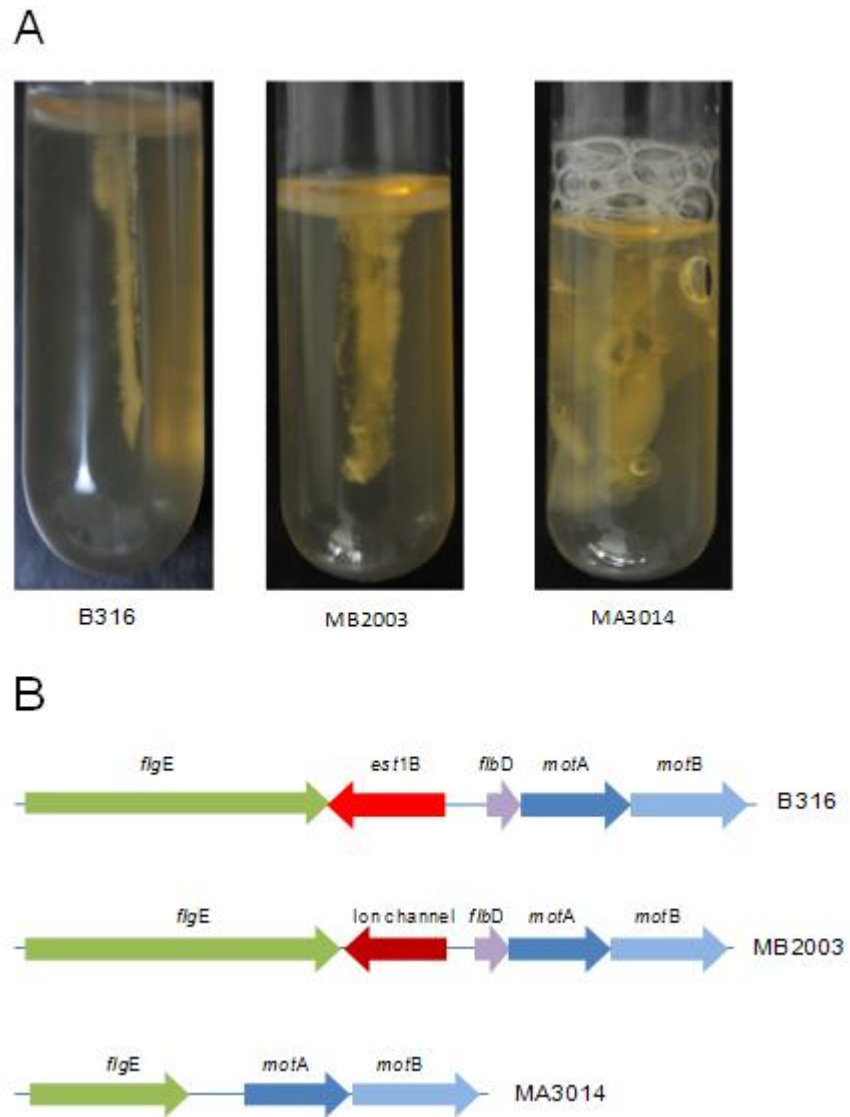


Figure 3.20. Motility assays and flagella biosynthesis operons in MB2003, MA3014 and B316. A, Motility assay; B, Genes encoded in the flagella biosynthesis operon.

3.11 Transmission Electron Microscopy (TEM)

To look for flagella, TEM was carried out on *B. hungatei* MB2003, *B. proteoclasticus* B316 and *P. xylanivorans* MA3014. All cultures were grown in RM02 and BY media overnight at 37°C with the GenRFV supplement, and electron microscopy grids were prepared as outlined in the Methods Section 2.2.10. Images in Figure 3.21 A and B show MB2003 cells in RM02 medium. In Figure 3.21 A there appears to be a long flagellum protruding from the middle of the bottom cell that stretches out and past the upper cell. This particular image is one of few found to contain such a structure; the majority of TEM images of MB2003 cells did not contain any apparent surface structures (Figure 3.21 B). There were no flagella detected on any of the MA3014 cells observed. Most MA3014 cells appeared shriveled and dehydrated (Figure 3.21 C). An interesting observation was the presence of unusual structures on the surface of some MA3014 cells (Figure 3.21 D). From a previous study, B316 have been shown to possess a single flagellum but they are non-motile. When grown in RM02 the B316 bacteria did not contain a single flagellum and appeared shriveled and dehydrated. However the RM02 media does not contain many of the substrates that encourage optimal growth for ruminant bacteria and thus the same TEM protocol was carried out on the three cultures grown in BY media.

The TEM images for all three cultures grown in BY media produced the same results as in the RM02 medium (Figure 3.22). The morphology of all three different bacteria was identical to those in RM02 and the MA3014 displayed inclusion bodies on the surface of the cells (Figure 3.22 D). There were no signs of flagella on any cell in any of the three cultures. The background staining on the TEM images was due to the negative stain binding to all of the molecules in the substrate rich BY media. All of the samples prepared for TEM imaging in both RM02 and BY were not centrifuged or had any kind of modification done to improve the quality of the image. Thus more work needs to be done with TEM imaging to completely rule out the possibility of visualising flagella in these particular cultures.

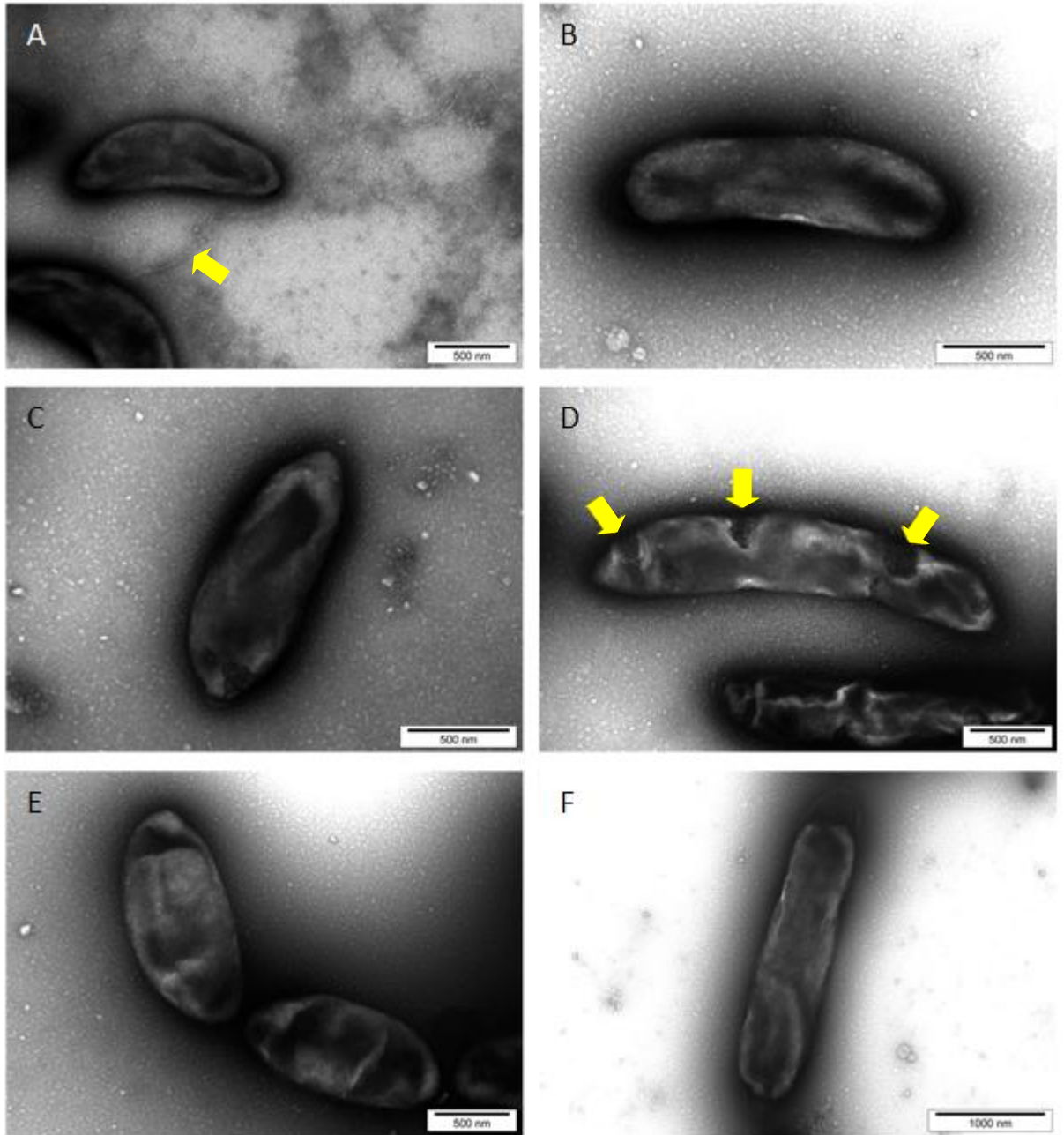


Figure 3.21. Transmission Electron Microscopy images of MB2003, MA3014 and B316 cells from cultures grown on RM02 media after overnight incubation. A and B, MB2003; C and D, MA3014; E and F, B316.

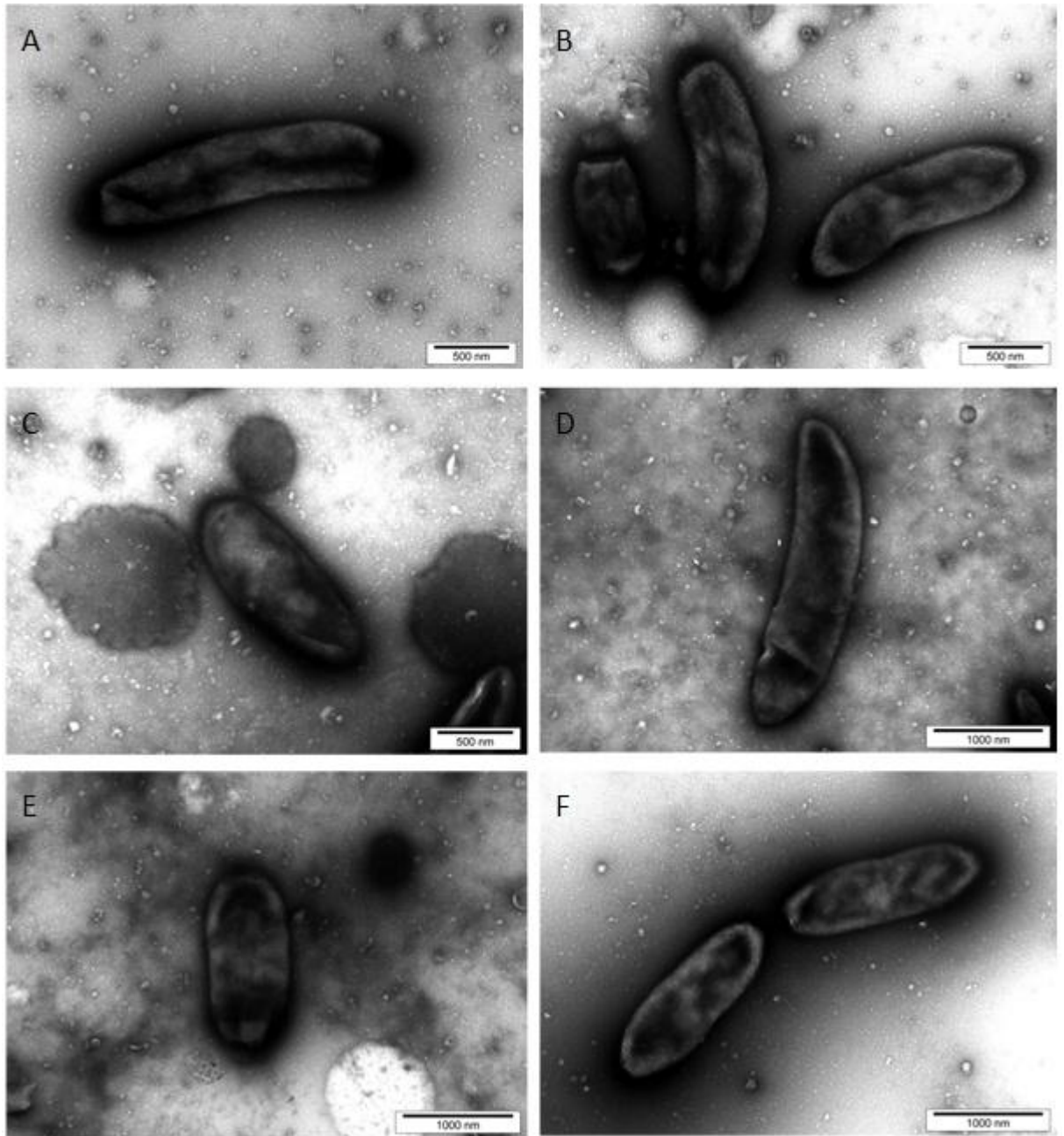


Figure 3.22. Transmission Electron Microscopy images of MB2003, MA3014 and B316 planktonic cells from cultures grown on BY media after overnight incubation. A and B, MB2003; C and D, MA3014; E and F, B316.

Chapter 4

Discussion

The work described in this thesis forms part of larger ARC and NERF programmes focused on understanding the role of plant adherent bacteria in the breakdown of plant cell wall polysaccharides in the rumen of pasture grazed cattle in New Zealand. These programs include novel cultivation approaches to isolate plant adherent bacteria coupled with metagenomic analyses to assess bacterial diversity, and genomic analysis of individual bacteria which are described here. The aim of this MSc was to select representative plant-adherent bacterial cultures based an analysis of metagenomic 16S rRNA sequence data, to determine their phenotypic and general genomic characteristics and to genome sequence isolates in order to learn more about their lignocellulose degradation systems.

The bacteria that adhere to plant material are thought to be the main fibrolytic bacteria in the rumen (Weimer *et al.*, 1993), with the best studied organisms being *Fibrobacter succinogenes*, *Prevotella* sp., *Ruminococcus flavefaciens*, *Ruminococcus albus* and *Butyrivibrio* sp. (Tajima *et al.*, 1999; Russell and Rychlik, 2001; Koike *et al.*, 2003b). However, the rumen harbours a dense bacterial population and the diversity of organisms involved in plant polysaccharide breakdown in the rumen is only now beginning to be revealed. A recent meta-analysis of all the curated 16S rRNA gene (*rrn*) sequences deposited in the RDP database (Kim *et al.*, 2011), has provided insight into the composition of the rumen bacterial microbiome. A total of 13,478 bacterial sequences of rumen origin were analysed with nineteen bacterial phyla represented. Firmicutes, Bacteroidetes and Proteobacteria were the predominant phyla accounting for 57.8%, 26.7% and 6.9% of the total bacterial sequences, respectively. Approximately 90.6% of the Firmicutes sequences were assigned to class Clostridia (Whitford *et al.*, 1998; Tajima *et al.*, 1999; Tajima *et al.*, 2000), with the largest families (Lachnospiraceae and Ruminococcaceae) accounting for 23.8% and 25.8% of the sequences. The predominant genera were *Butyrivibrio*, *Acetovibrio*, *Ruminococcus*, *Succinivibrio* and *Pseudobutyrvibrio*. An important finding from this study was that a large number of the Firmicutes sequences have not been classified to any existing family, order or genus within the Clostridia class.

Data from the metagenomic analysis of 16S rRNA gene sequences performed in the ARC project showed similar results. The most prevalent organisms in the plant adherent fraction were from the family Lachnospiraceae within the Firmicutes phylum, with

Butyrivibrio, and *Pseudobutyrvibrio* the most common genera. Consequently, these bacteria are predicted to be predominant in the rumen and the isolation and classification of novel strains belonging to these groups will greatly improve our knowledge of rumen function and our understanding of plant polysaccharide breakdown within the rumen.

A total of 141 bacterial isolates cultivated through NERF programme (Noel. S *et al.*, Unpublished) had their 16S rRNA genes sequenced and this information was combined with the sequences from the ARC metagenomic project into a single phylogenetic tree. 16S rRNA gene sequences of isolates corresponding to large clusters of gene sequences retrieved directly from rumen contents were selected as cultivated representatives of those clusters. In all, twenty isolates representing 5 major clusters were selected as candidates for genome sequencing. Genomic DNA was extracted from each isolate and used to produce full length 16S rRNA sequences to confirm their original classification. PFGE was performed to determine if there were differences within the cluster representatives at the genome level and to establish whether they contained extra chromosomal elements. Phenotypic characterisation experiments examined carbon source utilisation (both soluble and insoluble substrates), VFA production, and made observations of bacterial adherence to plant material using DAPI staining and SEM. These characterisations led to the description of *B. hungatei* MB2003 and *P. xylanivorans* MA3014 strains and were selected for whole-genome sequencing. Because of time constraints only the MB2003 genome analysis is reported here.

4.1 MB2003 general genome features

The *B. hungatei* MB2003 genome sequence was found to consist of four replicons: a chromosome (currently in 7 contigs arranged into one super scaffold), a chromid (Bhu II), a megaplasmid (pNP144) and a small plasmid (pNP6; Table 3.12). This is very similar genome architecture to the multi-replicon genome of *B. proteoclasticus* B316 (Kelly *et al.*, 2010). For many MB2003 genes, the closest homologues were to B316 genes. This is not surprising as B316 is the only other *Butyrivibrio* species genome sequenced. A MUMmer plot comparing the main chromosomes of MB2003 with B316 produced an X-shaped alignment. The most likely explanation for such an X-alignment is inversions of the chromosomes that cause the reversal of genomic sequence symmetry around the origin/terminus of DNA replication.

The main difference between MB2003 and B316 is in the size of the four replicons. The *B. proteoclasticus* B316 genome contains two megaplasmids of 360 Kb and 186 Kb compared to MB2003 which contains only one megaplasmid of 144 Kb. Megaplasmids in *Butyrivibrio* species are predicted to encode genes important for survival in the rumen, but no genes have been identified that contribute to the breakdown of lignocellulose (Yeoman, 2009).

The small MB2003 plasmid, pNP6 is unusual as extrachromosomal DNAs in *Butyrivibrio* are usually large. The pNP6 plasmid appears to encode only genes involved in its own replication and transfer and has no functional role in the degradation of polysaccharides. Its small size opens the possibility of development as a potential vector for *Butyrivibrio* species. First attempts at producing shuttle vectors for *Butyrivibrio* species (Ware *et al.*, 1992; Whitehead, 1992), produced poorly selectable plasmids that had problems with poor transformation efficiency and plasmid stability. Beard *et al.*, (1995) reported the construction of a stable 9.5 Kb shuttle vector, pBHerm, based on the *B. fibrisolvens* plasmid pRJF1 capable of successful transfer, replication and selection in both *E. coli* and *B. fibrisolvens*. The pBHerm system was demonstrated to be a highly stable transformation system and a practical tool for genetic alterations to *Butyrivibrio* species. A study by Gregg *et al.*, (1998) used the pBHerm shuttle vector augmented by the addition of a fluoroacetate dehalogenase gene attached to a copy of the erythromycin-resistance gene promoter to produce a *Butyrivibrio* strain capable of detoxifying fluoroacetate, and demonstrated this with *in vivo* experiments in sheep.

A novel feature of both the MB2003 and *B. proteoclasticus* B316 genomes is the presence of chromids or secondary chromosomes. Chromids are replicons that have %G+C content similar to that of their main chromosome, but have plasmid type maintenance and replication systems, and are smaller than the chromosome but usually larger than any other plasmids present. Chromids contain genes essential for growth and maintenance of the organism along with several core genus-specific genes that can be found on the chromosome in other species of bacteria (Harrison *et al.*, 2010). The Bhu II replicon has most of these characteristics and therefore has been designated as a chromid of MB2003. Since the Bhu II chromid of *B. hungatei* MB2003 is smaller than the BPc2 chromid of B316, it is now the smallest chromid reported for bacteria. The chromid description was developed for replicons found in a wide range of Gram

negative bacteria. The replicons found in *Butyrivibrio* species are smaller in size and warrant further investigation to determine if similar extrachromosomal elements occur in other Gram positive bacteria.

4.2 Polysaccharide Degradation

The ability to adhere to plant material is an essential characteristic for plant polysaccharide degrading bacteria within the rumen. Using light microscopy and SEM, MB2003 was observed to adhere to plant material (Figures 3.7 and 3.8). The mechanisms involved in rumen bacterial adherence to plant material are not fully understood. MB2003 contains genes for flagellum biosynthesis, exopolysaccharide (EPS) production and for a single carbohydrate-binding protein containing, CBM2 and CBM6 domains. However, only two of the polysaccharide degrading proteins, that are predicted to be secreted, have CBM domains. MB2003 has numerous genes predicted to encode proteins involved in production of exopolysaccharides. MB2003 produces large amounts of ropey filaments that could potentially be exopolysaccharide covering the cells, which can be seen under SEM observation of MB2003 and even more so in MA3014 grown in BY medium on NDF plant material (Figures 3.8 C, D and 3.9C, D). B316 has several potential mechanisms for adhesion to plant cell walls including a large number of CBM-containing proteins, and genes encoding for exopolysaccharide production, a single flagellum, pili and several cell surface proteins with a likely role in adhesion.

Both MB2003 and B316 are non-motile, while *P. xylanivorans* MA3014 displayed motility as shown using a motility assay (Figure 3.20 A and Attwood *et al.*, (1996)). Analysis of the genes in the flagella biosynthesis operon revealed the absence of two gene insertions in MA3014 (Figure 3.20 B). Flagellar gene organization is almost identical for both MB2003 and B316. The cluster of genes that encode for the flagellum structure and are responsible for motility functions are disrupted in both bacteria. It is hypothesised that these insertions impair motility in both MB2003 and B316. The TEM images from this study could not produce clear evidence of flagella for either MB2003 or B316 on NDF plant material (Figure 21 A, B, E and F and Figure 22 A, B, E and F). However it is possible that both bacteria use their flagellum for adherence rather than motility (Shimoyama *et al.*, 2009). From previous work it is known that B316 possesses

an flagella but is non-motile (Attwood *et al.*, 1996). MB2003 may be the same, however further investigation is required to resolve the flagellum issue.

Approximately 3% of the MB2003 genome (90 CDSs) is predicted to encode proteins dedicated to polysaccharide degradation (Table 3.18 and 3.19), similar to that found in B316. In total, MB2003 has 76 glycoside hydrolase and 86 glycosyl transferase genes (Table A4 in Appendix I), whereas *B. proteoclasticus* B316 has 113 GHs and 121 GTs (Kelly *et al.*, 2010). The numbers of GHs and GTs in MB2003 are comparable to B316 in terms of the number of genes relative to genome size.

Xylan is the preferred insoluble substrate that MB2003 utilises for growth and the genome encodes 29 enzymes predicted to metabolise this complex polysaccharide (Figure 4.1). Among these 7 are secreted: Cel5A (GH5), Xyn10A/B (GH10), Gh30A (GH30), Xsa43A (GH43), Est1A (CE1) and Est4C (CE4) (Table 3.18). All have significant matches to B316. The feruloyl esterase Est1A is the only gene encoding xylan degradation found on the chromid (Bhu II) of MB2003. The intracellular genes encoding proteins involved in xylan degradation and their catalytic domains are: Est10A (CE10), Xyl120A/B (GH120), Est1B (CE1), Xsa43B-G (GH43), Xyn8A (GH8), Xyl3A (GH3), Agu67A (GH67), Arf51A-C (GH51), Gh115A (GH115), Est2A (CE2), Est4A, B, D, and an unclassified acetyl-xylan esterase (Table 3.19). Of these all have significant matches to B316 except Xsa43C, Xsa43E, Arf51C and Est4A. Furthermore observation of the clusters of genes encoding xylan degradation in B316 and MB2003 (Figure 3.19) show similar arrangements. The difference is that the genes in MB2003 are found in a single cluster, whereas in B316 they form two separate clusters. In B316 genes from both clusters have been shown by microarray analysis to be strongly up-regulated in xylan grown cells (Kong, 2007).

Both MB2003 and B316 are unable to degrade cellulose but are able to utilise a range of other insoluble plant polysaccharides. The foundation of the B316 catalytic capability for the breakdown of pectin, starch and xylan is predicted to be centered around nine large cell-associated proteins (Amy13A, Agn53A, Est12B, Lic16A, Pel1A, Pme8B, Xsa43J, Xyn10B, and Bpr_I0264, (Kelly *et al.*, 2010)). These proteins contain multiple cell wall binding repeat domains at their C-termini that are predicted to anchor them to the peptidoglycan cell wall. MB2003 does not contain any genes encoding cell wall binding repeats and is thus markedly different from B316. As mentioned above

MB2003 prefers to utilise xylan as the primary growth substrate whereas the major sources of pyruvate for central metabolism in B316 are derived from breakdown products of both xylan and pectin, where pectin is the preferred growth substrate. B316 is thus a more versatile degrader of plant cell wall polysaccharides than MB2003 as it is able to utilise a wide range of insoluble polysaccharides for growth and metabolism.

When comparing the predicted number of secreted enzymes involved in polysaccharide degradation in Figures 4.2 and 4.3, there are many differences between the 44 enzymes of B316 and the 19 of MB2003. The overall size of the secreted enzymes in B316 (average 880 aa) are much larger than enzymes secreted by MB2003 (average 510 aa). B316 has nine enzymes larger than 1,000 aa and three larger than 2,000 aa whereas MB2003 has no secreted enzyme larger than 1,000 aa in size.

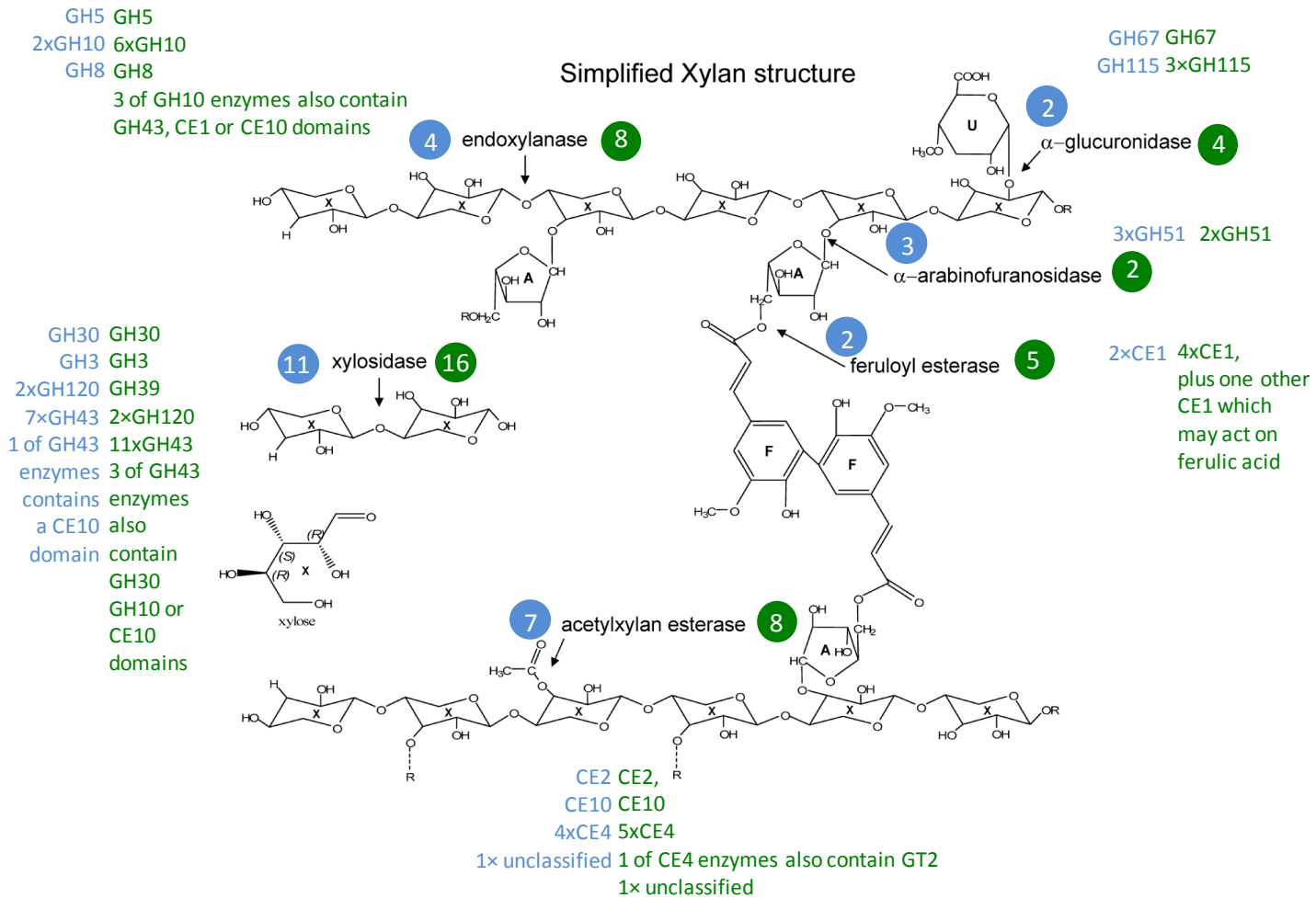


Figure 4.1. Diagrammatic representation of xylan, showing linkages attacked by hemicellulose-degrading enzymes in *B. hungatei* MB2003 (n = blue) and *B. proteoclasticus* B316 (n = green). A, arabinose; F, ferulic acid; U, 4-O-methylglucuronic acid; X, xylose; R, continuation of glycan chain. Modified image from Chesson and Forsberg (1997).

MB2003



Figure 4.2. Predicted domain architectures (PFAM) of secreted enzymes encoded by genes involved in polysaccharide breakdown in *B. hungatei* MB2003.

B316

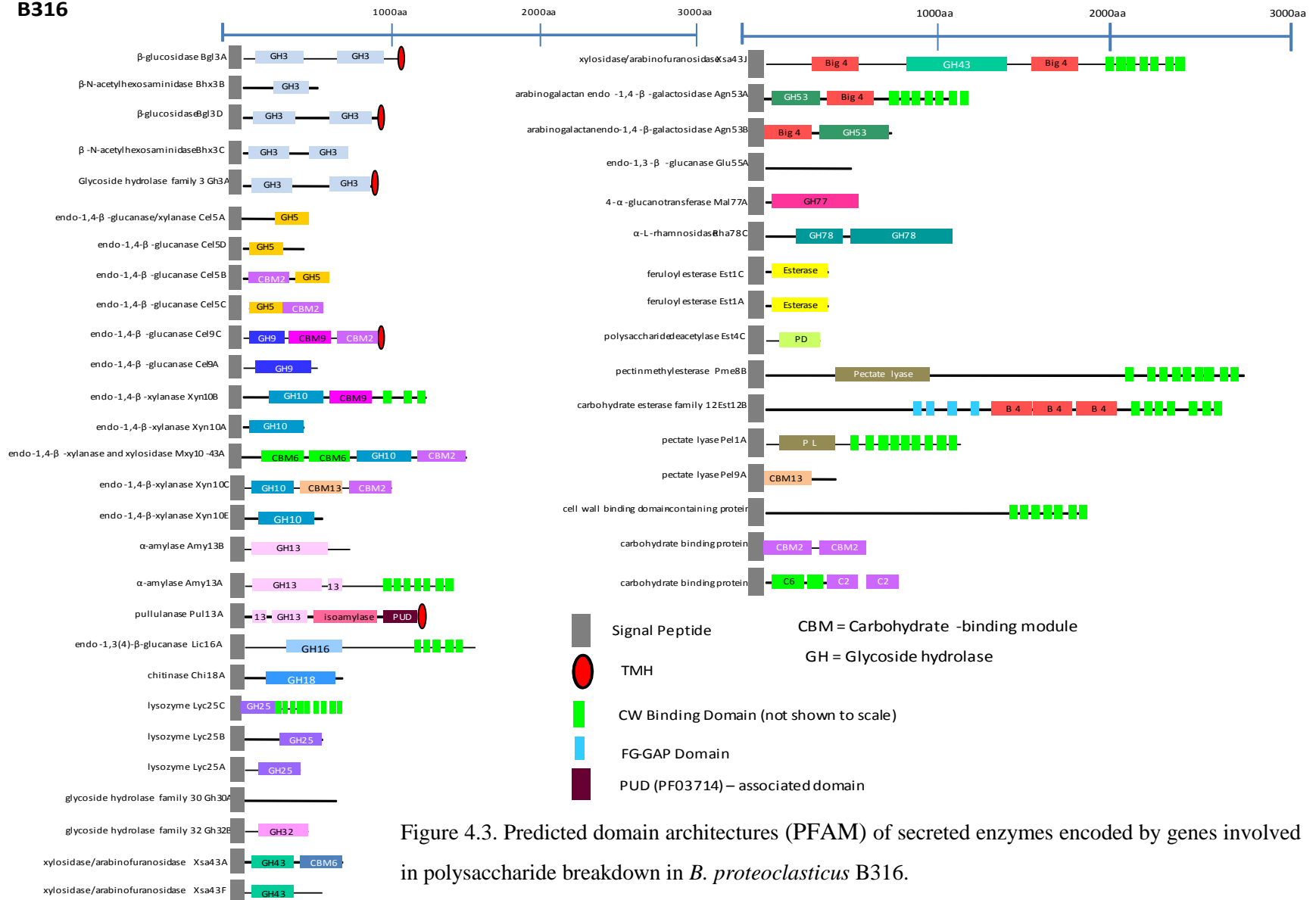


Figure 4.3. Predicted domain architectures (PFAM) of secreted enzymes encoded by genes involved in polysaccharide breakdown in *B. proteoclasticus* B316.

4.3 Comparison of polysaccharide-degrading strategies in ruminal bacteria

Genome sequence information is now available for *B. hungatei* MB2003, *B. proteoclasticus* B316 and for two well-studied cellulolytic rumen bacteria, *Fibrobacter succinogenes* S85 (Suen *et al.*, 2011) and *Ruminococcus flavefaciens* FD-1 (Miller *et al.*, 2009). These strains have a large repertoire of genes predicted to encode polysaccharide-degrading enzymes. In the current opinion these species are thought to work in synergy in the rumen to degrade plant biomass while using different approaches to breakdown polysaccharides (Flint *et al.*, 2008).

***Butyrivibrio* sp.**

B. hungatei MB2003 is capable of degrading and utilising hemicellulose as the main substrate, but not cellulose. Similarly *B. proteoclasticus* B316 is unable to degrade cellulose, but can degrade and utilise hemicellulose, and pectin (Kelly *et al.* 2010). For both, MB2003 and B316, enzymes that initiate polysaccharide breakdown are secreted, but the majority of the enzymes involved in carbohydrate metabolism are predicted to be intracellular. This suggests a model of plant cell wall breakdown in which secreted enzymes generate a variety of complex oligosaccharides that are transported into the cell for further metabolism. In B316, the clustering of genes encoding intracellular polysaccharide degrading enzymes with genes for transporters, transcriptional regulators and environmental sensors in several polysaccharide utilisation loci (Kelly *et al.*, 2010) supports this model. Clustering of polysaccharide-degrading enzymes around transporters could be a possible mechanism to reduce competition by limiting the amount of readily useable polysaccharide products in the surrounding environment that other microbes can utilise. Both B316 and MB2003 are able to adhere to plant material which together with their polysaccharide-degrading capabilities implicates that they are typical of plant colonisers in the rumen environment. However, given that B316 has more extensive polysaccharide-degrading capabilities than MB2003 it is possible that B316 is an initial coloniser and degrader of plant material in the rumen, whereas MB2003 may best be described as a secondary degrader.

Fibrobacter succinogenes

F. succinogenes S85 is a cellulose-degrading specialist in the rumen. Although it has the ability to produce a variety of enzymes capable of hydrolyzing plant cell wall polysaccharides, it uses these enzymes solely to gain access to cellulose. The genome lacks the genes necessary to transport and metabolise non-cellulose polysaccharides, resulting in utilisation of only cellulose as an energy source. The carbohydrates not used by *F. succinogenes* should be available to support the growth of a variety of other bacteria. Three unique models have been proposed for cellulose degradation by *F. succinogenes*. One model describes the growth of *F. succinogenes* as a biofilm on the cellulose surface, similar to a case reported for ruminococci (Costerton, 1992). Fibroslime proteins (Gong *et al.*, 1996; Jun *et al.*, 2007a) and Type IV pilin structures (Jun *et al.*, 2007b) are hypothesised to initiate cellulose breakdown by facilitating cell-surface attachment to the substrate and mediating GH and CBM containing enzyme interactions with the polysaccharides. A phylogenetically-related bacterium, *Cytophaga hutchinsonii*, is known to strongly associate with cellulose and has been proposed to degrade cellulose by disrupting cellulose fibers and absorbing through the outer cell membrane (Wilson, 2009). It is predicted to do this by cleaving the individual cellulose chains using various endoglucanases, thus enabling direct access to the hydrolytic products of cellulose (e.g. glucose, cellobiose, and cellodextrins). Lastly, a model similar to that proposed for the α -proteobacterium *Sphingomonas* sp. A1 (Hashimoto *et al.*, 2005a; Hashimoto *et al.*, 2005b) may apply, where the organism forms “pits” over its cell membrane acting as channels that can import the products of cellulose via high-affinity periplasmic proteins coupled with ABC transporters. In this model the degradation of the glucose, cellobiose, and cellodextrins would occur in the cytoplasm.

Ruminococcus flavefaciens

R. flavefaciens FD-1 can utilise a range of plant cell wall polysaccharides and produces a multi-enzyme cellulosome complex in which enzymes are linked to a non-catalytic scaffold structure via dockerin domains. FD-1 secretes at least 75% of its GHs and most of these are predicted to be cellulosome-associated, implying that most polysaccharide breakdown occurs extracellularly. The cellulosome organization in FD-1 is extremely complex and more than 200 dockerin-containing proteins have been identified from the draft genome sequence (Miller *et al.*, 2009).

4.4 Conclusion

The work described in this thesis was the first opportunity to conduct comparative genome analysis of two rumen bacteria belonging to the same genus and cultivated directly from the rumen microbial fraction of New Zealand cattle. The preliminary genotypic and phenotypic analysis of cultured representatives of large clusters of rumen organisms, confirmed *Butyrivibrio* and *Pseudobutyrvibrio* as important rumen bacteria under pasture grazed conditions. These characterisations led to the selection of *Butyrivibrio hungatei* MB2003, a strain that adheres to plant fibre and efficiently degrades xylan as the prime candidate for genome sequencing.

The finding that *Butyrivibrio hungatei* MB2003 has a multi-replicon genome, similar to that found in *B. proteoclasticus* B316 and the observation of other extrachromosomal DNAs in PFGE analyses, indicates that this type of genome arrangement is common among rumen *Butyrivibrio* species. Most of the genes encoding core metabolic processes in both these species are encoded on the chromosome or chromid, but the megaplastids and plasmid are largely cryptic, which leaves a question mark as to the role of these extrachromosomal elements in *Butyrivibrio*, especially the large megaplastids which encode a significant proportion of the genome.

The MB2003 genome encodes a large repertoire of enzymes predicted to metabolise hemicellulose which is consistent with its ability to grow well on xylan. Although both MB2003 and B316 have similar phenotypic characteristics and occupy the same habitat, the genome of MB2003 is significantly smaller and is predicted to contain much fewer extracellular polysaccharide degrading enzymes.

This comparison offers an alternate view of the genes required for a xylan/hemicellulose-degrading lifestyle in the rumen and poses an interesting question about the purpose of the greater range of polysaccharide degrading enzymes found in B316. It can be concluded that MB2003 is likely to be a secondary hemicellulose degrader in the rumen, relying on organisms such as B316 for enzymatic cleavage of polysaccharide linkages that it does not encode itself.

4.5 Future work

There are several areas identified in this thesis where future work is justified. The remaining gaps in the *B. hungatei* MB2003 genome will be closed and the final sequence will be reanalysed via GAMOLA to confirm the gene annotations and will be prepared for NCBI submission. Similarly, the draft *P. xylanivorans* MA3014 genome will be closed and fully annotated, to allow an in depth investigation into the polysaccharide degradation machinery of this bacterium. Genome comparisons of *Pseudobutyrvibrio* MA3014 will then be possible with MB2003, B316 and the human strain *Pseudobutyrvibrio* (*B. fibrisolvens*) 16/4 genomes which will hopefully offer new insights into the polysaccharide-degrading activities of this important group of rumen bacteria.

The adherence of MB2003 and MA3014 to plant substrates warrants further investigation along with the further attempts to visualise flagella. A better understanding of the metabolic capacity of *Butyrvibrio* and *Pseudobutyrvibrio* species would certainly be assisted via a side-by-side comparison of their growth on a range of soluble substrates in parallel with whole genome gene and protein expression profiling, coupled with end-product and metabolite analyses. The development of shuttle vectors using the pNP6 plasmid would also be useful to allow gene manipulation studies to be conducted in MB2003 and related strains in the future.

Appendix I

Gene Lists

Table A1. Summary of manual annotation of the genes within the chromid (Bhu II) of MB2003.

Gene number	Start	Stop	Size of gene (bp)	Gene Annotation
1	71	655	584	transcriptional regulator TetR family
4	820	2307	1487	radical SAM domain-containing protein
5	2397	2582	185	>hypothetical protein<
7	2745	4958	2213	GGDEF/EAL domain-containing protein
9	5086	6171	1085	tyrosine recombinase XerC
10	6598	6798	200	hypothetical protein
11	6888	7478	590	hypothetical protein
12	7590	8072	482	HTH domain-containing protein
13	8222	8827	605	transcriptional regulator TetR family
14	8830	9909	1079	conserved hypothetical protein
15	10019	12040	2021	beta-galactosidase Bga42A
16	12111	13622	1511	PAP2 family protein
17	13619	14356	737	acyltransferase
18	14487	14666	179	hypothetical protein
19	14736	15020	284	DNA-binding protein
21	15247	15918	671	hypothetical protein
22	16015	16851	836	degV family protein
25	17246	17542	296	anti-sigma factor antagonist
26	17578	18777	1199	hypothetical protein
27	19014	19619	605	hypothetical protein
28	19773	20411	638	orotate phosphoribosyltransferase PyrE
29	20456	20692	236	hypothetical protein
30	20905	21888	983	phosphate acetyltransferase Pta
33	21993	23348	1355	MATE efflux family protein
34	23339	24745	1406	GGDEF domain-containing protein
35	24832	25779	947	hydrogenase accessory protein HypB
36	25849	27042	1193	hypothetical protein
37	27081	27422	341	hydrogenase nickel insertion protein HypA
39	27736	29043	1307	AAA family ATPase
40	29085	30251	1166	replication initiation protein RepB1
42	32881	33708	827	plasmid partitioning protein ParA
43	33708	34241	533	hypothetical protein
44	34289	34573	284	DNA-binding protein
46	34718	36088	1370	MATE efflux family protein
48	36138	36986	848	MerR family transcriptional regulator
49	37101	37322	221	acyl carrier protein
50	37339	38271	932	enoyl-(acyl-carrier-protein) reductase FabK
51	38282	39217	935	malonyl CoA-acyl carrier protein transacylase FabD
52	39210	39950	740	3-oxoacyl-(acyl-carrier-protein) reductase FabG
53	39943	41175	1232	3-oxoacyl-(acyl-carrier-protein) synthase FabF
54	41188	41610	422	acetyl-CoA carboxylase biotin carboxyl carrier protein AccB
55	41622	42062	440	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ
57	42064	43404	1340	acetyl-CoA carboxylase biotin carboxylase AccC
58	43395	44225	830	acetyl-CoA carboxylase carboxyl transferase beta subunit AccD
59	44248	45003	755	acetyl-CoA carboxylase carboxyl transferase alpha subunit AccA
60	45009	46670	1661	acetyl-CoA synthetase AseA
61	46700	48463	1763	methyl-accepting chemotaxis protein McpA
62	48476	49777	1301	sugar ABC transporter substrate-binding protein
63	49979	50347	368	response regulator domain-containing protein
64	50383	50787	404	ACT domain-containing protein
65	50814	52001	1187	aminotransferase
67	52119	52370	251	GIY-YIG domain-containing protein
68	52371	53411	1040	S-adenosyl-methyltransferase MraW family protein
69	53424	54347	923	radical SAM domain-containing protein
71	54501	56024	1523	SPFH domain/band 7 family protein
72	56085	56264	179	hypothetical protein
73	56472	58616	2144	UvrD/REP family ATP-dependent DNA helicase
75	58926	59843	917	CorA-like Mg ²⁺ transporter protein
76	59887	60537	650	potassium uptake protein TrkA family protein
79	60548	61915	1367	potassium uptake protein TrkH family protein
81	62276	64687	2411	beta-glucosidase Bgl3A
84	64842	65066	224	hypothetical protein
85	65087	65476	389	UBA/TS-N domain-containing protein
86	65547	66725	1178	Na ⁺ /H ⁺ antiporter
87	66894	67661	767	hypothetical protein
88	67751	68392	641	hypothetical protein
89	68502	70067	1565	hypothetical protein
90	70070	70282	212	HTH domain-containing protein
91	70299	70775	476	hypothetical protein
92	71001	71624	623	polysaccharide deacetylase Est4A
93	71880	72545	665	phosphoglycerate mutase family protein
94	72565	73281	716	hypothetical protein
95	73346	74134	788	metallo-beta-lactamase family protein
96	74226	75773	1547	lysozyme Lyc25A
97	75829	76830	1001	GNAT family acetyltransferase

98	76979	78034	1055	feruloyl esterase Est1A
99	78077	79900	1823	ABC transporter ATP-binding protein
100	79902	80906	1004	sugar ABC transporter substrate-binding protein
103	80921	82918	1997	methyl-accepting chemotaxis protein McpB
104	83321	83977	656	hypothetical protein
105	83970	85907	1937	transposase
106	85911	86909	998	hypothetical protein
108	86893	88218	1325	hypothetical protein
109	88297	88557	260	hicA family protein
110	88554	88901	347	hicB family protein
111	89181	90107	926	GNAT family acetyltransferase/NUDIX domain- containing protein
112	90140	90508	368	CrcB-like protein
113	90555	91757	1202	MFS transporter

Table A2. Summary of manual annotation of the genes within the plasmid (pNP6) of MB2003.

Gene number	Start	Stop	Size of gene (bp)	Gene Annotation
2	826	1563	737	>plasmid mobilization protein<
3	1535	1966	431	hypothetical protein
8	3980	4735	755	hypothetical protein
11	5194	5724	530	plasmid replication protein
4	1976	3658	1682	TraG/TraD family protein

Table A3. Summary of manual annotation of the genes within the megaplasmid (pNP144) of MB2003.

Gene number	Start	Stop	Size of gene (bp)	Gene Annotation
1	283	852	569	signal peptidase I LepB
3	1066	1710	644	hypothetical protein
5	1707	3152	1445	hypothetical protein
7	4284	4853	569	restriction endonuclease
8	4876	5628	752	hypothetical protein
9	5606	6814	1208	hypothetical protein
10	6817	7668	851	hypothetical protein
11	7668	10469	2801	TraG/TraD family protein
12	11476	12288	812	partitioning protein ParA
13	12281	12739	458	hypothetical protein
14	12783	14399	1616	hypothetical protein
16	14526	14723	197	hypothetical protein
17	14950	16395	1445	hypothetical protein
19	17410	17568	158	hypothetical protein
21	18430	19395	965	hypothetical protein
22	19557	19721	164	hypothetical protein
23	19799	19951	152	hypothetical protein
25	20131	20754	623	hypothetical protein
26	20826	21566	740	hypothetical protein
29	21782	25969	4187	cell surface protein
30	25969	26799	830	sortase B family protein
31	26857	29181	2324	ATPase involved in pili biogenesis
32	29650	31032	1382	cell surface protein
33	31045	31446	401	hypothetical protein
34	31465	32433	968	hypothetical protein
35	32488	33564	1076	hypothetical protein
37	33655	34617	962	hypothetical protein
38	34689	35657	968	hypothetical protein
39	36071	37207	1136	hypothetical protein
40	37318	40824	3506	hypothetical protein
41	40918	42189	1271	hypothetical protein
42	42469	42780	311	hypothetical protein
43	42824	43195	371	hypothetical protein
44	43260	43598	338	hypothetical protein
45	43631	43972	341	hypothetical protein
46	43972	45048	1076	hypothetical protein
48	45060	47414	2354	hypothetical protein
51	47426	50071	2645	hypothetical protein
52	50204	51418	1214	hypothetical protein
53	51703	53022	1319	hypothetical protein
54	53177	53686	509	hypothetical protein
55	53676	54167	491	hypothetical protein
56	54182	55150	968	hypothetical protein
57	55178	56200	1022	hypothetical protein
58	56251	57315	1064	hypothetical protein
59	57334	58536	1202	hypothetical protein
60	58563	61085	2522	hypothetical protein
61	61226	61564	338	hypothetical protein
62	61567	62457	890	hypothetical protein
63	62775	64124	1349	transposase
64	64147	64458	311	hypothetical protein
65	64458	65510	1052	hypothetical protein
66	65507	65764	257	hypothetical protein
67	65821	66819	998	hypothetical protein
68	67400	67624	224	hypothetical protein
69	67596	67892	296	hypothetical protein
70	67906	68568	662	hypothetical protein
71	68641	70317	1676	hypothetical protein
72	70314	71078	764	hypothetical protein
73	71075	71524	449	hypothetical protein
74	71589	72320	731	ThiF family protein
75	72331	73203	872	hypothetical protein
76	73224	74225	1001	hypothetical protein
77	74419	76809	2390	ATP-dependent DNA helicase UvrD/REP family
78	76858	77757	899	DNA polymerase III, delta subunit
79	77788	78228	440	single-stranded DNA binding protein Ssb1
81	78604	78831	227	hypothetical protein
82	78936	79907	971	hypothetical protein
83	80264	81208	944	hypothetical protein
84	81205	81924	719	DNA polymerase III delta' subunit
85	81943	83601	1658	DNA polymerase III gamma and tau subunits DnaX
87	83723	84004	281	>hypothetical protein<
88	84194	85093	899	phage integrase family protein
89	85186	86250	1064	phage integrase family protein
90	86400	87395	995	transglutaminase domain-containing protein

91	87492	89993	2501	hypothetical protein
93	91145	91441	296	hypothetical protein
94	91641	92600	959	hypothetical protein
95	93021	93449	428	hypothetical protein
96	93816	94082	266	hypothetical protein
97	94075	94272	197	hypothetical protein
98	94407	95306	899	hypothetical protein
99	95775	96323	548	hypothetical protein
100	96394	96567	173	hypothetical protein
101	96646	96879	233	hypothetical protein
102	96986	97603	617	hypothetical protein
103	97828	98172	344	hypothetical protein
104	98876	98998	122	hypothetical protein
105	99505	99765	260	CRISPR-associated protein Cas2
107	99940	103689	3749	hypothetical protein
109	103849	104133	284	hypothetical protein
110	104172	105584	1412	archaeal ATPase family protein
111	105820	106776	956	hypothetical protein
112	107017	107427	410	hypothetical protein
113	107479	107745	266	Fic/DOC family protein
114	107862	108056	194	hypothetical protein
115	108408	108767	359	DNA-binding protein
116	108760	108945	185	hypothetical protein
118	109226	109381	155	hypothetical protein
119	109386	109679	293	hypothetical protein
120	109873	110082	209	hypothetical protein
121	110082	111152	1070	hypothetical protein
125	111762	111875	113	hypothetical protein
126	112214	112333	119	hypothetical protein
128	112640	112783	143	hypothetical protein
130	113023	113277	254	hypothetical protein
131	113356	113508	152	hypothetical protein
132	113611	113952	341	HTH domain-containing protein
133	113966	114139	173	hypothetical protein
134	114136	114270	134	hypothetical protein
135	114270	114824	554	hypothetical protein
136	115071	115301	230	hypothetical protein
137	115561	115842	281	DNA-binding protein
139	116160	116579	419	hypothetical protein
140	116633	116875	242	hypothetical protein
141	117027	117476	449	hypothetical protein
143	117738	117917	179	hypothetical protein
147	118695	119750	1055	HD-GYP domain-containing protein
148	119903	120274	371	hypothetical protein
149	120329	120553	224	hypothetical protein
150	120652	120849	197	hypothetical protein
151	121022	121651	629	DNA replication protein DnaD
152	121663	122043	380	hypothetical protein
153	122100	122330	230	hypothetical protein
154	122323	122802	479	hypothetical protein
155	122841	123008	167	hypothetical protein
156	123080	123691	611	hypothetical protein
157	123711	124124	413	hypothetical protein
158	124148	124627	479	hypothetical protein
159	124694	124921	227	hypothetical protein
160	125287	125577	290	hypothetical protein
161	125606	126538	932	phosphoadenosine phosphosulfate reductase family protein
162	126913	127470	557	hypothetical protein
163	127484	128017	533	hypothetical protein
164	128014	128316	302	hypothetical protein
165	128335	128610	275	hypothetical protein
167	128885	129292	407	hypothetical protein
169	131535	132215	680	hypothetical protein
170	132215	132505	290	hypothetical protein
171	132529	132669	140	hypothetical protein
172	132680	133057	377	hypothetical protein
174	133135	135129	1994	hypothetical protein
175	135144	135278	134	hypothetical protein
176	135305	135619	314	hypothetical protein
177	135633	135830	197	hypothetical protein
178	136150	136920	770	hypothetical protein
179	137109	137396	287	hypothetical protein
180	137397	137843	446	hypothetical protein
181	137972	138409	437	hypothetical protein
182	138964	139761	797	HTH domain-containing protein
183	140028	140342	314	hypothetical protein
184	140420	142033	1613	hypothetical protein
185	142127	142372	245	hypothetical protein
186	142403	143599	1196	replication initiation protein RepB2

Table A4. Summary of manual annotation of the genes within the chromosome of MB2003.

Gene number	Start	Stop	Size of gene (bp)	Annotation
1	349	744	395	HTH domain-containing protein
1	399	1505	1106	hypothetical protein
1	15	140	125	hypothetical protein
1	462	1538	1076	hypothetical protein
1	138	5288	5150	hypothetical protein
1	4	483	479	hypothetical protein
1	7	903	896	prolipoprotein diacylglyceryl transferase Lgt
1	498	854	356	hypothetical protein
1	792	1355	563	phosphoglycerate mutase family protein
1	222	1562	1340	hypothetical protein
2	550	3024	2474	1,4-alpha-glucan branching enzyme GlgB
2	836	1969	1133	hypothetical protein
2	1526	2194	668	L-serine dehydratase beta subunit SdhB
2	438	881	443	hypothetical protein
2	351	650	299	hypothetical protein
2	509	1954	1445	hypothetical protein
2	1557	2657	1100	class II aldolase/adducin family protein
2	5441	5812	371	DtxR family transcriptional regulator
2	916	2013	1097	GTP-binding protein YchF
2	1410	2600	1190	AAA family ATPase
2	237	1115	878	AraC family transcriptional regulator
2	98	2221	2123	hypothetical protein
2	1233	1700	467	acetyltransferase
3	2203	3090	887	L-serine dehydratase alpha subunit SdhA
3	943	2247	1304	23S rRNA methyltransferase
3	2020	3114	1094	hypothetical protein
3	1299	2204	905	chemotaxis protein CheW
3	439	1071	632	hypothetical protein
3	3058	4317	1259	hypothetical protein
3	2682	3923	1241	gamma-glutamyl phosphate reductase ProA
3	5936	7480	1544	HD/KH domain-containing protein
3	2378	2599	221	hypothetical protein
3	1111	2625	1514	hypothetical protein
3	2793	3239	446	hypothetical protein
3	1610	2941	1331	MatE efflux family protein
3	1257	2348	1091	GDP-mannose 4,6-dehydratase gmd
3	2047	2445	398	hypothetical protein
3	1757	2863	1106	glycosyl transferase GT4 family protein
4	2160	3968	1808	threonyl-tRNA synthetase ThrS
4	784	2592	1808	Carboxylesterase type B
4	3143	3883	740	hypothetical protein
4	2311	3126	815	hypothetical protein
4	4402	5571	1169	D-3-phosphoglycerate dehydrogenase SerA
4	4095	4664	569	5-formyltetrahydrofolate cycloligase
4	7651	8262	611	recombination regulator RecX
4	3286	4152	866	NAD-dependent epimerase/dehydratase
4	3248	3721	473	cNMP binding domain-containing protein
4	2357	3310	953	NAD-dependent epimerase/dehydratase
4	2518	3771	1253	hypothetical protein
4	2242	3447	1205	flavin containing amine oxidoreductase family protein
4	2883	3545	662	2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase IspD
5	4175	4864	689	hypothetical protein
5	2460	3428	968	hypothetical protein
5	3922	5514	1592	hypothetical protein
5	5571	6653	1082	phosphoserine aminotransferase SerC
5	705	1364	659	deoxyribose-phosphate aldolase DeoC
5	2624	3220	596	hypothetical protein
5	2701	3687	986	L-asparaginase AnsA
5	3728	5230	1502	AAA family ATPase
5	3790	4854	1064	aldose 1-epimerase
5	3444	4151	707	glycosyl transferase GT2 family protein
5	3872	5014	1142	glycoprotease family protein
6	4871	4981	110	hypothetical protein
6	3110	4132	1022	glycosyl transferase GT2 family protein
6	3916	4707	791	hypothetical protein
6	2779	4281	1502	glycerol kinase GlpK
6	5538	8468	2930	peptidase M16 family protein
6	7063	8589	1526	YkuD domain-containing protein
6	4677	6560	1883	hypothetical protein
6	8316	9383	1067	DNA recombination protein RecA
6	1461	2822	1361	citrate synthase cit
6	3549	4322	773	SAM-dependent methyltransferase
6	3695	4054	359	hypothetical protein
6	4242	4760	518	phosphoglycerate mutase family protein
6	3365	3784	419	SCP domain-containing protein

6	4864	5463	599	recombinational DNA repair protein RecR
6	4218	7415	3197	isoleucyl-tRNA synthetase IleS
6	5025	5786	761	hypothetical protein
7	4286	4792	506	thioesterase superfamily protein
7	4737	5813	1076	hypothetical protein
7	4302	5750	1448	glycerol-3-phosphate dehydrogenase GlpA
7	8485	9441	956	GTP-binding protein Era
7	3210	5444	2234	hypothetical protein
7	2672	3892	1220	tyrosyl-tRNA synthetase TyrS
7	8603	10291	1688	acetolactate synthase large subunit IlvB
7	6629	8296	1667	FAD dependent oxidoreductase
7	2846	3112	266	hypothetical protein
7	4577	6424	1847	beta-lactamase family protein
7	4184	4804	620	hypothetical protein
7	4937	5533	596	hypothetical protein
7	5275	6624	1349	hypothetical protein
7	4012	5802	1790	hypothetical protein
7	5463	5810	347	hypothetical protein
7	5922	6503	581	hypothetical protein
8	3929	5662	1733	alpha-galactosidase Aga27A
8	5100	5579	479	hypothetical protein
8	4937	5857	920	phosphate butyryltransferase ptb
8	5923	6465	542	hypothetical protein
8	5747	7015	1268	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
8	9451	10098	647	DNA repair protein RecO
8	5495	6112	617	imidazole glycerol phosphate synthase glutamine amidotransferase subunit HisH
8	8628	9854	1226	FAD dependent oxidoreductase
8	9629	10165	536	hypothetical protein
8	3198	4691	1493	threonine synthase ThrC
8	6485	7036	551	hypothetical protein
8	5100	6764	1664	methyl-accepting chemotaxis protein Mcp
8	5582	6469	887	hypothetical protein
8	6682	10014	3332	TPR domain-containing protein
8	5838	6293	455	hypothetical protein
8	5814	7412	1598	DNA polymerase III gamma and tau subunits DnaX
8	6518	7444	926	ribonuclease Z
9	6505	8163	1658	cellodextrinase Cel9B
9	5741	6559	818	metallo-beta-lactamase family protein
9	5860	6924	1064	butyrate kinase buk
9	7027	7392	365	hypothetical protein
9	10212	10766	554	hypothetical protein
9	6125	6886	761	imidazole glycerol phosphate synthase cyclase subunit HisF
9	5681	6070	389	hypothetical protein
9	10328	12016	1688	dihydroxy-acid dehydratase IlvD
9	10189	10581	392	hypothetical protein
9	4811	5608	797	hypothetical protein
9	7049	7651	602	hypothetical protein
9	6514	6897	383	glyoxalase family protein
9	6306	7175	869	dTDP-4-dehydrorhamnose reductase rfbD
9	7453	8526	1073	6-phosphofructokinase PfkA
9	7757	8443	686	MIP family transporter
9	7448	8662	1214	glycoprotease family protein
10	6931	8082	1151	reducing end xylose-releasing exo-oligoxylanase Xyn8A
10	8652	10211	1559	polygalacturonase Pgl28B
10	6662	7507	845	GNAT family acetyltransferase/NUDIX domain-containing protein
10	8152	9090	938	PAP2 family protein
10	7611	8999	1388	argininosuccinate lyase ArgH
10	10788	12227	1439	glycyl-tRNA synthetase glyS
10	6893	7570	677	uracil-DNA glycosylase Ung
10	6235	6594	359	ArsR family transcriptional regulator
10	9857	11494	1637	hypothetical protein
10	10598	11329	731	hypothetical protein
10	5609	6490	881	hypothetical protein
10	6909	7445	536	phosphoglycerate mutase family protein
10	10161	11003	842	AraC family transcriptional regulator
10	7233	8405	1172	hypothetical protein
10	8954	10168	1214	hypothetical protein
10	8675	9112	437	hypothetical protein
11	7511	10078	2567	preprotein translocase SecA subunit secA
11	9183	9683	500	hypothetical protein
11	9129	10214	1085	aspartate-semialdehyde dehydrogenase Asd
11	12256	14910	2654	alanine-tRNA ligase alaS
11	7657	8943	1286	diaminopimelate decarboxylase LysA
11	6630	6848	218	heavy metal associated domain-containing protein
11	12023	13717	1694	hypothetical protein
11	11511	12014	503	hypothetical protein
11	11611	11937	326	hypothetical protein
11	7825	8403	578	xanthine phosphoribosyltransferase Xpt
11	8309	8626	317	thioredoxin
11	7462	7713	251	hypothetical protein
11	8461	8724	263	phosphocarrier HPr family
11	10268	11338	1070	threonine aldolase
11	9221	10006	785	hypothetical protein

12	7568	8845	1277	endo-1,4-beta-xylanase Xyn10E
12	6595	7878	1283	beta-N-acetylhexosaminidase Bhx3B
12	10156	11298	1142	peptide chain release factor 2 prfB
12	15035	15211	176	ribosomal protein S21 rpsU
12	6865	8766	1901	heavy metal translocating P-type ATPase
12	13692	14795	1103	3-isopropylmalate dehydrogenase LeuB
12	11995	13206	1211	hypothetical protein
12	8703	9743	1040	hypothetical protein
12	7770	8387	617	hypothetical protein
12	10993	12621	1628	hypothetical protein
12	8726	10075	1349	phosphoglucosamine mutase glmM
12	10190	11071	881	aldose 1-epimerase family protein
12	11358	11828	470	GAF domain-containing protein
13	11389	12447	1058	LacI family transcriptional regulator
13	9695	10483	788	hypothetical protein
13	10248	11624	1376	archaeal ATPase family protein
13	15352	16884	1532	Na ⁺ /H ⁺ antiporter
13	14871	15245	374	preprotein translocase subunit YajC
13	13209	14234	1025	PhoH family protein
13	12045	13106	1061	HD-GYP domain-containing protein
13	8636	9874	1238	serine hydroxymethyltransferase GlyA
13	8445	9437	992	ABC transporter permease
13	9752	10537	785	hypothetical protein
13	10094	11116	1022	YbbR family protein
13	11133	11882	749	hypothetical protein
13	11922	12866	944	oligopeptide ABC transporter ATP-binding protein
13	10410	11504	1094	HD-GYP domain-containing protein
14	12674	13594	920	sugar ABC transporter permease
14	11617	13776	2159	DNA topoisomerase III TopB
14	8772	10712	1940	beta-lactamase family protein
14	15310	16437	1127	queuine tRNA-ribosyltransferase Tgt
14	14244	14726	482	hypothetical protein
14	13107	14408	1301	penicillin-binding protein PbpA
14	10000	10971	971	diacylglycerol kinase family protein
14	10555	12618	2063	2-enoate reductase
14	8528	9148	620	hypothetical protein
14	11368	12117	749	hypothetical protein
15	10488	11030	542	hypothetical protein
15	13791	14480	689	acetyltransferase
15	17020	17826	806	metallophosphoesterase
15	8964	10706	1742	hypothetical protein
15	16503	17633	1130	D-alanyl-D-alanine carboxypeptidase VanY
15	14728	15753	1025	MORN repeat-containing protein
15	9437	10537	1100	ABC transporter permease
15	12618	13421	803	phospholipase
15	9292	9729	437	cytidine deaminase Cdd
15	12662	13582	920	hypothetical protein
15	12251	13192	941	hypothetical protein
15	11956	13026	1070	hypothetical protein
15	12850	13986	1136	oligopeptide ABC transporter ATP-binding protein OppD
15	11541	12830	1289	UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurA
16	13606	14520	914	sugar ABC transporter permease
16	11199	11951	752	tRNA pseudouridine synthase A TruA
16	14629	15321	692	hypothetical protein
16	17828	18577	749	chromosome segregation and condensation protein SepA
16	10797	11690	893	tyrosine recombinase XerD
16	17703	18812	1109	chorismate synthase AroC
16	15913	17298	1385	NADPH-dependent glutamate synthase GltA3
16	14527	15897	1370	rod shape-determining protein RodA
16	11035	11760	725	PAP2 family protein
16	10540	12072	1532	ABC transporter ATP-binding protein
16	14896	15390	494	hypothetical protein
16	10180	11001	821	ornithine carbamoyltransferase Arg
16	13197	13592	395	hypothetical protein
16	13103	13408	305	hypothetical protein
16	13994	15073	1079	oligopeptide ABC transporter permease protein OppC
17	14588	16240	1652	sugar ABC transporter substrate-binding protein
17	12017	13633	1616	hypothetical protein
17	15344	16363	1019	hypothetical protein
17	18581	19150	569	chromosome segregation and condensation protein B scpB
17	11710	13020	1310	aminotransferase domain-containing protein
17	18914	19978	1064	6-phosphogluconolactonase
17	17328	18203	875	oxidoreductase NAD-binding domain-containing protein
17	15902	18139	2237	peptidase U32 family protein
17	12172	13434	1262	Bmp family protein
17	13546	15126	1580	EAL domain-containing protein
17	11059	11658	599	hypothetical protein
17	13662	14849	1187	iron-sulfur cluster-binding protein
17	13602	14027	425	hypothetical protein
17	13383	13940	557	hypothetical protein
17	15075	16010	935	oligopeptide ABC transporter permease protein OppB
18	15808	16482	674	lysozyme M1
18	13637	15349	1712	glycosyl transferase GT39 family protein

18	16488	17933	1445	GGDEF domain-containing protein
18	19371	20723	1352	methylmalonyl-CoA decarboxylase alpha subunit mmdA
18	13083	14111	1028	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase
18	11551	12381	830	purine-nucleoside phosphorylase DeoD
18	20196	21269	1073	spermidine/putrescine ABC transporter ATP-binding protein PotA
18	18304	19752	1448	TPR domain-containing protein
18	18202	18609	407	cell division protein ZapA
18	11813	12790	977	glycosyl transferase GT2 family protein
18	13807	15546	1739	ABC transporter ATP-binding/permease protein
18	11715	12368	653	5'-nucleotidase domain-containing protein
18	14891	15799	908	diacylglycerol kinase family protein
18	14033	14713	680	comF family protein
18	13955	14362	407	hypothetical protein
18	16088	17782	1694	oligopeptide ABC transporter substrate-binding protein OppA
18	12919	14409	1490	metallopeptidase M23 family protein
19	15475	16137	662	hypothetical protein
19	17936	19342	1406	GGDEF domain-containing protein
19	20741	21607	866	sodium ion-translocating decarboxylase gamma subunit
19	12403	13707	1304	amidohydrolase
19	19766	20626	860	methylenetetrahydrofolate dehydrogenase/cyclohydrolase FdhD
19	18660	19670	1010	holliday junction DNA helicase RuvB
19	12845	13480	635	hypothetical protein
19	15543	17162	1619	ABC transporter ATP-binding/permease protein
19	16600	17157	557	GNAT family acetyltransferase
19	12499	13080	581	phosphoglycerate mutase family protein
19	15796	16995	1199	hypothetical protein
19	14739	16988	2249	exodeoxyribonuclease V alpha subunit RecD
19	14359	14883	524	RNA polymerase sigma factor sigma-70 family protein
19	17935	18060	125	conserved hypothetical
19	14532	14972	440	hypothetical protein
20	16314	17795	1481	sucrose-6-phosphate hydrolase Scr32A
20	16375	17217	842	polysaccharide deacetylase Est4
20	21621	21980	359	biotin attachment domain-containing protein
20	14198	16369	2171	adenine-specific DNA methylase
20	13711	15384	1673	MutS domain-containing protein
20	21256	22080	824	spermidine/putrescine ABC transporter permease PotB
20	20721	22397	1676	formate-tetrahydrofolate ligase Fhs
20	13548	14573	1025	DNA polymerase III delta subunit HoIA
20	17281	18744	1463	cobalt ABC transporter ATP-binding protein CbiO
20	17180	17347	167	hypothetical protein
20	13103	14020	917	ribosome small subunit-dependent GTPase A Rsg
20	17065	17457	392	hypothetical protein
20	15037	15456	419	HIT domain-containing protein
20	18403	19635	1232	hypothetical protein
20	15268	15921	653	RelA/SpoT domain-containing protein
21	17799	18767	968	carbohydrate kinase PfkB family protein
21	17232	18251	1019	hypothetical protein
21	19567	20007	440	NfeD family protein
21	22031	23173	1142	sodium ion-translocating decarboxylase beta subunit
21	16170	17276	1106	CDF family transporter
21	15384	15878	494	ribonuclease H
21	19700	20416	716	holliday junction DNA helicase RuvA
21	18741	19427	686	cobalt ABC transporter permease protein CbiQ
21	17397	17876	479	hypothetical protein
21	14028	16067	2039	PASTA domain-containing serine/threonine protein kinase
21	17021	17509	488	dihydrofolate reductase FdhA
21	17472	18266	794	flagellar basal-body rod protein FlgG
21	15484	16185	701	LytTR/response regulator domain-containing protein
21	19686	20354	668	hypothetical protein
21	15932	16861	929	glycosyl transferase GT2 family protein
22	18781	19863	1082	GGDEF domain-containing protein
22	18278	18973	695	uridylylase kinase PyrH
22	20010	20927	917	SPFH domain/band 7 family protein
22	23201	24613	1412	oxaloacetate decarboxylase alpha subunit oadA
22	15895	16527	632	metallophosphoesterase
22	22080	22868	788	spermidine/putrescine ABC transporter permease PotC
22	22398	25241	2843	FtsK/SpoIIIE family protein
22	20424	21473	1049	branched-chain amino acid aminotransferase IlvE
22	14650	16965	2315	DNA internalization-related competence protein ComEC
22	17869	18330	461	NUDIX domain-containing protein
22	16060	16650	590	serine/threonine protein phosphatase
22	18423	19253	830	flagellar basal-body rod protein FlgG
22	16206	16961	755	hypothetical protein
23	19956	21005	1049	hypothetical protein
23	19001	19552	551	ribosome recycling factor Frr
23	17326	18669	1343	GGDEF/EAL domain-containing protein
23	16517	17518	1001	tryptophanyl-tRNA synthetase trpS
23	22858	23928	1070	spermidine/putrescine ABC transporter substrate-binding protein PotD
23	25410	25982	572	hypothetical protein
23	21529	22944	1415	hypothetical protein
23	16986	17849	863	hypothetical protein
23	18451	19170	719	SAM-dependent methyltransferase
23	16868	17908	1040	radical SAM domain-containing protein

23	17571	18455	884	thymidylate synthase thyA
23	19317	20309	992	cell shape determining protein MreB
23	20344	21015	671	hypothetical protein
23	16957	17880	923	cell envelope-related transcriptional attenuator
24	21098	22099	1001	glycosyl transferase GT2 family protein
24	20974	22791	1817	ABC transporter ATP-binding/permease
24	17602	18255	653	transaldolase
24	24173	24859	686	hypothetical protein
24	26040	27908	1868	glucosamine-fructose-6-phosphate aminotransferase GlmS
24	22971	23798	827	hypothetical protein
24	17918	19312	1394	two component system histidine kinase
24	19451	20134	683	hypothetical protein
24	19214	19339	125	hypothetical protein
24	17905	19257	1352	rRNA small subunit methyltransferase B Sun
24	18436	20040	1604	RNA-metabolising metallo-beta-lactamase
24	21031	21945	914	hypothetical protein
24	17955	18284	329	hypothetical protein
25	24742	26922	2180	glycogen debranching enzyme GlgX
25	22099	23136	1037	CDP-glucose 4,6-dehydratase rfbG
25	19556	20281	725	undecaprenyl pyrophosphate synthase UppS
25	18965	20536	1571	beta-lactamase family protein
25	18540	22781	4241	CnaB domain-containing cell-surface protein
25	23815	24816	1001	glycosyl transferase GT2 family protein
25	19432	20112	680	two component system response regulator
25	20187	20999	812	hypothetical protein
25	20208	21353	1145	hypothetical protein
25	20474	23350	2876	excinuclease ABC subunit A uvrA
25	17004	17948	944	hypothetical protein
25	21942	22316	374	hypothetical protein
26	23149	24495	1346	DegT/DnrJ/EryC1/StrS family aminotransferase
26	20306	21130	824	phosphatidate cytidyltransferase CdsA
26	22956	24605	1649	ABC transporter ATP-binding/permease
26	26909	27514	605	hypothetical protein
26	20580	21149	569	hypothetical protein
26	22768	23382	614	sortase family protein
26	27987	28790	803	glutamate 5-kinase ProB
26	24907	25395	488	3-isopropylmalate dehydratase small subunit LeuD
26	20131	20778	647	competence protein ComE
26	19320	19835	515	DNA topology modulation protein FlaR-related protein
26	19311	20585	1274	Na ⁺ -dependent transporter SNF family
26	23396	25393	1997	excinuclease ABC subunit B uvrB
26	17990	18646	656	NUDIX domain-containing protein
26	18382	19788	1406	exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase
27	21364	23358	1994	xylosidase
27	24497	25405	908	NAD-dependent epimerase/dehydratase
27	24612	25514	902	metallophosphoesterase
27	27532	28209	677	two component system response regulator
27	21431	22978	1547	cysteine protease C1 family protein
27	25414	26685	1271	3-isopropylmalate dehydratase large subunit LeuC
27	20825	22267	1442	sugar ABC transporter substrate-binding protein
27	19891	20343	452	hypothetical protein
27	20689	21249	560	hypothetical protein
27	21369	22124	755	hypothetical protein
27	25464	25703	239	hypothetical protein
27	22682	26167	3485	GGDEF/EAL domain-containing protein
27	19785	20756	971	glycosyl transferase GT2 family protein
28	25414	26952	1538	glycosyl transferase GT2 family protein
28	21322	22464	1142	1-deoxy-D-xylulose-5-phosphate reductoisomerase Dxr
28	25539	26021	482	hypothetical protein
28	23082	23372	290	hypothetical protein
28	23475	25028	1553	SSS family transporter
28	29470	30648	1178	acyltransferase
28	26771	29233	2462	hypothetical protein
28	22324	23523	1199	acetylornithine aminotransferase ArgD
28	21429	22361	932	methionyl-tRNA formyltransferase fmt
28	25700	26656	956	serine/threonine protein kinase
28	18750	19559	809	FRG domain-containing protein
28	20760	22163	1403	glycosyl transferase GT2 family protein
29	27065	27694	629	hypothetical protein
29	22472	23524	1052	peptidase M50 family protein
29	26389	26673	284	chaperonin GroES
29	28280	29722	1442	two component system histidine kinase
29	23725	24912	1187	cobalamin biosynthesis protein CbiD
29	30684	31154	470	SsrA-binding protein SmpB
29	29302	30330	1028	ketol-acid reductoisomerase IlvC
29	23385	24242	857	sugar ABC transporter permease
29	22469	23344	875	amino acid ABC transporter substrate-binding protein
29	23272	23856	584	hypothetical protein
29	26659	27408	749	serine/threonine protein phosphatase
29	19702	20484	782	hypothetical protein
29	26382	27443	1061	glycosyl transferase GT2 family protein
29	22191	23930	1739	phosphoglucomutase/phosphomannomutase family protein
30	28048	28383	335	transcriptional regulator

30	23541	24599	1058	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase IspG
30	26726	28348	1622	chaperonin GroEL
30	29734	30615	881	hypothetical protein
30	24934	26379	1445	glycogen synthase ADP-glucose type GlgA
30	25040	25510	470	GNAT family acetyltransferase
30	30545	31048	503	acetolactate synthase small subunit IlvH
30	23545	24441	896	acetylglutamate kinase ArgB
30	24242	25153	911	hypothetical protein
30	23390	24130	740	amino acid ABC transporter ATP-binding protein
30	20561	22123	1562	sugar ABC transporter substrate-binding protein
30	27604	28740	1136	AAA family ATPase
30	24009	26063	2054	glycosyl transferase GT2 family protein
31	25566	26000	434	ACT domain-containing protein
31	31319	32629	1310	hypothetical protein
31	20528	22057	1529	GDSL-family lipase/acylhydrolase
31	27504	28769	1265	C-terminal processing peptidase prc
31	28773	30431	1658	hypothetical protein
31	26079	28190	2111	glycosyl transferase GT2 family protein
32	28504	29691	1187	UDP-galactopyranose mutase glf
32	24635	29164	4529	DNA polymerase III C PolC
32	28778	30226	1448	arginine decarboxylase SpeA
32	31345	31452	107	hypothetical protein
32	26389	28041	1652	polysaccharide biosynthesis protein
32	26038	27339	1301	phenylacetyl-CoA ligase PaaK
32	31173	33632	2459	ribonuclease R Rnr
32	32617	33369	752	hypothetical protein
32	24452	25690	1238	arginine biosynthesis bifunctional protein ArgJ
32	22101	22616	515	hypothetical protein
32	24143	24817	674	amino acid ABC transporter permease
32	23971	25338	1367	potassium uptake protein TrkH family protein
32	28794	30026	1232	metallopeptidase M23 family protein
32	22184	23014	830	sugar ABC transporter permease
32	28202	29137	935	SAM-dependent methyltransferase
33	29861	31165	1304	flagellar hook-associated protein FlhD
33	29744	30199	455	hypothetical protein
33	30428	31300	872	spermidine synthase SpeE
33	28474	29391	917	hypothetical protein
33	27366	27944	578	indolepyruvate ferredoxin oxidoreductase beta subunit IorB
33	33716	33958	242	hypothetical protein
33	25690	26133	443	GNAT family acetyltransferase
33	25285	26745	1460	sugar ABC transporter substrate-binding protein
33	22713	23156	443	hypothetical protein
33	24858	25364	506	peptide deformylase def
33	25420	26781	1361	potassium uptake protein TrkA family protein
33	30491	31687	1196	flavodoxin/beta-lactamase domain-containing protein
33	29113	30360	1247	SAM-dependent methyltransferase
34	30331	31335	1004	beta-lactamase family protein
34	29442	29720	278	hypothetical protein
34	27948	29702	1754	indolepyruvate ferredoxin oxidoreductase alpha subunit IorA
34	34046	35593	1547	2,3-bisphosphoglycerate-independent phosphoglycerate mutase GpmA
34	33530	37234	3704	hypothetical protein
34	26157	27221	1064	N-acetyl-gamma-glutamyl-phosphate reductase ArgC
34	23153	23566	413	hypothetical protein
34	23011	23901	890	sugar ABC transporter permease
34	31707	34802	3095	two component system histidine kinase/response regulator hybrid protein
34	30380	31705	1325	polysaccharide ABC transporter ATP-binding protein
35	31283	33157	1874	hypothetical protein
35	31535	31999	464	molecular chaperone Hsp
35	31404	32666	1262	carboxynorspermidine dehydrogenase
35	29892	31526	1634	AMP-binding enzyme
35	27461	28699	1238	argininosuccinate synthase ArgG
35	23846	24439	593	MFS transporter
35	25505	27751	2246	primosomal protein N' pri
35	26923	27465	542	flavodoxin family protein
35	30019	30765	746	cell division permease FtsX
35	23898	25004	1106	sugar ABC transporter ATP-binding protein
35	34823	36193	1370	sugar ABC transporter substrate-binding protein
35	31779	32768	989	polysaccharide ABC transporter permease
36	33280	33429	149	ribosomal protein L33 rpmG
36	32821	33879	1058	hypothetical protein
36	29899	30945	1046	glutamyl aminopeptidase M42 family
36	31871	32191	320	hypothetical protein
36	37677	38159	482	acetolactate synthase small subunit IlvH
36	27498	28100	602	hypothetical protein
36	30905	31645	740	cell division ATP-binding protein FtsE
36	25020	25190	170	hypothetical protein
36	36222	36737	515	hypothetical protein
36	32777	33322	545	dTDP-4-dehydrothamose 3,5-epimerase RibC2
37	33461	33739	278	preprotein translocase subunit secE
37	32079	32903	824	hypothetical protein
37	33881	34075	194	hypothetical protein
37	30942	32807	1865	fibronectin-binding A domain-containing protein
37	32206	34803	2597	heavy metal translocating P-type ATPase

37	35738	37459	1721	hypothetical protein
37	38202	39827	1625	acetolactate synthase large subunit IlvB
37	27771	29258	1487	UDP-N-acetylmuramoylalanyl-D-glutamate-2,6- diaminopimelate ligase Mur
37	31659	32741	1082	hypothetical protein
37	25187	25519	332	hypothetical protein
37	36748	37251	503	transcription elongation factor GreA/GreB2
37	33350	33811	461	hypothetical protein
38	34100	34828	728	ABC transporter ATP-binding protein
38	34833	35270	437	dUTP diphosphatase Dut
38	27115	28101	986	AraC family transcriptional regulator
38	29416	31212	1796	ssDNA exonuclease RecJ
38	28114	29724	1610	hypothetical protein
38	32970	33158	188	hypothetical protein
38	25532	27070	1538	2Fe-2S iron-sulfur cluster binding domain-containing protein
38	37323	38186	863	fructose-1,6-bisphosphate aldolase FbaA
38	33867	34736	869	glucose-1-phosphate thymidyltransferase RfbA
39	38274	40367	2093	alpha-amylase Amy13E
39	33756	34379	623	transcription termination/antitermination factor NusG
39	32915	33574	659	two component system response regulator
39	33020	34519	1499	altronate oxidoreductase UxaB
39	35353	36816	1463	glutamyl-tRNA synthetase GltX
39	29017	29781	764	response regulator domain-containing protein
39	28128	30449	2321	hypothetical protein
39	32331	34616	2285	protein-export membrane protein SecD
39	33206	34132	926	CAAX amino terminal protease family protein
39	34758	35843	1085	dTDP-glucose 4,6-dehydratase rfbB
40	34506	34931	425	ribosomal protein L11 rplK
40	36797	38725	1928	UvrD/REP family ATP-dependent DNA helicase
40	39846	42287	2441	methionine synthase Meth
40	29875	30711	836	degV family protein
40	30461	31159	698	hypothetical protein
40	34637	36073	1436	radical SAM domain-containing protein
40	29797	31680	1883	threonyl-tRNA synthetase thrZ
40	34142	34990	848	hypothetical protein
40	27205	28365	1160	hypothetical protein
41	34945	35634	689	ribosomal protein L1 rplA
41	33604	34458	854	hypothetical protein
41	34572	36131	1559	ABC transporter ATP-binding protein
41	37478	38671	1193	CDF family transporter
41	42284	43336	1052	hypothetical protein
41	30729	31214	485	hypothetical protein
41	31191	31670	479	hypothetical protein
41	35005	35739	734	acyl-ACP thioesterase
41	40568	42427	1859	ATP-dependent metalloproteinase HflB2
41	35876	37267	1391	Hypothetical protein
42	39194	41140	1946	glycogen debranching enzyme GlgX
42	34706	35356	650	PAP2 family protein
42	34818	36431	1613	hypothetical protein
42	38668	39339	671	thiamine pyrophosphokinase
42	43333	43821	488	hypothetical protein
42	31220	31579	359	Hpt domain-containing protein
42	31876	32307	431	hypothetical protein
42	31771	32250	479	SEC-C domain-containing protein
42	42452	42976	524	hypoxanthine phosphoribosyltransferase Hpt
43	37349	40990	3641	lysozyme Lyc
43	35812	36954	1142	capsule biosynthesis protein
43	36492	37655	1163	carboxynorspermidine decarboxylase NspC
43	41137	42381	1244	tripeptide aminopeptidase PepT
43	39339	40703	1364	hypothetical protein
43	43872	44297	425	ATP synthase F1 epsilon subunit AtpC
43	31670	32737	1067	peptide chain release factor 1 PrfA
43	32405	32926	521	hypothetical protein
43	32588	33415	827	hypothetical protein
43	28388	29434	1046	hypothetical protein
43	42981	44393	1412	tRNA(Ile)-lysidine synthetase TiiS
44	37167	37742	575	ribosomal protein L10 RplJ
44	37687	38925	1238	agmatine deiminase AguA
44	42385	43827	1442	tRNA-thiotransferase enzyme MiaB
44	44301	45665	1364	ATP synthase F1 beta subunit AtpD
44	32785	33624	839	methyltransferase HemK
44	33121	33876	755	hypothetical protein
44	36391	37671	1280	hypothetical protein
44	33501	34727	1226	L,L-diaminopimelate aminotransferase
44	35996	38107	2111	glutamine synthetase type III GlnA
44	44455	44766	311	septum formation initiator DivIC
45	35353	36699	1346	MatE efflux family protein
45	38944	39843	899	N-carbamoylputrescine amidohydrolase AguB
45	36191	38299	2108	hypothetical protein
45	40930	41574	644	sugar-phosphate isomerase LacAB/RpiB family
45	45679	46542	863	ATP synthase F1 gamma subunit AtpG
45	33631	34689	1058	hypothetical protein
45	37695	38009	314	L-rhamnose 1-epimerase rha
45	38206	39372	1166	peptidase U32 family protein

45	29456	30346	890	hypothetical protein
45	44835	45074	239	RNA-binding S4 domain-containing protein
46	41637	43529	1892	Alpha-L-arabinofuranosidase
46	37814	38191	377	ribosomal protein L7/L12 rpLL
46	36831	37745	914	LysR family transcriptional regulator
46	38375	38944	569	hypothetical protein
46	43857	46535	2678	DNA mismatch repair protein MutS
46	46598	48391	1793	ATP synthase F1 alpha subunit AtpA
46	34770	34976	206	ribosomal protein L31 RpmE
46	38033	39298	1265	histidyl-tRNA synthetase his
46	35006	35917	911	diaminopimelate epimerase dapF
46	39441	40088	647	O-methyltransferase
46	41116	42621	1505	hypothetical protein
47	37951	38886	935	cysteine synthase CysK
47	39877	40698	821	hypothetical protein
47	38975	39889	914	hypothetical protein
47	46536	48458	1922	DNA mismatch repair protein MutL
47	43546	44400	854	HAD superfamily hydrolase
47	48384	48875	491	ATP synthase F0 B subunit AtpF
47	35086	35709	623	RNA polymerase sigma factor sigma-70 family protein
47	33876	35204	1328	hypothetical protein
47	39372	40952	1580	EAL domain-containing protein
47	36165	38033	1868	ABC transporter ATP binding/permease
47	40078	40569	491	hypothetical protein
47	45075	46658	1583	PEGA domain-containing protein
48	38530	42351	3821	DNA-directed RNA polymerase beta subunit rpoB
48	40695	41930	1235	hypothetical protein
48	48455	49393	938	tRNA delta(2)-isopentenylpyrophosphate transferase MiaA
48	44393	45607	1214	ion transport protein
48	35715	36464	749	rRNA methylase SpoU family protein
48	41327	43279	1952	EAL domain-containing protein
48	30407	32185	1778	GGDEF domain-containing protein
48	46785	48143	1358	tRNA nucleotidyltransferase Cca
48	42655	44067	1412	hypothetical protein
49	42396	46133	3737	DNA directed RNA polymerase beta-prime subunit RpoC
49	39067	39975	908	LysR family transcriptional regulator
49	40091	41815	1724	hypothetical protein
49	49386	50669	1283	aluminum resistance protein
49	48890	49108	218	ATP synthase F0 C subunit AtpE
49	36480	36917	437	ribonuclease III family protein
49	43303	43605	302	hypothetical protein
49	38035	39843	1808	ABC transporter ATP binding/permease
49	32196	33869	1673	GGDEF domain-containing protein
49	44182	44526	344	hypothetical protein
50	46278	46697	419	ribosomal protein S12 rpsL
50	41985	42872	887	molecular chaperone Hsp33
50	42103	42951	848	hypothetical protein
50	50749	51462	713	DNA repair protein RadC
50	39936	40598	662	two component system histidine kinase
51	43654	45861	2207	beta-galactosidase Bga35
51	46817	47332	515	ribosomal protein S7 rpsG
51	40005	40907	902	AraC family transcriptional regulator
51	43469	44932	1463	Type I restriction-modification system methyltransferase subunit
51	51472	52476	1004	rod shape-determining protein MreB
51	45635	47479	1844	RNA helicase DEAD/DEAH box family protein
51	49655	49795	140	hypothetical protein
51	36902	38326	1424	cysteinyI-tRNA synthetase CysS
51	40603	41301	698	two component system response regulator
51	40591	42270	1679	RNA-metabolising metallo-beta-lactamase
51	33871	34398	527	hypothetical protein
51	48222	49940	1718	GGDEF domain-containing protein
51	44523	44906	383	hypothetical protein
52	47435	49504	2069	translation elongation factor G fusA
52	40982	42361	1379	MatE efflux family protein
52	44956	46140	1184	hypothetical protein
52	52477	53346	869	rod shape-determining protein MreC
52	47508	48050	542	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase PgsA
52	49130	49873	743	ATP synthase F0 A subunit AtpB
52	38352	38897	545	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase IspF
52	41314	42150	836	SAM-dependent methyltransferase
52	42332	42607	275	hypothetical protein
52	44912	45883	971	hypothetical protein
53	43061	44782	1721	methyl-accepting chemotaxis protein Mcp
53	53352	53876	524	rod shape-determining protein MreD
53	48050	49408	1358	MiaB-like tRNA modifying enzyme
53	38947	39222	275	branched-chain amino acid transport protein AziD
53	45887	46900	1013	LacI family transcriptional regulator
53	42154	42627	473	small multi-drug export family transporter
53	42621	43040	419	holliday junction resolvase YqgF
53	34452	35189	737	hypothetical protein
53	45899	47008	1109	hypothetical protein
54	44916	47654	2738	glycoside hydrolase family 2 Gh2F
54	49643	50833	1190	translation elongation factor Tu tufA

54	42453	44201	1748	hypothetical protein
54	49405	49683	278	DNA-directed RNA polymerase omega subunit RpoZ
54	49961	52270	2309	hypothetical protein
54	46915	47775	860	sugar ABC transporter permease
54	42627	43682	1055	histidinol-phosphate transaminase hisC
54	43054	43329	275	hypothetical protein
54	35291	36001	710	hypothetical protein
54	49933	51618	1685	GGDEF domain-containing protein
55	50949	51083	134	hypothetical protein
55	44265	45167	902	hypothetical protein
55	53869	56727	2858	penicillin-binding protein PbpA
55	49724	50347	623	guanylate kinase Gmk
55	52313	52633	320	hypothetical protein
55	39245	39961	716	branched-chain amino acid transport protein AziC
55	47799	49199	1400	sugar ABC transporter permease
55	43411	44763	1352	MiaB-like tRNA modifying enzyme MiaB
56	51127	51684	557	hypothetical protein
56	45183	46157	974	D-isomer specific 2-hydroxyacid dehydrogenase
56	47764	48402	638	CoA-binding domain-containing protein
56	46398	47408	1010	phage integrase family protein
56	56740	57864	1124	rod shape-determining protein RodA
56	50337	50507	170	hypothetical protein
56	52856	55375	2519	hypothetical protein
56	39985	41202	1217	alanyl-tRNA synthetase family protein
56	49295	50620	1325	sugar ABC transporter substrate-binding protein
56	44856	45086	230	hypothetical protein
56	36017	37003	986	hypothetical protein
56	51625	53058	1433	prolyl-tRNA synthetase ProS
57	51732	52151	419	GNAT family acetyltransferase
57	48419	49081	662	hypothetical protein
57	47408	48226	818	hypothetical protein
57	57869	58264	395	methylglyoxal synthase MgsA
57	50593	51507	914	hypothetical protein
57	55407	55814	407	anti-sigma regulatory factor
57	41279	41797	518	CarD family transcriptional regulator
57	50880	52037	1157	DNA protecting protein Dpr
57	44254	44961	707	bacterial sugar transferase
57	37073	37390	317	hypothetical protein
57	47047	48507	1460	MBOAT family acyltransferase
58	52246	52389	143	NUDIX domain-containing protein
58	49155	50711	1556	carbohydrate kinase family protein
58	48266	49378	1112	hypothetical protein
58	58239	59213	974	D-alanyl-D-alanine carboxypeptidase DacA
58	51566	52930	1364	hypothetical protein
58	55882	56559	677	hypothetical protein
58	41938	42492	554	GNAT family acetyltransferase
58	52027	53607	1580	Mg chelatase-like protein
58	45192	46394	1202	thiamine biosynthesis/tRNA modification protein Thil
58	37488	38033	545	hypothetical protein
58	53655	54206	551	anaerobic ribonucleoside triphosphate reductase activating protein NrdG
58	48609	49322	713	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase IspD
59	52386	52928	542	hypothetical protein
59	47036	47800	764	hypothetical protein
59	50712	51869	1157	alanine racemase Alr
59	59285	59704	419	hypothetical protein
59	56593	57639	1046	thiamin biosynthesis lipoprotein ApbE
59	42494	42931	437	hypothetical protein
59	44986	45717	731	hypothetical protein
59	46409	47566	1157	cysteine desulfurase IscS
59	38033	38593	560	hypothetical protein
59	54296	56527	2231	anaerobic ribonucleoside triphosphate reductase NrdD
59	49342	50184	842	LicD family protein
60	57636	59906	2270	glycoside hydrolase family 31 Gh31A
60	53049	53924	875	metallophosphoesterase
60	48797	49258	461	hypothetical protein
60	49412	52765	3353	Type I site-specific restriction- modification system R (restriction) subunit and related helicases
60	59758	61365	1607	hypothetical protein
60	52954	53226	272	ACT domain-containing protein
60	53771	54832	1061	twitching motility protein pil
60	45740	46414	674	sortase B family protein
60	47571	48323	752	hypothetical protein
60	38628	39107	479	YbaK family protein
60	56647	57465	818	glutamate racemase MurI
60	50367	51158	791	hypothetical protein
61	59923	62127	2204	glycoside hydrolase family 95 Gh95A
61	49299	50441	1142	transcription termination factor NusA
61	52810	53493	683	hypothetical protein
61	61446	61973	527	hypothetical protein
61	53315	54112	797	metallo-beta-lactamase family protein
61	42928	44757	1829	EAL domain-containing protein
61	54851	55936	1085	HD-GYP domain-containing protein
61	46488	47546	1058	cell envelope-related transcriptional attenuator

61	48323	49360	1037	ribosomal protein L11 methyltransferase prmA
61	39135	39959	824	radical SAM domain-containing protein
61	57556	57945	389	hypothetical protein
61	51180	52481	1301	hypothetical protein
62	62161	64023	1862	alpha-galactosidase Aga36B
62	44890	46194	1304	alpha-amylase Amy13C
62	54086	54580	494	hypothetical protein
62	50461	50700	239	hypothetical protein
62	53616	54665	1049	hypothetical protein
62	62043	63146	1103	3-dehydroquinate synthase AroB
62	56046	56642	596	HD domain-containing protein
62	47556	48020	464	hypothetical protein
62	49441	50622	1181	chaperone protein DnaJ
62	40145	41104	959	membrane transport protein family transporter
62	57947	58660	713	bacterial SH3 domain-containing protein
62	53032	53649	617	uridine kinase Udk
63	54624	56234	1610	hypothetical protein
63	50744	51031	287	ribosomal protein L7A
63	55271	55603	332	hypothetical protein
63	63161	63700	539	signal peptidase II LspA
63	54188	56347	2159	cell division protein FtsA
63	64131	65891	1760	sugar ABC transporter substrate-binding protein
63	46589	47902	1313	hypothetical protein
63	56673	57851	1178	oxygen-independent coproporphyrinogen III oxidase Hem
63	48039	48872	833	polysaccharide biosynthesis protein
63	58701	59021	320	thioredoxin
63	53639	55126	1487	hypothetical protein
64	56442	57500	1058	hypothetical protein
64	56347	56967	620	dephospho-CoA kinase CoaE
64	65933	66820	887	sugar ABC transporter permease
64	48886	49608	722	polysaccharide export protein
64	50742	52646	1904	chaperone protein DnaK
64	41101	41940	839	AraC family transcriptional regulator
64	59021	59674	653	tRNA (guanine-N7)-methyltransferase TrmB
65	57546	58070	524	isochorismatase family hydrolase
65	56018	56902	884	hypothetical protein
65	63700	64629	929	pseudouridine synthase RluA family protein
65	66851	67786	935	sugar ABC transporter permease
65	49637	50329	692	polysaccharide biosynthesis protein
65	52674	53318	644	chaperone protein GrpE
65	42005	42952	947	adenosine deaminase Add
65	59749	61218	1469	D-alanyl-D-alanine carboxypeptidase DacA3
65	55142	56602	1460	MBOAT family acyltransferase
66	67926	70403	2477	beta-galactosidase Bga2B
66	58213	58908	695	hypothetical protein
66	51032	53920	2888	translation initiation factor IF-2 InfB
66	56950	57687	737	hypothetical protein
66	64664	65299	635	hypothetical protein
66	56972	59086	2114	DNA polymerase I PolA
66	48365	49504	1139	hypothetical protein
66	57851	59659	1808	GTP-binding protein Lep
66	53362	54393	1031	heat-inducible transcription repressor HrcA
66	42973	43824	851	aminoglycoside 6-adenyltransferase
67	59069	59287	218	hypothetical protein
67	57799	58359	560	hypothetical protein
67	65527	66111	584	hypothetical protein
67	59109	59432	323	hypothetical protein
67	70441	72750	2309	two component system response regulator
67	49675	50169	494	hypothetical protein
67	59871	60134	263	ribosomal protein S20 rps
67	54611	55117	506	hypothetical protein
67	43853	44647	794	SAM-dependent methyltransferase
67	61234	62172	938	exonuclease family protein
67	56595	57539	944	glycosyl transferase GT2 family protein
68	62188	63015	827	acetyl-xylan esterase Est10A
68	59295	60530	1235	hypothetical protein
68	53957	54394	437	ribosome-binding factor A RbfA
68	58505	58720	215	hypothetical protein
68	66175	67185	1010	phage integrase family protein
68	59550	60998	1448	sugar transporter GPH family protein
68	50737	53358	2621	ATP-dependent chaperone protein ClpB
68	55257	56447	1190	aspartate/tyrosine/aromatic aminotransferase
68	44674	45105	431	NUDIX domain-containing protein
68	57610	59187	1577	flavin containing amine oxidoreductase
69	54391	55341	950	DHH domain-containing protein
69	58743	59060	317	HNH endonuclease domain-containing protein
69	60998	62470	1472	hypothetical protein
69	72957	75764	2807	pyruvate phosphate dikinase PpdK
69	60208	61221	1013	hypothetical protein
69	53431	54576	1145	cobalamin adenosyltransferase CobA
69	56624	57577	953	hypothetical protein
69	63171	63713	542	glutathione peroxidase
69	59349	60581	1232	GDSL-family lipase/acylhydrolase

70	60720	61457	737	RelA/SpoT domain-containing protein
70	55411	56367	956	tRNA pseudouridine synthase B TruB
70	59106	59447	341	ArsC family protein
70	67347	68492	1145	D-alanyl-D-alanine carboxypeptidase DacA
70	75986	76810	824	glycerophosphoryl diester phosphodiesterase
70	50433	51287	854	hypothetical protein
70	61251	62210	959	cobalamin biosynthesis protein Cob
70	45255	46208	953	hypothetical protein
70	63720	64529	809	RNA-binding S4 domain-containing protein
70	60711	61946	1235	polysaccharide biosynthesis protein
71	61570	63798	2228	hypothetical protein
71	59507	60142	635	hypothetical protein
71	68613	70136	1523	DHH domain-containing protein
71	62629	63714	1085	HD-GYP domain-containing protein
71	77100	77744	644	hypothetical protein
71	51284	52909	1625	hypothetical protein
71	62217	63431	1214	cobalamin biosynthesis protein Cob
71	54577	55182	605	hypothetical protein
71	57850	58956	1106	hypothetical protein
71	64531	65106	575	NUDIX domain-containing protein
71	61936	62904	968	glycosyl transferase GT2 family protein
72	63918	65249	1331	capsule biosynthesis protein
72	56378	57286	908	riboflavin biosynthesis protein RibF
72	70194	70610	416	LysM domain-containing protein
72	63904	64500	596	NUDIX domain-containing protein
72	52890	53507	617	hypothetical protein
72	63499	65373	1874	hypothetical protein
72	55236	56033	797	thiamine biosynthesis protein ThiF
72	58953	59276	323	hypothetical protein
72	46780	48330	1550	GMP synthase GuaA
73	65378	66184	806	hypothetical protein
73	57305	58243	938	membrane transport protein family transporter
73	60415	60645	230	hypothetical protein
73	70784	71404	620	LexA repressor
73	64708	65010	302	hypothetical protein
73	77816	79255	1439	MBOAT family acyltransferase
73	65560	66054	494	heptaprenyl diphosphate synthase component I family protein
73	48480	49796	1316	NLPC/P60 domain-containing protein
73	65744	67618	1874	cysteine protease C11 family protein
74	66268	66789	521	peptidyl-prolyl cis-trans isomerase cyclophilin- type
74	58304	59998	1694	phosphoglucomutase/phosphomannomutase family protein
74	60704	61153	449	hypothetical protein
74	71451	71786	335	hypothetical protein
74	79318	80799	1481	hypothetical protein
74	53821	54147	326	hypothetical protein
74	66074	66466	392	hypothetical protein
74	56119	56718	599	GTP-binding protein
74	59425	60339	914	hypothetical protein
74	49842	51233	1391	asparaginyl-tRNA synthetase AsnS
74	67679	67927	248	hypothetical protein
75	60157	60354	197	PspC domain-containing protein
75	71821	72405	584	HD domain-containing protein
75	65097	66743	1646	manganese-dependent inorganic pyrophosphatase Ppa
75	80830	82899	2069	hypothetical protein
75	66557	67342	785	chemotaxis protein methyltransferase Che
75	60339	61583	1244	hypothetical protein
75	51259	52485	1226	peptidyl-prolyl cis-trans isomerase FKBP-type
75	63010	64872	1862	hypothetical protein
76	60509	61600	1091	hypothetical protein
76	61361	62218	857	hypothetical protein
76	72402	73010	608	nicotinate (nicotinamide) nucleotide adenyltransferase NadD
76	66843	68573	1730	ABC transporter ATP-binding/permease
76	83026	84444	1418	GGDEF domain-containing protein
76	67496	68638	1142	ribosomal protein S1 rps
76	56772	58553	1781	ABC transporter ATP-binding/permease
76	52659	53132	473	hypothetical protein
76	64894	65913	1019	radical SAM domain-containing protein
77	67842	69044	1202	hypothetical protein
77	62257	63366	1109	hypothetical protein
77	73041	73352	311	hypothetical protein
77	68570	70294	1724	ABC transporter ATP-binding/permease
77	54150	55667	1517	hypothetical protein
77	68619	69473	854	hydroxymethylbutenyl pyrophosphate reductase Isp
77	58553	60238	1685	ABC transporter ATP-binding/permease
77	61675	62097	422	hypothetical protein
77	53160	53612	452	hypothetical protein
77	68399	68839	440	hypothetical protein
77	65954	67447	1493	hypothetical protein
78	69397	69639	242	hypothetical protein
78	61604	62209	605	hypothetical protein
78	63366	64604	1238	hypothetical protein
78	73349	74701	1352	GTP-binding protein Obg/CgtA
78	70287	71810	1523	RNA methylase NOL1/NOP2/sun family protein

78	84574	85881	1307	MFS transporter
78	55792	56712	920	hypothetical protein
78	69482	70141	659	cytidylate kinase cmk
78	60606	61883	1277	ATP-dependent Clp protease ATP-binding subunit ClpX
78	62253	63245	992	hypothetical protein
78	53625	54494	869	4-diphosphocytidyl-2C-methyl-D-erythritol kinase IspE
79	69728	70270	542	hypothetical protein
79	62369	62635	266	ribosomal protein S15 RpsO
79	64626	65018	392	hypothetical protein
79	74765	75052	287	ribosomal protein L27 RpmA
79	56949	57251	302	hypothetical protein
79	70161	71426	1265	FAD dependent oxidoreductase
79	61905	62507	602	ATP-dependent Clp protease proteolytic subunit ClpP
80	70284	71345	1061	hypothetical protein
80	62870	64960	2090	polyribonucleotide nucleotidyltransferase
80	65345	66265	920	peptidyl-prolyl cis-trans isomerase FKBP-type
80	75056	75385	329	hypothetical protein
80	85908	87548	1640	hypothetical protein
80	57251	57586	335	hypothetical protein
80	63375	63989	614	CAAX amino terminal protease family protein
80	54668	55291	623	hypothetical protein
81	57928	59391	1463	endo-1,4-beta-glucanase Cel5C
81	71503	72531	1028	CAAX amino terminal protease family protein
81	75407	75715	308	ribosomal protein L21 RplU
81	88069	88956	887	degV family protein
81	71438	72379	941	HTH/SIS domain-containing protein
81	63986	64234	248	hypothetical protein
81	69246	70016	770	hypothetical protein
81	67482	68735	1253	hypothetical protein
82	72609	74219	1610	alpha-amylase Amy13B
82	65235	66839	1604	peptide chain release factor 3 PrfC
82	66439	68229	1790	ABC transporter ATP-binding protein
82	89002	89250	248	hypothetical protein
82	59724	61769	2045	UvrD/REP family ATP-dependent DNA helicase
82	72376	72927	551	TetR family transcriptional regulator
82	62719	63318	599	hypothetical protein
82	64221	64772	551	hypothetical protein
82	55367	56272	905	dihydroorotate dehydrogenase PyrD
82	70032	70733	701	hypothetical protein
82	68790	70112	1322	mannose-6-phosphate isomerase/mannose-1-phosphate guanylyl transferase
83	74397	75068	671	HAD superfamily hydrolase
83	66978	68291	1313	hypothetical protein
83	68456	69964	1508	ABC transporter permease protein
83	75801	76967	1166	ribonuclease Rne/Rng family
83	89374	91032	1658	GGDEF/EAL domain-containing protein
83	73047	73676	629	MgtC family protein
84	75092	76891	1799	oligoendopeptidase pepF
84	68354	68728	374	hypothetical protein
84	76960	77688	728	hypothetical protein
84	61835	62812	977	CAAX amino terminal protease family protein
84	63438	64946	1508	trigger factor tig
84	64917	65636	719	hypothetical protein
84	70548	71687	1139	DegT/DnrJ/EryCI/StrS family aminotransferase
85	76905	77522	617	flavin reductase domain-containing protein
85	68744	69148	404	transcription antitermination factor NusB
85	77681	79561	1880	radical SAM domain-containing protein
85	91222	91500	278	DNA-binding protein
85	62918	63523	605	YkuD domain-containing protein
85	74063	76621	2558	GGDEF/EAL domain-containing protein
85	65258	65656	398	GNAT family acetyltransferase
85	65715	67448	1733	ABC transporter ATP-binding/permease
85	56288	57088	800	dihydroorotate dehydrogenase electron transfer subunit
85	71050	71757	707	HTH domain-containing protein
85	71792	72727	935	glycosyl transferase
86	69186	70364	1178	exodeoxyribonuclease VII large subunit XseA
86	69946	70539	593	hypothetical protein
86	79571	80788	1217	bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase protein HisIE
86	91645	92904	1259	6-phosphofructokinase PfkA
86	63646	64986	1340	MatE efflux family protein
86	76607	77137	530	hypothetical protein
86	65736	66302	566	DJ-1 family protein
86	57108	58421	1313	dihydroorotase PyrC
86	72736	73161	425	WxcM-like protein
87	77522	78103	581	hypothetical protein
87	70664	71341	677	hypothetical protein
87	80790	81380	590	imidazoleglycerol-phosphate dehydratase HisB
87	92972	94054	1082	endonuclease/exonuclease/phosphatase family protein
87	65169	66956	1787	arginyl-tRNA synthetase ArgS
87	77139	80060	2921	helicase SNF2 family
87	66431	66961	530	hypothetical protein
87	67466	71035	3569	transcription-repair coupling factor mfd
87	58554	59297	743	1-acyl-sn-glycerol-3-phosphate acyltransferase PlsC

87	71830	73629	1799	HD-GYP/cache domain-containing protein
88	73716	75443	1727	glycoside hydrolase family 30 Gh30A
88	78399	78827	428	translation initiation factor IF-3 infC
88	70345	70578	233	exodeoxyribonuclease VII small subunit XseB
88	71422	72759	1337	cobalt ABC transporter ATP-binding protein CbiO
88	94128	94334	206	hypothetical protein
88	66994	67632	638	hypothetical protein
88	80279	80779	500	shikimate kinase Aro
88	67067	68095	1028	LacI family transcriptional regulator
88	71039	71659	620	peptidyl-tRNA hydrolase pth
88	73173	74801	1628	hypothetical protein
89	80913	83864	2951	carbohydrate binding protein
89	75454	76512	1058	glycoside hydrolase
89	78897	79097	200	ribosomal protein L35 rpml
89	70578	71504	926	geranylgeranyl pyrophosphate synthase IspA
89	72809	73339	530	hypothetical protein
89	81564	82853	1289	histidinol dehydrogenase HisD
89	67611	68195	584	hypothetical protein
89	71764	72051	287	SpoVG family protein
89	59343	60551	1208	D-alanine-D-alanine ligase DdlA
89	74777	75895	1118	NAD-dependent epimerase/dehydratase
90	84329	86659	2330	glycoside hydrolase
90	68121	69641	1520	4-alpha-glucanotransferase Mal77A
90	79148	79504	356	ribosomal protein L20 rpIT
90	73336	74271	935	cobalamin biosynthesis protein CobW
90	82855	83379	524	ATP phosphoribosyltransferase HisG
90	94376	95443	1067	hypothetical protein
90	68188	68703	515	hypothetical protein
90	76532	77659	1127	aldo/keto reductase family oxidoreductase
91	79610	81061	1451	aminoacyl-histidine dipeptidase pepD
91	71521	73371	1850	deoxyxylulose-5-phosphate synthase Dxs
91	74425	75642	1217	iron-containing alcohol dehydrogenase
91	83376	83492	116	ATP phosphoribosyltransferase HisG
91	95511	96374	863	YihY family protein
91	68901	70301	1400	shikimate 5-dehydrogenase/shikimate kinase
91	69753	70694	941	carbohydrate kinase PfkB family protein
91	72088	73206	1118	glucose-1-phosphate adenyltransferase glgD
91	60590	62170	1580	UDP-N-acetylmuramyl pentapeptide synthase MurF
91	75943	77295	1352	undecaprenyl-phosphate glucose phosphotransferase
92	81095	81661	566	TetR family transcriptional regulator
92	73378	74277	899	hemolysin TlyA family protein
92	75877	76263	386	GntR family transcriptional regulator
92	83504	84616	1112	ATP phosphoribosyltransferase regulatory subunit HisZ
92	96730	98553	1823	metallopeptidase M24 family
92	70435	70848	413	hypothetical protein
92	86895	87605	710	triosephosphate isomerase Tpi
92	70706	71308	602	TetR family transcriptional regulator
92	73203	74477	1274	glucose-1-phosphate adenyltransferase glgC
92	62183	62977	794	alpha/beta fold family hydrolase
92	77849	78736	887	EAL domain-containing protein
92	77430	79034	1604	polysaccharide biosynthesis protein
93	63055	64629	1574	endo-1,4-beta-glucanase Cel9A
93	76260	77165	905	ABC transporter ATP-binding protein
93	84692	85240	548	recombination protein U RecU
93	70996	71751	755	peptide/nickel ABC transporter ATP-binding protein
93	87745	88983	1238	phosphoglycerate kinase Pfk
93	71485	71775	290	hypothetical protein
93	78770	80419	1649	ABC transporter, ATP-binding protein
93	79061	80470	1409	glycosyl transferase GT26 family protein
94	81691	84039	2348	alpha-galactosidase Aga36A
94	85242	86216	974	RNA pseudouridylylase synthase
94	71762	72562	800	peptide/nickel ABC transporter ATP-binding protein
94	89148	90173	1025	glyceraldehyde-3-phosphate dehydrogenase Gap
94	71825	72679	854	hypothetical protein
94	74737	76131	1394	UDP-N-acetylmuramate--alanine ligase murC
94	64670	66961	2291	aconitate hydratase AcnA
95	84071	85105	1034	aldo/keto reductase family oxidoreductase
95	74298	75140	842	ATP-NAD kinase PpnK
95	77233	79326	2093	hypothetical protein
95	98672	99154	482	hypothetical protein
95	72579	73364	785	peptide/nickel ABC transporter permease
95	90398	91216	818	histidinol phosphate phosphatase His
95	72702	74408	1706	glutaminyl-tRNA synthetase GlnS
95	76473	77597	1124	primosome protein DnaD
95	80422	82248	1826	ABC transporter ATP-binding/permease protein
95	80472	82133	1661	hypothetical protein
96	99315	101069	1754	carbohydrate esterase family 12 Est12A
96	85155	85460	305	hypothetical protein
96	75169	75624	455	arginine repressor ArgR
96	79385	80251	866	ABC transporter permease
96	86228	86830	602	hypothetical protein
96	73364	74314	950	peptide/nickel ABC transporter permease
96	91297	92034	737	GGDEF domain-containing protein

96	74463	77177	2714	GGDEF/EAL domain-containing protein
96	77740	78723	983	DNA replication protein DnaC
96	82310	83686	1376	hypothetical protein
96	82163	83125	962	radical SAM domain-containing protein
97	75664	77355	1691	DNA repair protein RecN
97	86840	87502	662	HAD-superfamily hydrolase
97	101285	102925	1640	ABC transporter ATP-binding protein
97	74335	75867	1532	peptide/nickel ABC transporter substrate-binding protein
97	77212	78663	1451	IMP dehydrogenase GuaB
97	78723	79937	1214	ribose-phosphate diphosphokinase prs
97	67046	68560	1514	hypothetical protein
97	83520	84209	689	hypothetical protein
98	77360	78007	647	hypothetical protein
98	80273	81169	896	ABC transporter permease
98	87506	88234	728	RNA pseudouridylate synthase
98	76035	76547	512	NrdR family transcriptional regulator
98	92009	92995	986	transglutaminase domain-containing protein
98	80016	80672	656	HAD superfamily hydrolase
99	85524	86393	869	sugar ABC transporter permease
99	81198	82715	1517	RND family transporter
99	88257	89306	1049	N-acetylmuramoyl-L-alanine amidase
99	103010	103882	872	phosphoribosylaminoimidazole-succinocarboxamide synthase PurC
99	93014	94066	1052	type II secretion system protein F Gsp
99	78684	79196	512	hypothetical protein
99	68563	69084	521	CMP/dCMP deaminase zinc-binding domain-containing protein
99	83877	84629	752	two component system response regulator
100	78007	80466	2459	ATP-dependent DNA helicase PcrA
100	82715	83989	1274	ABC transporter permease
100	89363	90061	698	hypothetical protein
100	76723	77970	1247	cell division protein FtsZ
100	94076	94420	344	hypothetical protein
100	79508	80305	797	hypothetical protein
100	80653	81435	782	phosphoglycerate mutase family protein
100	69178	69588	410	Rrf2 family transcriptional regulator
100	84638	86365	1727	two component system histidine kinase
101	80549	80815	266	hypothetical protein
101	83993	86476	2483	ABC transporter substrate-binding protein
101	90080	90565	485	panetheine-phosphate adenyltransferase CoaD
101	103951	105981	2030	methyl-accepting chemotaxis protein McpA
101	94430	94918	488	hypothetical protein
101	80358	81677	1319	aspartate kinase
101	81450	81743	293	hypothetical protein
101	69821	70912	1091	hypothetical protein
101	84310	85860	1550	hypothetical protein
102	81769	84873	3104	beta-galactosidase Bga2A
102	86393	87490	1097	sugar ABC transporter permease
102	80828	82240	1412	23S rRNA methyltransferase
102	86596	87081	485	mechanosensitive ion channel protein MscL family
102	90562	91119	557	hypothetical protein
102	106239	108278	2039	methyl-accepting chemotaxis protein McpA
102	78081	78782	701	cell division protein FtsQ
102	81703	82923	1220	homoserine dehydrogenase
102	70999	72108	1109	ABC transporter ATP-binding protein
102	86368	87411	1043	iron ABC transporter substrate-binding protein
103	82625	82843	218	hypothetical protein
103	87153	88607	1454	cobyric acid synthase CobQ
103	91116	93353	2237	DNA topoisomerase IV subunit A ParC
103	108707	109309	602	phosphoserine phosphatase/homoserine phosphotransferase bifunctional protein ThrH
103	78901	80043	1142	cell division protein FtsW
103	95089	95457	368	hypothetical protein
103	82948	83412	464	ACT domain-containing protein
103	84877	87438	2561	hypothetical protein
103	72307	73098	791	hypothetical protein
103	87415	87759	344	hypothetical protein
104	87568	88944	1376	sugar ABC transporter substrate-binding protein
104	88612	89691	1079	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase TrmU
104	93375	95297	1922	DNA topoisomerase IV subunit B ParE
104	80065	81441	1376	UDP-N-acetylmuramoylalanine-D-glutamate ligase MurD
104	83551	84057	506	HTH domain-containing protein
104	73123	73527	404	GNAT family acetyltransferase
104	87919	88287	368	hypothetical protein
104	85872	87404	1532	hypothetical protein
105	89269	90282	1013	LacI family transcriptional regulator
105	82904	83293	389	hypothetical protein
105	89932	90957	1025	LacI family transcriptional regulator
105	95333	96202	869	degV family protein
105	109321	110712	1391	tryptophan synthase beta subunit TrpB
105	81473	82438	965	phospho-N-acetylmuramoyl-pentapeptide-transferase MraY
105	95992	96354	362	response regulator domain-containing protein
105	87491	87970	479	hypothetical protein
105	73604	74557	953	alpha/beta fold family hydrolase
105	89082	89726	644	hypothetical protein
105	87414	88706	1292	hypothetical protein

106	84309	86333	2024	glycoside hydrolase family 31 Gh31C
106	90480	91454	974	hypothetical protein
106	83829	83996	167	hypothetical protein
106	90963	92381	1418	glucuronate isomerase UxaC
106	96273	98342	2069	ATP-dependent DNA helicase RecG
106	110768	111631	863	PHP domain-containing protein
106	82457	84715	2258	cell division protein FtsI
106	96445	97068	623	hypothetical protein
106	87973	89154	1181	cell wall hydrolase
106	90059	90436	377	hypothetical protein
106	88850	89980	1130	hypothetical protein
107	91552	93213	1661	sucrose phosphorylase SucI3P
107	84044	84334	290	hypothetical protein
107	84763	85203	440	cell division protein FtsL
107	97101	97412	311	hypothetical protein
107	86691	88250	1559	sugar ABC transporter substrate-binding protein
107	89167	89649	482	deoxycytidylate deaminase
107	74645	76681	2036	methionyl-tRNA synthetase MetG
107	90552	91721	1169	hypothetical protein
107	90076	91161	1085	hypothetical protein
108	84544	85566	1022	hypothetical protein
108	92570	94072	1502	altronate hydrolase UxaA
108	98411	100180	1769	dihydroxyacetone kinase family protein
108	111757	112725	968	hypothetical protein
108	85236	86171	935	S-adenosyl-methyltransferase MraW
108	97705	98112	407	hypothetical protein
108	88341	89285	944	sugar ABC transporter permease
108	76708	77598	890	methionine aminopeptidase Map
108	91854	92789	935	HAMP/SpoIIE domain-containing protein
108	91163	91804	641	hypothetical protein
109	85579	88236	2657	McrBC restriction endonuclease system protein McrB
109	94243	95091	848	xylose isomerase domain-containing protein
109	100217	100579	362	hypothetical protein
109	86181	86621	440	MraZ protein
109	98144	98860	716	endonuclease V
109	89314	90123	809	sugar ABC transporter permease
109	89674	90303	629	uracil phosphoribosyltransferase upp
109	77633	79165	1532	5'-nucleotidase domain-containing protein
109	91829	93358	1529	polysaccharide biosynthesis protein
110	88249	89556	1307	McrBC restriction endonuclease system protein McrC
110	100788	100973	185	ribosomal protein L28 RpmB
110	112730	114235	1505	hypothetical protein
110	90240	91262	1022	LacI family transcriptional regulator
110	90320	90748	428	ribose 5-phosphate isomerase RpiB
110	79170	79994	824	phosphonate ABC transporter permease protein PhnE
110	92786	93430	644	SpoIIE domain-containing protein
110	93366	94400	1034	acyltransferase
111	89596	90342	746	hypothetical protein
111	95177	95950	773	hypothetical protein
111	101028	101444	416	hypothetical protein
111	90788	91846	1058	Sua5/YciO/YrdC/YwlC family protein
111	79997	80818	821	phosphonate ABC transporter permease protein PhnE
112	91851	93554	1703	chitinase Chi18A
112	90476	91747	1271	ATP-dependent metallopeptidase HfIB
112	95990	98113	2123	hypothetical protein
112	114270	115088	818	HAD superfamily hydrolase
112	98853	99302	449	hypothetical protein
112	80824	81597	773	phosphonate ABC transporter ATP-binding protein phnC
112	93645	93812	167	hypothetical protein
113	91649	94093	2444	cellobiose phosphorylase Cbp94A
113	91821	92396	575	hypothetical protein
113	98167	99537	1370	hypothetical protein
113	101446	102213	767	RNA polymerase sigma factor sigma-70 family protein
113	115090	115344	254	hypothetical protein
113	94256	95056	800	MecA family protein
113	81747	82856	1109	phosphonate ABC transporter substrate-binding protein
113	93879	94586	707	purine nucleoside phosphorylase DeoD
113	94402	96297	1895	hypothetical protein
114	93926	95839	1913	hypothetical protein
114	92412	93119	707	hypothetical protein
114	99518	100714	1196	acetate kinase AckA
114	102226	103104	878	hypothetical protein
114	115360	115944	584	septum formation protein Maf
114	99452	100450	998	septum site-determining protein MinD
114	94222	95001	779	metallopeptidase M23 family protein
114	94604	95470	866	hypothetical protein
114	96281	97669	1388	hypothetical protein
115	96107	97171	1064	hypothetical protein
115	93217	93960	743	hypothetical protein
115	100886	101746	860	AraC family transcriptional regulator
115	103122	105857	2735	two component system histidine kinase/response regulator hybrid protein
115	115962	118112	2150	polyphosphate kinase Ppk
115	100468	101157	689	hypothetical protein

115	95018	95473	455	hypothetical protein
115	95109	98639	3530	pyruvate:ferredoxin oxidoreductase
115	83051	83707	656	ribulose-phosphate 3-epimerase Rpe
115	95529	96830	1301	seryl-tRNA synthetase SerS
115	97672	98160	488	hypothetical protein
116	101783	102901	1118	acetyl-xylan esterase Est2A
116	93962	94948	986	hypothetical protein
116	101276	105826	4550	FtsK/SpoIIIE family protein
116	99063	99233	170	ferredoxin
116	83707	84648	941	transketolase subunit B Tkt
116	96880	97401	521	hypothetical protein
116	98160	99515	1355	hypothetical protein
117	95079	95651	572	hypothetical protein
117	102919	103377	458	thioesterase superfamily protein
117	105956	108328	2372	HD-GYP domain-containing protein
117	118137	119684	1547	exopolyphosphatase Ppx
117	105840	106166	326	hypothetical protein
117	84669	85499	830	transketolase subunit A Tkt
117	97415	98317	902	chromosome partitioning protein ParB
118	95754	96566	812	ADP-ribosylglycohydrolase family protein
118	103569	103685	116	ribosomal protein L32 RpmF
118	119702	120850	1148	chorismate mutase/prephenate dehydratase PheA
118	106179	106469	290	hypothetical protein
118	95695	95979	284	TrpR-like protein YerC/YecD
118	99303	100646	1343	replicative DNA helicase dnaB
118	85705	86400	695	L-ribulose-5-phosphate 4-epimerase AraD
118	98331	99110	779	chromosome partitioning protein ParA
119	97837	98826	989	Lacl family transcriptional regulator
119	96660	97958	1298	hypothetical protein
119	103857	104867	1010	fatty acid/phospholipid synthesis protein PlsX
119	108338	109111	773	response regulator domain-containing protein
119	120933	121937	1004	hypothetical protein
119	106484	106867	383	hypothetical protein
119	96254	96454	200	cold shock domain protein CspD
119	100650	101096	446	ribosomal protein L9 rplI
119	86422	88017	1595	L-ribulokinase AraB
119	99399	100136	737	methyltransferase GidB
119	99533	101005	1472	MBOAT family acyltransferase
120	98957	101266	2309	glycogen phosphorylase GlgP
120	104920	105156	236	acyl carrier protein
120	109124	113338	4214	two component system histidine kinase/response regulator hybrid protein
120	106914	107249	335	hypothetical protein
120	96624	97397	773	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase HisA
120	101100	103181	2081	DHH domain-containing protein
120	88004	89116	1112	GntR family transcriptional regulator
120	100126	102042	1916	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
120	101005	102453	1448	hypothetical protein
121	105234	105947	713	ribonuclease III Rnc
121	121979	122935	956	aspartate carbamoyltransferase PyrB
121	107252	107551	299	hypothetical protein
121	97440	98195	755	dihydrodipicolinate reductase DapB
121	89264	90103	839	8-oxoguanine DNA glycosylase Ogg
121	102044	103465	1421	tRNA modification GTPase TrmE
122	102269	103993	1724	hypothetical protein
122	98802	99014	212	HTH domain-containing protein
122	105928	109485	3557	chromosome segregation protein Smc
122	113351	115486	2135	HD-GYP domain-containing protein
122	122963	123406	443	aspartate carbamoyltransferase regulatory subunit PyrI
122	107616	109055	1439	hypothetical protein
122	98197	99087	890	dihydrodipicolinate synthase dapA
122	90100	90558	458	restriction endonuclease family protein
122	103537	104262	725	single stranded nucleic acid binding protein
123	104148	106004	1856	GGDEF/cache domain-containing protein
123	99008	99445	437	hypothetical protein
123	123484	124524	1040	S-adenosylmethionine:tRNA ribosyltransferase-isomerase QueA
123	99439	100713	1274	hypothetical protein
123	90661	91770	1109	hypothetical protein
123	104286	105590	1304	membrane insertion protein
123	102462	103991	1529	hypothetical protein
124	109485	110414	929	signal recognition particle-docking protein FtsY
124	124529	126028	1499	peptidase M18 family protein
124	109124	109375	251	hypothetical protein
124	100728	102182	1454	peptidase U62 family protein
124	103211	106078	2867	hypothetical protein
125	106087	106692	605	TetR family transcriptional regulator
125	99662	99826	164	hypothetical protein
125	110433	111659	1226	threonine dehydratase IlvA
125	115714	119961	4247	CoA-substrate-specific enzyme activase
125	126171	128594	2423	phenylalanyl-tRNA synthetase beta subunit PheT
125	109683	110180	497	hypothetical protein
125	102369	104201	1832	GTP-binding protein TypA
125	106125	107024	899	sugar ABC transporter permease
125	91813	92505	692	hypothetical protein

125	106224	106358	134	ribosomal protein L34 RpmH
125	104026	105231	1205	hypothetical protein
126	106903	109017	2114	RND family transporter
126	99899	100528	629	abortive infection bacteriophage resistance protein
126	111662	112540	878	hypothetical protein
126	120009	120491	482	hypothetical protein
126	128632	129516	884	phenylalanyl-tRNA synthetase alpha subunit PheS
126	110238	110690	452	hypothetical protein
126	104268	104930	662	hypothetical protein
126	107040	107981	941	sugar ABC transporter permease
126	92533	92967	434	AbrB family transcriptional regulator
127	112543	113631	1088	lysozyme Lyc25A
127	120567	121391	824	polysaccharide deacetylase Est4D
127	109068	111794	2726	cell surface protein
127	129513	129656	143	hypothetical protein
127	110738	111241	503	hypothetical protein
127	105127	107730	2603	penicillin-binding protein 1A family Pbp1A
127	108193	109899	1706	sugar ABC transporter substrate binding protein
127	106900	108300	1400	chromosomal replication initiator protein DnaA
127	105250	106473	1223	hypothetical protein
128	100968	101321	353	hypothetical protein
128	121406	122371	965	homoserine O-succinyltransferase MetA
128	111365	111868	503	hypothetical protein
128	110040	111059	1019	transcriptional regulator AraC family
128	92964	94040	1076	acyltransferase
128	108592	109701	1109	DNA polymerase III beta subunit DnaN
128	106531	107046	515	hypothetical protein
129	101446	102393	947	hypothetical protein
129	113636	114481	845	AraC family transcriptional regulator
129	122394	124394	2000	NAD-dependent DNA ligase LigA
129	111892	112290	398	hypothetical protein
129	108042	109406	1364	MATE efflux family protein
129	111118	112926	1808	two component system histidine kinase
129	94194	95873	1679	Site-specific recombinases DNA invertase Pin homologs
129	109725	109940	215	S4 domain protein YaaA
130	111859	114432	2573	flagellin FlhC
130	124457	125170	713	RNA pseudouridylate synthase
130	112307	112777	470	hypothetical protein
130	110515	110997	482	metallophosphoesterase
130	113068	113331	263	ribosomal protein S18 rpsR
130	95890	96960	1070	hypothetical protein
130	109937	111040	1103	DNA replication and repair protein RecF
130	107031	107594	563	hypothetical protein
131	114578	115603	1025	hypothetical protein
131	102837	103556	719	hypothetical protein
131	114562	114912	350	HTH domain-containing protein
131	125241	126119	878	hypothetical protein
131	112796	113350	554	hypothetical protein
131	111019	111639	620	non-canonical purine NTP pyrophosphatase rdgB/HAM1 family
131	97158	97511	353	HTH domain-containing protein
131	111044	112993	1949	DNA gyrase subunit B GyrB
131	107591	107920	329	hypothetical protein
132	114912	116282	1370	signal recognition particle protein Ffh
132	126175	127185	1010	hydrogenase expression/formation protein HypE
132	113373	113987	614	hypothetical protein
132	111636	112607	971	N-acetylmuramoyl-L-alanine amidase
132	113349	113804	455	single-stranded DNA binding protein ssb
132	97566	97778	212	hypothetical protein
132	108031	109029	998	glycosyl transferase GT2 family protein
133	115631	117037	1406	aldehyde dehydrogenase family protein
133	103791	104399	608	hypothetical protein
133	127199	128245	1046	hydrogenase expression/formation protein HypD
133	112648	113802	1154	RNA methylase family protein
133	113824	114111	287	ribosomal protein S6 rpsF
133	97796	98242	446	hypothetical protein
133	113012	115570	2558	DNA gyrase subunit A GyrA
133	109076	110314	1238	nucleotide sugar dehydrogenase
134	117071	118186	1115	N-acetylglucosamine-6-phosphate deacetylase NagA
134	117071	118186	1115	N-acetylglucosamine-6-phosphate deacetylase NagA
134	104525	104878	353	hypothetical protein
134	116362	116607	245	ribosomal protein S16 RpsP
134	128245	128463	218	hydrogenase assembly chaperone HypC
134	113987	117571	3584	hypothetical protein
134	114010	115149	1139	HD-GYP/response regulator domain-containing protein
134	114394	115152	758	SIR2 family protein
134	98414	98743	329	hypothetical protein
134	115578	117011	1433	GGDEF domain-containing protein
134	110367	111137	770	hypothetical protein
135	118276	119661	1385	extradiol ring-cleavage dioxygenase class III protein subunit B
135	104882	105172	290	hypothetical protein
135	116652	116879	227	RNA binding protein
135	128465	130744	2279	hydrogenase maturation protein HypF
135	117790	117975	185	hypothetical protein

135	115181	116362	1181	thiolase ThIA
135	117059	117772	713	exodeoxyribonuclease III (xth) xth
136	119781	120641	860	radical SAM domain-containing protein
136	116972	117682	710	tRNA (guanine-N1)-methyltransferase TrmD
136	130744	131247	503	ech hydrogenase subunit F EchF
136	117991	118992	1001	hypothetical protein
136	116498	117244	746	cobalt ABC transporter ATP-binding protein CbiO
136	115271	116983	1712	sugar ABC transporter substrate-binding protein
136	99351	100436	1085	NUDIX domain-containing protein
136	117831	118757	926	alpha/beta fold family hydrolase
136	111145	112626	1481	polysaccharide biosynthesis protein
137	118914	120428	1514	alpha-L-arabinofuranosidase Arf51B
137	117686	118198	512	16S rRNA processing protein RimM
137	131273	132277	1004	ech hydrogenase subunit E EchE
137	119009	119467	458	hypothetical protein
137	117241	118014	773	cobalt ABC transporter permease CbiQ
137	117116	118045	929	sugar ABC transporter permease
137	112705	114585	1880	asparagine synthase glutamine-hydrolyzing AsnB
138	120967	122502	1535	alpha-amylase Amy13G
138	118289	119065	776	uridine phosphorylase Udp
138	132368	132700	332	ech hydrogenase subunit D EchD
138	119521	119871	350	hypothetical protein
138	118030	119034	1004	cobalamin biosynthesis protein CbiM
138	118059	119063	1004	sugar ABC transporter permease
138	100674	101408	734	metallo-beta-lactamase family protein
138	120479	121006	527	GNAT family acetyltransferase
138	114599	115783	1184	glycosyl transferase GT4 family protein
139	106108	106668	560	TetR family transcriptional regulator
139	132700	133383	683	ech hydrogenase subunit C EchC
139	119864	120931	1067	hypothetical protein
139	119201	120286	1085	methylthioribose-1-phosphate isomerase MtnA
139	119257	120255	998	AraC family transcriptional regulator
139	101412	102086	674	HAD superfamily hydrolase
139	121019	121786	767	phosphorylase Pnp/Udp family protein
140	123167	124276	1109	arabinogalactan endo-1,4-beta-galactosidase Agn53B
140	120387	121436	1049	unsaturated rhamnogalacturonyl hydrolase Gh105A
140	119114	120304	1190	Bmp family protein
140	133407	134276	869	ech hydrogenase subunit B EchB
140	120296	120490	194	hypothetical protein
140	115785	116987	1202	glycosyl transferase GT4 family protein
141	121467	123635	2168	lacto-N-biose phosphorylase
141	124533	125240	707	IMP cyclohydrolase purO
141	106803	108296	1493	radical SAM domain-containing protein
141	120458	121999	1541	ABC transporter ATP-binding protein
141	120955	121464	509	hypothetical protein
141	120581	121630	1049	electron transfer flavoprotein alpha subunit EtfA
141	102095	103408	1313	hypothetical protein
141	116984	117916	932	glycosyl transferase
142	125273	126451	1178	5-aminoimidazole-4-carboxamide ribonucleotide transformylase
142	108364	108606	242	hypothetical protein
142	134276	136123	1847	ech hydrogenase subunit A EchA
142	121486	122499	1013	hypothetical protein
142	121659	122450	791	electron transfer flavoprotein beta subunit EtfB
142	103601	105097	1496	23S rRNA methyltransferase
142	122110	123486	1376	MatE efflux family protein
142	117919	119385	1466	hypothetical protein
143	122134	123141	1007	ABC transporter permease
143	136255	136557	302	hypothetical protein
143	123007	123462	455	hypothetical protein
143	122487	123662	1175	butyryl-CoA dehydrogenase Bed
143	123706	124614	908	S1 RNA binding domain-containing protein
143	123523	124641	1118	citrate transporter family protein
144	127041	127604	563	hypothetical protein
144	108944	109510	566	acyl carrier protein phosphodiesterase AcpD
144	136686	137129	443	MarR family transcriptional regulator
144	123481	124269	788	hypothetical protein
144	124637	125635	998	hypothetical protein
144	105113	105607	494	rRNA methylase SpoU family protein
144	124675	125394	719	5'/3'-nucleotidase
144	119389	120273	884	glycosyl transferase
145	127635	128180	545	hypothetical protein
145	109552	110502	950	hypothetical protein
145	123252	124082	830	ABC transporter permease
145	137219	137497	278	septum site-determining protein MinE
145	124347	124853	506	hypothetical protein
145	123674	124549	875	3-hydroxybutyryl-CoA dehydrogenase Hbd
145	105634	107070	1436	two component system histidine kinase
145	125563	126588	1025	hypothetical protein
145	120335	121855	1520	MBOAT family acyltransferase
146	128299	129333	1034	GntR family transcriptional regulator
146	124088	124510	422	cytidine deaminase Cdd
146	137513	138292	779	septum site-determining protein MinD
146	124589	125383	794	crotonase Crt

146	126599	127483	884	hypothetical protein
147	129411	129734	323	hypothetical protein
147	110667	111137	470	hypothetical protein
147	124614	124964	350	ribosomal protein L19 RplS
147	138305	139066	761	septum site-determining protein MinC
147	125136	125639	503	hypothetical protein
147	125582	126787	1205	thiolase ThIA
147	127260	127685	425	hypothetical protein
147	107071	107772	701	two component system response regulator
147	127526	128227	701	hypothetical protein
147	121880	122614	734	hypothetical protein
148	129886	132018	2132	GGDEF/EAL domain-containing protein
148	139226	140179	953	translation elongation factor Ts Tsf
148	126960	127490	530	ribosomal subunit interface protein
148	122601	122990	389	hypothetical protein
149	111463	112119	656	PAP2 family protein
149	125035	125631	596	signal peptidase I LepB
149	140256	140990	734	ribosomal protein S2 RpsB
149	126028	126315	287	hypothetical protein
149	107789	110188	2399	MutS2 family protein
149	128250	128855	605	hypothetical protein
149	123028	123267	239	hypothetical protein
150	132116	135592	3476	ATP-dependent nuclease subunit B addB
150	112126	112746	620	hypothetical protein
150	125696	126523	827	GTP-binding protein
150	126455	129091	2636	hypothetical protein
150	128879	129532	653	hypothetical protein
150	123270	124823	1553	hydrolase HAD superfamily
151	127828	129336	1508	alpha-L-arabinofuranosidase Arf51A
151	135622	139311	3689	ATP-dependent nuclease subunit A addA
151	112767	113396	629	GNAT family acetyltransferase
151	126556	127089	533	signal peptidase I LepB
151	141310	142065	755	hypothetical protein
151	110348	111712	1364	serine protease HtrA family protein
151	129569	131080	1511	MBOAT family acyltransferase
151	125070	126137	1067	hypothetical protein
152	139298	139999	701	proton-coupled thiamine transporter YuaJ
152	113456	114334	878	GNAT family acetyltransferase
152	142081	143079	998	hypothetical protein
152	129351	130286	935	sugar ABC transporter permease
152	111741	112685	944	HD domain-containing protein
152	126436	127572	1136	glycosyl transferase GT4 family protein
153	140235	141575	1340	GTP-binding protein HflX
153	114453	115865	1412	MatE efflux family protein
153	127089	127850	761	ribonuclease HII RnhB
153	129381	131411	2030	ABC transporter ATP-binding/permease
153	112738	113895	1157	aspartate/tyrosine/aromatic aminotransferase
153	131090	132184	1094	hypothetical protein
153	127573	128538	965	glycosyl transferase GT2 family protein
154	141628	142638	1010	aspartate--ammonia ligase asnA
154	115923	116903	980	aminoglycoside phosphotransferase
154	127892	129511	1619	hypothetical protein
154	143152	143541	389	hypothetical protein
154	131413	133095	1682	ABC transporter ATP-binding/permease
154	130273	131196	923	sugar ABC transporter permease
154	113892	114380	488	AsnC family transcriptional regulator
154	132250	132498	248	hypothetical protein
155	142663	142953	290	hypothetical protein
155	129526	129810	284	FlhB domain-containing protein
155	143584	145161	1577	hypothetical protein
155	133180	133650	470	MarR family transcriptional regulator
155	131267	132925	1658	sugar ABC transporter substrate-binding protein
155	114420	115394	974	thiamine-monophosphate kinase ThiL
155	132523	133605	1082	undecaprenyldiphospho- muramoylpentapeptide beta-N-acetylglucosaminyltransferase MurG
155	128576	130150	1574	hypothetical protein
156	117256	118746	1490	hypothetical protein
156	145195	145968	773	flagellar biosynthesis sigma factor FliA
156	115434	116993	1559	VanW-like family protein
156	133668	134147	479	S-ribosylhomocysteinase LuxS
156	130176	131204	1028	hypothetical protein
157	143059	145512	2453	Rad3-related DNA helicase
157	118730	119503	773	hypothetical protein
157	130218	131090	872	hypothetical protein
157	145995	146330	335	hypothetical protein
157	133118	134146	1028	LacI family transcriptional regulator
157	134156	135022	866	HTH/SIS domain-containing protein
157	131201	132475	1274	polysaccharide ABC transporter ATP-binding/permease
158	146032	147603	1571	GGDEF/TPR repeat domain-containing protein
158	131154	131876	722	two component system response regulator
158	146327	146812	485	chemotaxis protein CheD
158	135165	135596	431	hypothetical protein
159	148000	150672	2672	valyl-tRNA synthetase valS

159	131990	134335	2345	two component system histidine kinase
159	146827	147459	632	chemotaxis protein CheC family protein
159	133817	134584	767	tryptophan synthase alpha subunit TrpA
159	117117	117719	602	hypothetical protein
159	135612	136190	578	rubrerythrin Rbr1
159	132518	133369	851	polysaccharide ABC transporter permease
160	119595	120935	1340	MFS transporter
160	134402	135601	1199	ribose-phosphate pyrophosphokinase PrsA
160	147474	147968	494	chemotaxis protein CheW family protein
160	134577	135767	1190	tryptophan synthase beta subunit TrpB
160	134399	135676	1277	MFS transporter
161	120961	122034	1073	HTH domain-containing protein
161	135737	137101	1364	Fe-S oxidoreductase
161	135808	136440	632	phosphoribosylanthranilate isomerase TrpF
161	136326	137021	695	MTA/SAH nucleosidase
161	133429	134568	1139	hypothetical protein
162	150978	152363	1385	tetrahydrofolate biosynthesis FoC
162	137105	138433	1328	GTP-binding protein
162	148005	150116	2111	chemotaxis histidine kinase CheA
162	136444	137229	785	indole-3-glycerol phosphate synthase TrpC
162	117777	119282	1505	galactose-1-phosphate uridylyltransferase GalT
162	134578	136023	1445	MBOAT family acyltransferase
163	152380	152544	164	hypothetical protein
163	122399	123403	1004	ribosome small subunit-dependent GTPase A RsgA
163	138438	139139	701	hypothetical protein
163	150256	151326	1070	chemotaxis-specific methyltransferase CheB
163	137226	138266	1040	anthranilate phosphoribosyltransferase trpD
163	135928	136512	584	sugar-phosphate isomerase LacAB/RpiB family
163	119337	120536	1199	galactokinase GalK
163	137018	139270	2252	hypothetical protein
163	136196	137293	1097	hypothetical protein
164	123462	124016	554	HAD superfamily hydrolase
164	139153	140175	1022	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase GpsA
164	151329	152051	722	flagellar protein
164	136531	136689	158	hypothetical protein
164	120695	121858	1163	toxic anion resistance family protein
164	139338	140918	1580	glucose-6-phosphate isomerase Gpi
165	141083	142021	938	xylosidase/arabinofuranosidase Xsa43G
165	152547	153965	1418	TPR domain-containing protein
165	124074	124514	440	LytTR domain-containing protein
165	152087	153019	932	flagellar biosynthesis protein FlhG
165	138819	141635	2816	NADPH-dependent glutamate synthase GltA
165	136857	137021	164	hypothetical protein
165	137478	137705	227	hypothetical protein
166	154161	155357	1196	MBOAT family acyltransferase
166	124516	124980	464	hypothetical protein
166	153044	154303	1259	flagellar biosynthesis protein FlhF
166	141706	142842	1136	hypothetical protein
166	137154	138032	878	acetyltransferase GNAT family
166	121902	123128	1226	hypothetical protein
167	155459	156313	854	GDLSL-family lipase/acylhydrolase
167	125077	125775	698	ABC transporter ATP-binding protein
167	142852	143472	620	chemotaxis protein CheC family protein
167	138084	139421	1337	NLPC/P60 domain-containing protein
167	123298	124203	905	AraC family transcriptional regulator
167	142110	143621	1511	two component system histidine kinase
167	138008	139519	1511	BCCT family transporter
168	124195	126669	2474	glycogen phosphorylase GlgP
168	156331	156576	245	hypothetical protein
168	125853	126569	716	ABC transporter permease
168	154305	156362	2057	flagellar biosynthesis protein FlhA
168	143751	144497	746	DeoR family transcriptional regulator
168	143662	144333	671	two component system response regulator
168	139521	140243	722	cholinephosphate cytidyltransferase
169	156601	156978	377	endoribonuclease L-PSP
169	126718	127119	401	hypothetical protein
169	156403	157557	1154	flagellar biosynthetic protein FlhB
169	140236	142083	1847	choline kinase
170	156991	158040	1049	TPR domain-containing protein
170	157532	158317	785	flagellar biosynthetic protein FlhR
170	144512	145417	905	1-phosphofructokinase pfkB
170	139526	140464	938	thioredoxin-disulfide reductase trxB
170	144449	145138	689	RNA pseudouridylyltransferase
171	158061	158858	797	hypothetical protein
171	127242	128309	1067	sugar ABC transporter substrate-binding protein
171	158327	158599	272	flagellar biosynthetic protein FlhQ
171	142080	143000	920	hypothetical protein
172	158865	160778	1913	hypothetical protein
172	128425	129780	1355	two component system histidine kinase
172	158611	159546	935	flagellar biosynthetic protein FlhP
172	145433	147523	2090	PTS system IIABC fructose-specific family protein
172	140562	141338	776	hydrolase HAD superfamily
172	145156	146628	1472	hypothetical protein

172	143020	144246	1226	glycosyl transferase
173	160809	161966	1157	unsaturated glucuronyl hydrolase Ugl88A
173	159548	159886	338	flagellar protein FlIZ
173	147584	147847	263	TS system HPr phosphocarrier
173	141338	142060	722	3-oxoacyl-(acyl-carrier-protein) reductase FabG
173	146603	147088	485	hypothetical protein
173	144252	145394	1142	hypothetical protein
174	129797	131392	1595	two component system response regulator
174	159898	161133	1235	flagellar motor switch protein FljY
174	147860	149566	1706	PTS system I PEP-phosphotransferase
175	161966	164032	2066	hypothetical protein
175	149632	149757	125	hypothetical protein
175	142170	142523	353	response regulator domain-containing protein
175	147177	148481	1304	two component system histidine kinase
175	145425	146546	1121	glycosyl transferase GT4 family protein
176	164029	164577	548	GNAT family acetyltransferase
176	131413	132456	1043	xylose ABC transporter substrate binding protein
176	161137	162081	944	flagellar motor switch protein FljM
176	149941	150852	911	hypothetical protein
176	142537	143421	884	hypothetical protein
176	148509	149183	674	two component system response regulator
176	146598	147629	1031	hypothetical protein
177	164555	165256	701	two component system response regulator
177	132642	133814	1172	xylose ABC transporter substrate-binding protein
177	162169	162705	536	hypothetical protein
177	151005	152339	1334	Glu/Leu/Phe/Val dehydrogenase
178	133912	135465	1553	xylose ABC transporter ATP-binding protein
178	162725	163558	833	flagellar motor protein MotB
178	144013	144261	248	hypothetical protein
178	147637	149514	1877	hypothetical protein
179	165470	166582	1112	two component system histidine kinase
179	135458	136648	1190	xylose ABC transporter permease
179	144281	144613	332	hypothetical protein
179	149601	150515	914	hypothetical protein
180	166714	167700	986	GGDEF domain-containing protein
180	136896	138590	1694	hypothetical protein
180	144628	145086	458	GNAT family acetyltransferase
180	150546	151304	758	hypothetical protein
180	149514	151211	1697	hypothetical protein
181	163558	164223	665	flagellar motor protein MotA
181	152534	153847	1313	cell wall hydrolase
181	151221	152363	1142	acyltransferase
182	168664	169443	779	hypothetical protein
182	138781	141048	2267	oligopeptide ABC transporter substrate-binding protein OppA
182	164427	164600	173	flagellar protein FljD
182	145251	146582	1331	amino acid carrier protein AGCS family
183	169549	169980	431	peptide methionine sulfoxide reductase MsrA
183	141176	142759	1583	oligopeptide ABC transporter permease OppB
183	153937	155052	1115	GFO/IDH/MOCA family oxidoreductase
183	146776	147516	740	SAM-dependent methyltransferase
183	151500	151793	293	hypothetical protein
183	152602	153768	1166	hypothetical protein
184	170056	172131	2075	methyl-accepting chemotaxis protein McpA
184	142785	143762	977	oligopeptide ABC transporter permease OppC
184	165172	165906	734	Ion channel.
184	155079	156365	1286	bacterial SH3 domain-containing protein
185	172330	175278	2948	beta-glucosidase Bgl3D
185	143775	145541	1766	oligopeptide ABC transporter ATP-binding protein OppD
185	165965	169225	3260	flagellar hook protein FlgE
185	156584	157267	683	hypothetical protein
186	145534	146835	1301	oligopeptide ABC transporter ATP-binding protein OppF
186	169340	169774	434	flagellar operon protein
186	157267	158151	884	hypothetical protein
186	147960	148490	530	GDSL-family lipase/acylhydrolase
187	175350	177776	2426	beta-glucosidase Bgl3E
187	169842	170540	698	flagellar hook capping protein FlgD
187	148746	149942	1196	phosphopantothienoylcysteine decarboxylase CoaBC
187	151893	152348	455	ferric uptake regulator family protein
188	177806	178363	557	hypothetical protein
188	146913	147653	740	hypothetical protein
188	158277	158834	557	translation elongation factor P efp
188	149955	150476	521	Baf family transcriptional regulator
188	152362	153258	896	hypothetical protein
189	178517	179257	740	phosphoribulokinase/uridine kinase family protein
189	170563	172081	1518	flagellar hook length control protein FljK
189	158959	159201	242	hypothetical protein
189	150748	151398	650	hypothetical protein
189	155274	156482	1208	hypothetical protein
190	179276	180787	1511	DHH domain-containing protein
190	147829	148080	251	hypothetical protein
190	172150	173088	938	hypothetical protein
190	159215	159454	239	hypothetical protein
190	153553	154314	761	FeS assembly ATPase SufC

190	156502	157182	680	hypothetical protein
191	159637	161877	2240	Polysaccharide lyase Pli1A
191	148170	148319	149	hypothetical protein
191	151531	151857	326	hypothetical protein
191	154329	155747	1418	FeS assembly protein SufB
192	148547	150013	1466	methyl-accepting chemotaxis protein McpH
192	173091	173549	458	flagellar export protein FliJ
192	151893	152210	317	hypothetical protein
192	155760	156929	1169	FeS assembly protein SufD
192	157179	157922	743	hypothetical protein
193	161896	162630	734	carbohydrate esterase family 12 Est12B
193	180833	183967	3134	hypothetical protein
193	150247	152316	2069	methyl-accepting chemotaxis protein McpG
193	173553	174911	1358	flagellar protein export ATPase FliI
193	152247	152948	701	hypothetical protein
193	156949	158169	1220	cysteine desulfurase SufS
194	174924	175778	854	flagellar assembly protein FliH
194	162635	165220	2585	hypothetical protein
194	152962	153300	338	hypothetical protein
194	158159	158602	443	FeS assembly protein SufE
194	158103	159356	1253	hypothetical protein
195	184125	184760	635	endonuclease III nth
195	152397	153464	1067	glucose-1-phosphate adenylyltransferase GlgD
195	175798	176847	1049	flagellar motor switch protein FliG
195	165292	166872	1580	sugar ABC transporter permease
195	153455	154081	626	HAD superfamily hydrolase
195	159078	159854	776	TatD family hydrolase
195	159422	160540	1118	hypothetical protein
196	153512	154654	1142	glucose-1-phosphate adenylyltransferase GlgC
196	176860	178479	1619	flagellar M-ring protein FliF
196	154140	155444	1304	hypothetical protein
196	159855	160715	860	dimethyladenosine transferase KsgA
196	160694	161500	806	hypothetical protein
197	184854	185087	233	hypothetical protein
197	154762	155679	917	hypothetical protein
197	178511	178855	344	flagellar hook-basal body complex protein FliE
197	166939	167853	914	sugar ABC transporter permease
197	155489	156787	1298	peptidase M18 family protein
198	185491	185754	263	hypothetical protein
198	155779	157005	1226	HDOD domain-containing protein
198	167871	168791	920	sugar ABC transporter permease
198	161481	162629	1148	acyltransferase
199	185959	186855	896	hypothetical protein
199	157016	157711	695	4-phosphopantetheinyl transferase
199	178879	179328	449	flagellar basal-body rod protein FlgC
199	168957	171212	2255	AraC family transcriptional regulator
199	156888	158156	1268	Na ⁺ /H ⁺ antiporter
199	160775	163630	2855	hypothetical protein
199	162677	163204	527	hypothetical protein
200	157736	158767	1031	hypothetical protein
200	179344	179736	392	flagellar basal-body rod protein FlgB
200	163208	164581	1373	hypothetical protein
201	163662	165653	1991	1,4-alpha-glucan branching enzyme GlgB
201	187489	189918	2429	leucyl-tRNA synthetase leuS
201	158883	159437	554	phosphoglycerate mutase family protein
201	179944	180723	779	GTP-sensing transcriptional pleiotropic repressor CodY
201	171407	171841	434	3-dehydroquinate dehydratase type II aro
202	190339	194061	3722	hypothetical protein
202	180859	182973	2114	DNA topoisomerase I TopA
202	158452	158730	278	ribonuclease inhibitor
202	165744	168866	3122	hypothetical protein
202	164843	165277	434	hypothetical protein
203	194076	194885	809	serine/threonine protein phosphatase
203	183106	185973	2867	FAD-dependent pyridine nucleotide-disulphide oxidoreductase
203	171857	172519	662	hypothetical protein
203	158730	159863	1133	hypothetical protein
203	168899	170062	1163	flagellin FliC
203	165291	165848	557	hypothetical protein
204	194982	196580	1598	serine/threonine protein kinase
204	159616	161400	1784	hypothetical protein
204	172542	175808	3266	carbamoyl-phosphate synthase large subunit carB
204	159872	160336	464	hypothetical protein
204	170081	170557	476	flagellar protein FliS
205	196598	197017	419	hypothetical protein
205	186002	186346	344	hydrogenase nickel insertion protein HypA
205	175808	177019	1211	carbamoyl-phosphate synthase small subunit carA
205	166281	166925	644	flavin reductase domain-containing protein
206	161481	162419	938	hypothetical protein
206	186467	187135	668	hydrogenase accessory protein HypB
206	177046	180852	3806	phosphoribosylformylglycinamide synthase Pur
206	171077	172	-170905	ATPase
206	167091	167690	599	hypothetical protein
207	197070	198449	1379	GGDEF domain-containing protein

207	162464	162682	218	hypothetical protein
207	187135	188256	1121	hypothetical protein
207	180931	181527	596	TetR family transcriptional regulator
207	160482	167819	7337	cell surface protein
207	172068	173312	1244	Flp pilus assembly protein ATPase TadA
208	198433	200076	1643	GGDEF/EAL/PAS domain-containing protein
208	188430	189845	1415	oxygen-independent coproporphyrinogen III oxidase HemN
208	181619	182086	467	hypothetical protein
208	167873	168088	215	hypothetical protein
208	173422	174015	593	Flp pilus assembly protein TadB
208	167806	169398	1592	hypothetical protein
209	200076	200813	737	hypothetical protein
209	162809	164995	2186	hypothetical protein
209	189865	190527	662	metallo-beta-lactamase family protein
209	174141	175487	1346	Flp pilus assembly protein TadC
209	169438	170403	965	glycosyl transferase GT2 family protein
210	164997	165902	905	xylosidase/arabinofuranosidase Xsa43I
210	183363	185660	2297	mannose 6-phosphate isomerase/glycoside hydrolase family 1 protein
210	201032	201280	248	hypothetical protein
210	168294	170198	1904	thiamine pyrophosphate enzyme
210	175545	175763	218	hypothetical protein
210	170458	170865	407	WxcM-like protein
211	201434	201631	197	hypothetical protein
211	165930	167591	1661	hypothetical protein
211	185683	186828	1145	hypothetical protein
211	175819	177186	1367	hypothetical protein
212	186838	187860	1022	glycoside hydrolase
212	201662	202324	662	hypothetical protein
212	167810	168337	527	hypothetical protein
212	190527	192806	2279	RelA/SpoT family protein
212	170205	170984	779	glucose-1-phosphate cytidylyltransferase RfbF
213	187877	189046	1169	glycoside hydrolase
213	202558	203181	623	hypothetical protein
213	168430	168714	284	hypothetical protein
213	192813	193337	524	adenine phosphoribosyl transferase Apt
213	171027	172178	1151	hypothetical protein
213	170958	172301	1343	glycosyl transferase GT4 family protein
214	189260	190273	1013	LacI family transcriptional regulator
214	177183	178076	893	hypothetical protein
214	172349	173149	800	glycosyl transferase GT4 family protein
215	190305	192497	2192	alpha-galactosidase Aga
215	203327	204097	770	NCAIR mutase-related protein
215	172175	173125	950	xylose isomerase domain-containing protein
215	178090	178521	431	hypothetical protein
215	173580	175196	1616	hypothetical protein
216	204157	204897	740	hypothetical protein
216	168763	169728	965	hypothetical protein
216	192596	193522	926	hypothetical protein
216	173122	174447	1325	hypothetical protein
216	178559	178963	404	hypothetical protein
217	204894	205472	578	ABC transporter ATP-binding protein
217	193599	194507	908	LysR family transcriptional regulator
217	174488	176863	2375	hypothetical protein
217	178963	180273	1310	FHA domain-containing protein
217	175649	176782	1133	glycosyl transferase GT4 family protein
218	180312	181322	1010	hypothetical protein
218	176806	177825	1019	hypothetical protein
219	205491	206960	1469	iron-only hydrogenase
219	170125	171267	1142	GGDEF domain-containing protein
219	194565	196280	1715	methyl-accepting chemotaxis protein Mcp
219	176937	178763	1826	hypothetical protein
219	181335	181997	662	membrane-associated peptidase rhomboid family
220	171407	172393	986	sugar ABC transporter substrate-binding protein
220	196522	197817	1295	adenylosuccinate synthetase pur
220	178883	179518	635	HTH domain-containing protein
220	181997	182407	410	HIT domain-containing protein
220	177838	179250	1412	MBOAT family acyltransferase
221	207378	209984	2606	two component system histidine kinase/response regulator hybrid protein
221	179818	180018	200	hypothetical protein
221	182463	183260	797	fumarate hydratase alpha subunit FumA
222	210014	210784	770	response regulator domain-containing protein
222	172395	173864	1469	two component system histidine kinase
222	179264	180079	815	hypothetical protein
223	210801	211727	926	orotidine 5'-phosphate decarboxylase pyrF
223	173845	175101	1256	two component system response regulator
223	183272	183829	557	fumarate hydratase beta subunit FumB
223	180287	180445	158	rubredoxin Rub
224	211743	212423	680	orotate phosphoribosyltransferase pyrE
224	198166	198663	497	hypothetical protein
224	180083	180946	863	hypothetical protein
224	184369	184617	248	hypothetical protein
225	212450	212602	152	hypothetical protein
225	175241	176224	983	sugar ABC transporter substrate-binding protein

225	181128	182438	1310	glycosyl transferase GT2 family protein
226	212602	212967	365	hypothetical protein
226	176311	177810	1499	sugar ABC transporter ATP-binding protein
226	198806	199735	929	hypothetical protein
226	181066	182361	1295	NAD-dependent epimerase/dehydratase
227	199863	201458	1595	glycoside hydrolase
227	212979	215060	2081	peptidase M29 family protein
227	177810	178880	1070	sugar ABC transporter permease
227	182438	183781	1343	hypothetical protein
227	183082	183618	536	hypothetical protein
228	215136	215636	500	phosphoribosylaminoimidazole carboxylase PurE
228	178877	179902	1025	sugar ABC transporter permease
228	201490	202380	890	HAD superfamily hydrolase
228	182380	183606	1226	glycosyl transferase GT4 family protein
229	215698	216729	1031	phosphoribosylformylglycinamide cyclo-ligase purM
229	179923	180372	449	chemotaxis protein CheW family protein
229	202593	203510	917	ArsR family transcriptional regulator
229	184096	184950	854	hypothetical protein
229	183667	184755	1088	NAD dependent epimerase/dehydratase
230	216729	217373	644	phosphoribosylglycinamide formyltransferase purN
230	180365	180811	446	chemotaxis protein CheW family protein
230	203601	206300	2699	translation elongation factor EF-G-like protein
230	184947	186251	1304	hypothetical protein
231	217394	218677	1283	phosphoribosylamine--glycine ligase purD
231	181197	181676	479	hypothetical protein
231	186273	187418	1145	glycosyl transferase GT2 family protein
231	184801	186183	1382	exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase
232	218674	220185	1511	bacterial SH3 domain-containing protein
232	181694	183706	2012	methyl-accepting chemotaxis protein McpA
232	187512	188630	1118	hypothetical protein
233	220323	221615	1292	hypothetical protein
233	183914	184819	905	LysR family transcriptional regulator
233	206391	208163	1772	methyl-accepting chemotaxis protein Mcp
233	188748	189938	1190	glycosyl transferase
233	186183	187352	1169	UDP-N-acetylglucosamine 2-epimerase
234	208158	209402	1244	6-phosphofructokinase PfkA
234	189972	191072	1100	glycosyl transferase
234	187365	188564	1199	DegT/DnrJ/EryC1/StrS family aminotransferase
235	184865	186025	1160	hypothetical protein
235	209513	210388	875	GDSL-family lipase/acylhydrolase
235	191096	191431	335	hypothetical protein
236	221794	222531	737	hypothetical protein
236	210477	211802	1325	two component system histidine kinase
237	211954	213741	1787	ABC transporter ATP-binding/permease
237	191440	193314	1874	carbamoyl transferase
237	189009	190241	1232	hypothetical protein
238	222644	223021	377	GntR family transcriptional regulator
238	186173	187483	1310	hypothetical protein
238	213758	215539	1781	ABC transporter ATP-binding/permease
238	193311	195113	1802	hypothetical protein
239	223039	225651	2612	ATP-dependent Clp protease ATP-binding subunit ClpC
239	195242	196288	1046	SAM-dependent methyltransferase
239	190706	191500	794	hypothetical protein
240	225741	226151	410	hypothetical protein
240	187549	188334	785	hypothetical protein
240	215734	217422	1688	GGDEF domain-containing protein
240	196325	197383	1058	SAM-dependent methyltransferase
241	217832	218101	269	hypothetical protein
241	197528	198562	1034	hypothetical protein
241	191691	191885	194	hypothetical protein
242	226258	227634	1376	DNA repair protein RadA
242	188509	189216	707	hypothetical protein
242	218200	219141	941	hypothetical protein
242	198794	200362	1568	hypothetical protein
243	189304	189801	497	hypothetical protein
243	200377	201525	1148	CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase
243	192731	193399	668	1-acyl-sn-glycerol-3-phosphate acyltransferase PlsC
244	227827	230277	2450	RND family transporter
244	219668	220552	884	hypothetical protein
244	201630	202388	758	glycerophosphoryl diester phosphodiesterase
244	193532	194566	1034	5'-3' exonuclease C-terminal domain-containing protein
245	190024	190515	491	hypothetical protein
245	220743	221162	419	hypothetical protein
245	202421	203410	989	glycerol-3-phosphate cytidyltransferase
246	190853	193414	2561	GGDEF/EAL domain-containing protein
246	221167	221490	323	hypothetical protein
246	203439	204791	1352	glycerol-3-phosphate cytidyltransferase tagD
247	221564	222199	635	hypothetical protein
247	204814	205791	977	hypothetical protein
247	194699	195481	782	mechanosensitive ion channel protein MscS family
248	230297	233965	3668	GLUG domain-containing protein
248	193597	194328	731	hypothetical protein
248	222217	222708	491	hypothetical protein

248	205802	207508	1706	LicD family protein
249	234051	235922	1871	hypothetical protein
249	194333	194875	542	hypothetical protein
249	222764	223912	1148	malate dehydrogenase Mdh
249	207522	209237	1715	LicD family protein
249	195514	196782	1268	hypothetical protein
250	236154	236990	836	hypothetical protein
250	194877	195365	488	flavodoxin family protein
250	223946	225745	1799	hypothetical protein
250	209256	209975	719	nucleotidyl transferase
251	237259	237951	692	transglutaminase domain-containing protein
251	195389	196006	617	Hypothetical protein
251	225765	227096	1331	recombination factor protein Rar
251	209978	210304	326	hypothetical protein
252	237964	238359	395	hypothetical protein
252	196024	196281	257	hypothetical protein
252	227178	228035	857	radical SAM domain-containing protein
252	210301	211029	728	nucleotidyl transferase
252	197262	198314	1052	hypothetical protein
253	238626	239504	878	sugar ABC transporter substrate-binding protein
253	196283	196843	560	hypothetical protein
253	211026	211658	632	beta-phosphoglucomutase family hydrolase
253	198466	199011	545	thymidine kinase Tdk
254	197241	198068	827	hypothetical protein
254	228167	229609	1442	MatE efflux family protein
254	211648	212298	650	hypothetical protein
254	199288	199746	458	molecular chaperone Hsp20
255	239501	241072	1571	hypothetical protein
255	198072	198428	356	hypothetical protein
255	212397	213344	947	hypothetical protein
255	199853	200059	206	hypothetical protein
256	241398	242756	1358	hypothetical protein
256	198443	198997	554	hypothetical protein
256	229606	230556	950	hypothetical protein
256	213339	214346	1007	glycosyl transferase GT2 family protein
256	200076	201383	1307	glutamine synthetase, type I, glnA
257	242791	243861	1070	hypothetical protein
257	201539	201967	428	ANTAR domain-containing protein
258	243983	244750	767	hypothetical protein
258	199500	200837	1337	hypothetical protein
258	214351	215451	1100	hypothetical protein
258	202034	203167	1133	beta-lactamase family protein
259	244747	245550	803	hypothetical protein
259	200855	201634	779	hypothetical protein
259	230580	231698	1118	hypothetical protein
259	215517	216686	1169	hypothetical protein
259	203216	204493	1277	OAH/OAS sulfhydrylase
260	245558	245917	359	hypothetical protein
260	201822	202403	581	phosphotyrosine protein phosphatase
260	231785	232450	665	sugar fermentation stimulation protein Sfs
260	216688	217527	839	glycosyl transferase GT2 family
260	204527	205888	1361	hypothetical protein
261	245988	246176	188	hypothetical protein
261	202621	203469	848	degV family protein
261	217539	218489	950	glycosyl transferase GT8 family
262	246226	247362	1136	hypothetical protein
262	203571	203945	374	desulfoferrodoxin Dfx
262	233015	233842	827	hypothetical protein
262	218491	219087	596	hypothetical protein
262	206059	207045	986	diaminopimelate dehydrogenase
263	247341	247508	167	hypothetical protein
263	203946	204254	308	hypothetical protein
263	233829	234626	797	hypothetical protein
263	219059	220168	1109	glycosyl transferase group I
263	207085	208323	1238	EAL domain-containing protein
264	247649	248422	773	hypothetical protein
264	204271	205311	1040	UmuC-like DNA repair protein
264	234664	234984	320	rhodanese domain-containing protein
264	220289	221260	971	glycosyl transferase GT2 family protein
264	208623	209555	932	hypothetical protein
265	248862	249509	647	hypothetical protein
265	205291	205602	311	hypothetical protein
265	221269	221643	374	hypothetical protein
265	209593	210258	665	hypothetical protein
266	250196	251011	815	CAAX amino terminal protease family protein
266	205631	207226	1595	hypothetical protein
266	234997	236454	1457	hypothetical protein
266	221645	223105	1460	hypothetical protein
266	210595	211395	800	hypothetical protein
267	250996	251580	584	hypothetical protein
267	207461	207907	446	hypothetical protein
267	236441	237316	875	glycosyl transferase GT2 family
267	211388	211618	230	hypothetical protein

268	208261	209028	767	flagellin FlhC
268	237300	238574	1274	hypothetical protein
268	223117	225216	2099	hypothetical protein
268	211710	213020	1310	hypothetical protein
269	251742	252107	365	hypothetical protein
269	209517	210356	839	oligopeptide ABC transporter permease protein OppB
269	238591	239502	911	hypothetical protein
269	225235	226284	1049	glycosyl transferase
269	213223	215286	2063	hypothetical protein
270	210374	211432	1058	oligopeptide ABC transporter permease protein OppC
270	239502	240380	878	glycosyl transferase GT2 family
270	226301	227260	959	conserved hypothetical
271	252120	253202	1082	hypothetical protein
271	211457	212461	1004	oligopeptide ABC transporter ATP-binding protein OppD
271	240430	241314	884	glycosyl transferase
271	227273	230125	2852	glycosyl transferase
271	215408	216256	848	hypothetical protein
272	253294	255357	2063	metalloendopeptidase M13 family
272	212461	213447	986	oligopeptide ABC transporter ATP-binding protein OppF
272	241293	242414	1121	glycosyl transferase GT4 family protein
272	230129	232144	2015	glycosyl transferase GT2 family protein
272	216299	216748	449	ribose 5-phosphate isomerase RpiB
273	255413	256258	845	hypothetical protein
273	232173	235379	3206	glycosyl transferase GT2 family protein
273	216779	217555	776	hypothetical protein
274	213552	215399	1847	oligopeptide ABC transporter substrate-binding protein OppA
274	242416	243351	935	radical SAM domain-containing protein
274	235386	236612	1226	hypothetical protein
274	217568	218227	659	hypothetical protein
275	256426	256986	560	hypothetical protein
275	215883	218963	3080	hypothetical protein
275	218242	218634	392	hypothetical protein
276	256983	257726	743	hypothetical protein
276	218978	220633	1655	sulfate permease SulP family
276	236774	237592	818	flagellin FlhC
277	220668	221432	764	PHP domain-containing protein
277	244289	245077	788	HD domain-containing protein
277	237821	238639	818	flagellin FlhC
277	218753	219190	437	hypothetical protein
278	257751	258587	836	hypothetical protein
278	221474	221977	503	hypothetical protein
278	245425	245790	365	hypothetical protein
278	219246	220106	860	hypothetical protein
279	258644	259489	845	hypothetical protein
279	245827	246456	629	hypothetical protein
279	238887	240059	1172	hypothetical protein
279	220118	220660	542	hypothetical protein
280	222155	222367	212	hypothetical protein
280	246491	247024	533	hypothetical protein
280	240071	240880	809	glycosyl transferase
280	220684	220980	296	anti-sigma factor antagonist
281	259488	261152	1664	GGDEF/EAL domain-containing protein
281	222546	223394	848	hypothetical protein
281	247105	248724	1619	D-mannonate oxidoreductase Uxu
281	240937	241416	479	hypothetical protein
282	261199	261753	554	hypothetical protein
282	248756	249832	1076	D-mannonate dehydratase Uxu
282	241440	241841	401	flagellar protein FlhS
283	261956	263464	1508	ABC transporter ATP-binding protein
283	241883	243418	1535	flagellar hook-associated protein FlhD
284	224575	224766	191	hypothetical protein
284	249979	250905	926	AraC family transcriptional regulator
284	243497	243970	473	flagellar protein FlaG
284	221162	223501	2339	ferrous iron transport protein FeoA/B
285	263630	264316	686	ABC transporter permease
285	224763	225332	569	hypothetical protein
285	250941	251417	476	methylglyoxal synthase mgs
285	244133	245320	1187	hypothetical protein
285	223630	223932	302	hypothetical protein
286	264331	265329	998	tRNA dihydrouridine synthase A dusA
286	251598	252317	719	hypothetical protein
287	265387	265743	356	hypothetical protein
287	225501	226112	611	hypothetical protein
287	245393	245608	215	global regulator family protein
288	226418	226576	158	hypothetical protein
288	252330	253328	998	hypothetical protein
288	245652	246110	458	flagellar assembly protein FlhW
288	224084	225040	956	hypothetical protein
289	265819	266727	908	serine O-acetyltransferase cysE
289	226738	227028	290	hypothetical protein
289	253331	253666	335	PadR family transcriptional regulator
289	246207	248282	2075	flagellar hook-associated protein FlgL
290	253932	254561	629	hypothetical protein

290	248387	250234	1847	flagellar hook-associated protein FlgK
291	266965	268293	1328	flagellar hook protein FlgE
291	227916	229568	1652	hypothetical protein
291	254635	254853	218	hypothetical protein
291	225148	226095	947	L-lactate dehydrogenase
292	268881	269468	587	hypothetical protein
292	229945	231435	1490	anthranilate synthase component I TrpE
292	254971	256284	1313	erythromycin esterase
292	250250	251827	1577	flagellar hook-associated protein FlgK
293	269476	270057	581	signal peptidase I lepB
293	231453	232031	578	anthranilate synthase component II TrpG
293	256281	258284	2003	ABC transporter permease
293	252043	252570	527	FlgN family protein
293	226233	226952	719	4Fe-4S ferredoxin iron-sulfur binding domain- containing protein
294	232150	233541	1391	PhoH family protein
294	258274	259044	770	ABC transporter ATP-binding protein
294	252597	252884	287	negative regulator of flagellin synthesis FlgM
295	270070	274959	4889	LPxTG motif-containing cell surface protein
295	259180	260046	866	two component system histidine kinase
295	227050	227748	698	cNMP binding domain-containing protein
296	274984	276513	1529	LPxTG motif-containing cell surface protein
296	260096	260770	674	two component system response regulator
296	227767	228072	305	hypothetical protein
297	276653	27	-276626	sortase
297	234546	234758	212	hypothetical protein
297	260840	261631	791	GNAT family acetyltransferase
297	253691	256003	2312	hypothetical protein
298	234789	235010	221	hypothetical protein
298	261645	262139	494	SAM-dependent methyltransferase/GNAT family acetyltransferase
298	256000	256638	638	hypothetical protein
299	256655	257368	713	branched-chain amino acid ABC transporter ATP- binding protein LivF
299	228607	229320	713	hypothetical protein
300	235071	237359	2288	ferrous iron transport protein FeoA/B
300	262449	263453	1004	hypothetical protein
300	257397	258152	755	branched-chain amino acid ABC transporter ATP- binding protein LivG
301	277554	278669	1115	LPxTG motif-containing cell surface protein
301	258149	259189	1040	branched-chain amino acid ABC transporter permease LivM
302	263507	264829	1322	MATE efflux family protein
302	259220	259909	689	branched-chain amino acid ABC transporter permease LivH
302	229523	230284	761	hypothetical protein
303	278735	27	-278708	sortase
303	237875	238138	263	hypothetical protein
303	264876	266876	2000	hypothetical protein
303	260264	261451	1187	branched-chain amino acid ABC transporter substrate-binding protein LivJ
304	279601	280374	773	hypothetical protein
304	238151	238390	239	hypothetical protein
304	267015	268442	1427	hypothetical protein
304	261569	262123	554	hypothetical protein
304	230533	231303	770	histidinol phosphate phosphatase HisJ family
305	280390	281187	797	hypothetical protein
305	262202	262507	305	hypothetical protein
305	231410	234565	3155	hypothetical protein
306	238467	239405	938	RNA pseudouridylylate synthase
306	268475	269095	620	methylated-DNA-protein-cysteine S- methyltransferase Ogt
306	234675	236780	2105	metalloendopeptidase M13 family
307	281252	281584	332	hypothetical protein
307	239464	240768	1304	UmuC-like DNA repair protein
307	269391	270143	752	hypothetical protein
308	281786	282829	1043	iron ABC transporter ATP-binding protein
308	270208	271143	935	hypothetical protein
308	262949	264241	1292	voltage gated chloride channel protein
309	271325	272068	743	hypothetical protein
309	264257	265204	947	sugar ABC transporter substrate-binding protein
309	237294	238901	1607	hypothetical protein
310	282844	283869	1025	iron ABC transporter permease protein
310	272049	272402	353	hypothetical protein
310	265257	267278	2021	methyl-accepting chemotaxis protein McpL
311	284064	284672	608	iron ABC transporter ATP-binding protein
311	241162	241941	779	CAAX amino terminal protease family protein
311	272399	272884	485	hypothetical protein
311	238912	239505	593	hypothetical protein
312	285346	285816	470	hypothetical protein
312	242185	242961	776	sortase B family protein
312	273200	273748	548	hypothetical protein
312	267385	269052	1667	spermidine synthase family protein
313	285900	286553	653	HAD superfamily hydrolase
313	243127	243684	557	SCP domain-containing protein
313	273749	274120	371	hypothetical protein
313	269090	270085	995	cobalamin biosynthesis protein CbiD
313	239617	240345	728	hypothetical protein
314	286597	288258	1661	sugar ABC transporter substrate-binding protein
314	243764	243976	212	hypothetical protein
314	274296	274712	416	ribosomal protein S9 Rps

315	288657	289100	443	hypothetical protein
315	243995	244336	341	PilZ domain-containing protein
315	274721	275149	428	ribosomal protein L13 rpl
315	270095	270886	791	Cobalamin-5-phosphate synthase
315	240854	241258	404	NUDIX domain-containing protein
316	289133	289723	590	hypothetical protein
316	244444	245628	1184	nucleotide sugar dehydrogenase
316	275393	276169	776	tRNA pseudouridine synthase A tru
316	270883	271830	947	hypothetical protein
317	289762	290073	311	hypothetical protein
317	245634	246419	785	bifunctional biotin--acetyl-CoA-carboxylase ligase/biotin operon
317	276191	276994	803	ABC transporter permease
317	241684	241902	218	hypothetical protein
318	277011	277859	848	ABC transporter ATP-binding protein
318	271832	272653	821	>nicotinate (nicotinamide) nucleotide adenyllyltransferase NadD<
318	241901	243415	1514	hypothetical protein
319	290999	291940	941	dGTP triphosphohydrolase Dgt
319	277877	278725	848	ABC transporter ATP-binding protein
319	272658	272996	338	>nicotinate (nicotinamide) nucleotide adenyllyltransferase NadD<
319	243401	244957	1556	SpoIIE domain-containing protein
320	291944	293722	1778	DNA primase DnaG
320	246598	247161	563	hypothetical protein
321	278848	279726	878	hypothetical protein
322	293739	295463	1724	RNA polymerase sigma factor RpoD
322	279800	280375	575	ribosomal protein L17 rpl
322	273084	274163	1079	histidinol-phosphate aminotransferase HisC
323	295531	296262	731	hypothetical protein
323	280495	281451	956	DNA-directed RNA polymerase alpha subunit rpoA
323	274271	275194	923	lipoprotein NLP A family
323	244959	246731	1772	MatE efflux family protein
324	296272	297927	1655	phosphoribulokinase/uridine kinase family protein
324	248038	249459	1421	L-arabinose isomerase
324	281497	282090	593	ribosomal protein S4 rpsD
324	275365	276363	998	ABC transporter ATP-binding protein
324	246872	247873	1001	hypothetical protein
325	297930	298973	1043	NAD-dependent epimerase/dehydratase
325	249496	250929	1433	rhamnulokinase RhaB
325	282250	282660	410	ribosomal protein S11 rpsK
325	276405	277025	620	ABC transporter permease
325	247903	248208	305	hypothetical protein
326	299020	299610	590	hypothetical protein
326	250954	251829	875	rhamnulose-1-phosphate aldolase RhaD
326	282762	283130	368	ribosomal protein S13 rpsM
326	277075	277860	785	SAM-dependent methyltransferase
326	248208	248735	527	NUDIX domain-containing protein
327	299712	300035	323	hypothetical protein
327	283323	283436	113	ribosomal protein L36 rpmJ
327	277891	278547	656	HAD superfamily hydrolase
327	248939	250087	1148	PP-loop family protein
328	300185	301720	1535	von Willebrand factor type A domain-containing protein
328	251965	252939	974	AraC family transcriptional regulator
328	278618	278773	155	Multimeric flavodoxin WrbA
329	301781	302254	473	RNA polymerase sigma factor sigma-70 family protein
329	252953	253840	887	hypothetical protein
329	283635	284555	920	hypothetical protein
329	278895	280163	1268	MATE efflux family protein
329	250124	250720	596	radical SAM domain-containing protein
330	302370	302783	413	hypothetical protein
330	253925	255658	1733	GGDEF/EAL domain-containing protein
330	280298	281638	1340	hypothetical protein
331	255705	256040	335	hypothetical protein
331	284742	286121	1379	uracil-xanthine permease
331	282411	283217	806	hypothetical protein
331	250732	251259	527	hypothetical protein
332	303397	304284	887	5-keto 4-deoxyuronate isomerase KduI
332	286447	287307	860	aldo/keto reductase family oxidoreductase
332	283663	284421	758	NAD synthetase nadE
332	251261	252124	863	pyridoxine kinase
333	287374	288099	725	xylosidase/arabinofuranosidase Xsa
333	304361	305239	878	2-deoxy-D-gluconate 3-dehydrogenase kduD
334	305264	306226	962	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase, eda
334	256154	257068	914	3-phosphoshikimate 1-carboxyvinyltransferase AroA
334	288285	289034	749	two component system response regulator
334	284453	285658	1205	nicotinate phosphoribosyltransferase pncB
334	252296	252883	587	hypothetical protein
335	306267	307292	1025	2-dehydro-3-deoxygluconokinase KdgK
335	289012	290613	1601	two component system histidine kinase
335	285672	286184	512	cytidyltransferase-related domain-containing protein
336	257497	258594	1097	prephenate dehydrogenase TyrA
336	290706	292376	1670	sugar ABC transporter substrate-binding protein
336	286199	2867	-283332	amidase
336	253222	253866	644	hypothetical protein
337	307615	308631	1016	GGDEF domain-containing protein

337	258825	260915	2090	translation elongation factor EF-G-like protein
337	292381	293250	869	sugar ABC transporter permease
337	286705	287433	728	NUDIX domain-containing protein
338	260996	262939	1943	EAL domain-containing protein
338	293268	294212	944	sugar ABC transporter permease
338	253868	254449	581	hypothetical protein
339	308635	309900	1265	HD-GYP domain-containing protein
339	288195	288989	794	electron transport complex RnfB
339	254446	255393	947	hypothetical protein
340	309991	311040	1049	lacI family transcriptional regulator
340	263124	264557	1433	pyruvate carboxyltransferase domain-containing protein
340	294401	296200	1799	methyl-accepting chemotaxis protein McpK
340	255582	256169	587	SAM-dependent methyltransferase
341	264574	265254	680	GntR family transcriptional regulator
341	289014	289601	587	electron transport complex RnfA
341	256172	256951	779	SAM-dependent methyltransferase
342	311275	312231	956	sugar ABC transporter permease
342	265282	266484	1202	isocitrate dehydrogenase Icd
342	296345	297361	1016	TRAP transporter solute receptor DctP family dct
342	256941	258035	1094	radical SAM domain-containing protein
343	312253	313176	923	sugar ABC transporter permease
343	266516	268492	1976	ABC transporter ATP-binding protein
343	297358	298146	788	two component system response regulator
343	289604	290383	779	electron transport complex RnfE
344	268627	269259	632	CAAX amino terminal protease family protein
344	290383	290991	608	electron transport complex rnfG
344	258351	259139	788	hypothetical protein
345	313420	314994	1574	sugar ABC transporter substrate-binding protein
345	269285	270664	1379	MFS transporter
345	298152	299507	1355	two component system histidine kinase
345	290994	291869	875	electron transport complex rnfD
345	259154	259633	479	RNA polymerase sigma factor sigma-70 family protein
346	315261	317129	1868	beta-galactosidase Bga35A
346	270788	271738	950	protease activity modulator HflK
346	299729	300151	422	hypothetical protein
346	291954	293270	1316	electron transport complex rnfC
346	259820	260383	563	hypothetical protein
347	317452	319323	1871	beta-galactosidase Bga2B
347	271748	272632	884	protease activity modulator HflC
347	300221	301366	1145	TRAP transporter DctM subunit dctM
347	260407	261495	1088	hypothetical protein
348	272862	274343	1481	transcription termination factor Rho
348	301516	302040	524	hypothetical protein
348	261626	262564	938	hypothetical protein
349	274391	275272	881	PII domain-containing protein
349	302079	303140	1061	TRAP transporter DctP subunit dctP
349	293492	294679	1187	hypothetical protein
349	262796	263290	494	hypothetical protein
350	275378	276073	695	alanine racemase domain-containing protein
350	294745	295014	269	hypothetical protein
351	303381	303599	218	translation initiation factor IF-1 inf
351	295152	295889	737	MerR family transcriptional regulator
352	276845	277726	881	GGDEF domain-containing protein
352	303621	304379	758	methionine aminopeptidase, type I, map
352	295886	296791	905	hypothetical protein
352	263533	264426	893	hypothetical protein
353	296939	298090	1151	unsaturated glucuronyl hydrolase Ugl88B
353	304372	305016	644	adenylate kinases adk
354	305106	306410	1304	preprotein translocase subunit Sec
355	277922	28050	-249872	xylosidase
355	306425	306865	440	ribosomal protein L15 RplO
355	298115	299641	1526	polysaccharide biosynthesis protein
355	264596	265006	410	peptide deformylase def
356	280531	281802	1271	hypothetical protein
356	306989	307177	188	ribosomal protein L30 RpmD
356	265067	266299	1232	hypothetical protein
357	281808	282473	665	hypothetical protein
357	307194	307703	509	ribosomal protein S5 RpsE
358	307722	308090	368	ribosomal protein L18 RplR
358	299641	300570	929	glycosyl transferase GT11 family protein
358	266604	266936	332	hypothetical protein
359	308110	308649	539	ribosomal protein L6 RplF
359	300598	301740	1142	polysaccharide pyruvyl transferase
360	282557	283711	1154	hypothetical protein
360	308665	309066	401	ribosomal protein S8 RpsH
360	301767	302744	977	glycosyl transferase GT2 family protein
360	266933	267931	998	HNH endonuclease domain-containing protein
361	283846	285135	1289	alpha-L-fucosidase Fuc29A
361	309207	309377	170	ribosomal protein S14 RpsN
361	302729	304099	1370	hypothetical protein
361	267943	268806	863	hypothetical protein
362	309411	309905	494	ribosomal protein L5 RplE
362	268803	269891	1088	hypothetical protein

363	285449	286591	1142	hypothetical protein
363	309973	310302	329	ribosomal protein L24 rplX
363	304119	305297	1178	polysaccharide biosynthesis protein
364	310315	310683	368	ribosomal protein L14 rplN
364	305318	306274	956	glycosyl transferase GT17 family protein
364	270328	270729	401	hypothetical protein
365	310705	310959	254	30S ribosomal protein S17 rpsQ
365	306264	307298	1034	glycosyl transferase GT2 family protein
366	310992	311201	209	ribosomal protein L29 rpmC
366	307313	308032	719	hypothetical protein
366	270912	272651	1739	hypothetical protein
367	311191	311622	431	ribosomal protein L16 rplP
367	308087	309124	1037	glycosyl transferase GT11 family protein
367	272735	27588	-245147	helicase
368	311628	312299	671	ribosomal protein S3 rpsC
368	309138	310931	1793	ABC transporter ATP-binding/permease
369	312319	312714	395	ribosomal protein L22 rplV
369	310951	311823	872	glycosyl transferase GT11 family protein
370	312745	313035	290	ribosomal protein S19 rpsS
370	311858	313753	1895	asparagine synthase glutamine-hydrolyzing AsnB
371	313047	313844	797	ribosomal protein L2 rplB
371	313825	314913	1088	hypothetical protein
372	313985	314284	299	ribosomal protein L23 RplW
373	314284	314907	623	ribosomal protein L4 RplD
373	276609	276929	320	hypothetical protein
374	314935	315618	683	ribosomal protein L3 rplC
374	314930	316462	1532	MBOAT family acyltransferase
375	315760	316071	311	ribosomal protein S10 rpsJ
375	316494	317696	1202	glycosyl transferase GT4 family protein
376	316589	317470	881	hypothetical protein
376	277755	278639	884	hypothetical protein
377	317687	319573	1886	asparagine synthase glutamine-hydrolyzing AsnB
377	278670	280226	1556	hypothetical protein
378	317538	318602	1064	hypothetical protein
378	319642	320751	1109	glycosyl transferase GT4 family protein
379	320758	321957	1199	glycosyl transferase GT4 family protein
379	280486	280719	233	hypothetical protein
380	280789	281520	731	feruloyl esterase Est1
380	318771	319532	761	TPR domain-containing protein
381	321957	323399	1442	glycosyl transferase GT4 family protein
381	281563	282012	449	D-tyrosyl-tRNA(Tyr) deacylase dtd
382	323727	324461	734	polysaccharide deacetylase Est4E
382	319536	320441	905	von Willebrand factor type A domain-containing protein
382	282033	282848	815	nitroreductase family protein
383	320495	321589	1094	von Willebrand factor type A domain-containing protein
384	321589	321984	395	hypothetical protein
384	324690	325664	974	NAD-dependent epimerase/dehydratase
384	282951	283883	932	cobalamin biosynthesis protein CobW
385	322052	322834	782	hypothetical protein
385	325866	327209	1343	nucleotide sugar dehydrogenase
385	283942	285009	1067	NADP-dependent alcohol dehydrogenase
386	322964	323959	995	AAA family ATPase
386	327279	328376	1097	glycosyl transferase GT4 family protein
386	285232	286527	1295	MFS transporter
387	323973	324836	863	5,10-methylenetetrahydrofolate reductase MetF
387	328379	329374	995	hypothetical protein
387	286517	287632	1115	HTH domain-containing protein
388	324947	326461	1514	adenylosuccinate lyase purB
388	329379	330005	626	bacterial sugar transferase
389	326562	328007	1445	glutamine phosphoribosylpyrophosphate amidotransferase PurF
389	330033	331304	1271	DegT/DnrJ/EryC1/StrS family aminotransferase
389	287791	288069	278	hypothetical protein
390	328127	329320	1193	hypothetical protein
390	331307	333247	1940	polysaccharide biosynthesis protein
390	288066	288488	422	hypothetical protein
391	329343	330182	839	hypothetical protein
391	288488	289108	620	hypothetical protein
392	330200	331501	1301	hypothetical protein
393	333263	335896	2633	hypothetical protein
393	289232	289927	695	hypothetical protein
394	331501	335094	3593	hypothetical protein
394	335978	337603	1625	hypothetical protein
394	289917	290219	302	hypothetical protein
395	335209	336420	1211	ABC transporter permease
395	337653	338597	944	glycosyl transferase GT2 family protein
395	290491	290730	239	hypothetical protein
396	336435	337646	1211	ABC transporter permease
396	338613	339128	515	GNAT family acetyltransferase
396	290746	291171	425	hypothetical protein
397	337662	339254	1592	RND family transporter
397	291237	291629	392	hypothetical protein
398	339283	339981	698	ABC transporter ATP-binding protein
398	339141	340901	1760	hypothetical protein

398	291626	292276	650	hypothetical protein
399	340916	342052	1136	TDP-4-keto-6-deoxy-D-glucose transaminase
400	342067	342525	458	hypothetical protein
400	292530	292874	344	HxlR family transcriptional regulator
401	340390	341175	785	formate/nitrite transporter fnt
401	342545	344185	1640	hypothetical protein
401	292947	293189	242	hypothetical protein
402	341280	341987	707	hypothetical protein
402	293337	294617	1280	OAH/OAS sulphydrylase
403	344284	345396	1112	hypothetical protein
403	294757	295617	860	PP-loop family protein
404	342142	344217	2075	molecular chaperone Hsp90
404	345417	347168	1751	ABC transporter ATP-binding/permease
405	295741	297648	1907	glycoside hydrolase family 31 Gh31D
405	344471	344911	440	hypothetical protein
406	347193	348218	1025	lipooligosaccharide sialyltransferase
406	297776	299524	1748	hypothetical protein
407	348271	349332	1061	glycosyl transferase GT2 family protein
407	299727	300698	971	MCP signaling domain-containing protein
408	344967	345464	497	hypothetical protein
408	349329	350411	1082	glycosyl transferase GT28 family protein
409	350427	351500	1073	N-acetylneuraminate synthase NeuB
409	300717	301832	1115	hypothetical protein
410	346320	347126	806	two component system response regulator
410	351501	353429	1928	CMP-N-acetylneuraminic acid synthetase NeuA
410	301919	303256	1337	hypothetical protein
411	347135	348913	1778	two component system histidine kinase
411	303493	304101	608	hypothetical protein
412	349015	349863	848	sugar ABC transporter permease
412	353426	354628	1202	hypothetical protein
413	349866	350780	914	sugar ABC transporter permease
413	304098	304736	638	hypothetical protein
414	354678	356234	1556	hypothetical protein
415	350865	352226	1361	sugar ABC transporter substrate-binding protein
415	356238	357245	1007	glycosyl transferase GT2 family protein
415	305082	305558	476	YbaK/prolyl-tRNA synthetase domain-containing protein
416	352294	352836	542	hypothetical protein
416	305723	306547	824	hypothetical protein
417	353076	353816	740	hypothetical protein
417	357238	358461	1223	hypothetical protein
417	306643	307689	1046	hypothetical protein
418	358469	359551	1082	glycosyl transferase GT4 family protein
419	353842	355650	1808	hypothetical protein
419	359674	360918	1244	glycosyl transferase GT4 family protein
419	307938	308234	296	hypothetical protein
420	355668	356558	890	peptidyl-prolyl cis-trans isomerase PpiC-type
421	356726	358123	1397	FAD dependent oxidoreductase
421	360948	361841	893	glycosyl transferase GT2 family protein
421	308330	310342	2012	heavy metal translocating P-type ATPase
423	358269	360260	1991	2-isopropylmalate synthase leuA
423	361853	362965	1112	polysaccharide biosynthesis protein
423	310421	310759	338	hypothetical protein
424	360551	362104	1553	xylosidase/arabino-furanosidase Xsa43D
424	362966	363787	821	glycosyl transferase GT2 family protein
424	310825	311358	533	rhodanese domain-containing protein
425	362224	362694	470	hypothetical protein
425	363787	364893	1106	polysaccharide export protein
425	311371	311913	542	hypothetical protein
426	364911	366311	1400	exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase
426	311993	313015	1022	peptidase S15 family protein
427	362750	364054	1304	thiamine biosynthesis protein ThiC
427	366330	367076	746	glycosyl transferase GT2 family protein
427	313012	313251	239	hypothetical protein
428	364108	364863	755	phosphomethylpyrimidine kinase thiD
428	313350	314222	872	LysR family transcriptional regulator
429	364946	365581	635	thiamine-phosphate pyrophosphorylase thiE
429	367076	367957	881	glycosyl transferase GT2 family protein
429	314287	314883	596	SAM-dependent methyltransferase
430	365591	366409	818	hydroxyethylthiazole kinase thiM
430	314887	315507	620	hypothetical protein
431	366410	367075	665	thiazole transporter ThiW
431	315586	315867	281	hypothetical protein
433	315963	316940	977	hypothetical protein
434	317598	318377	779	hypothetical protein
436	318619	319206	587	hypothetical protein
437	319203	319946	743	hypothetical protein
439	319959	320393	434	hypothetical protein
442	320438	321157	719	hypothetical protein
443	321177	322025	848	hypothetical protein
445	322037	322624	587	hypothetical protein
447	323367	323861	494	hypothetical protein
449	323869	324612	743	hypothetical protein
450	324626	325507	881	hydrolase alpha/beta fold family

451	325561	326289	728	hypothetical protein
454	326778	328019	1241	hypothetical protein
455	328016	329002	986	hypothetical protein
456	329173	329952	779	hypothetical protein
457	330032	330955	923	hypothetical protein
459	330957	331856	899	hypothetical protein
461	331850	332455	605	hypothetical protein
462	332472	333074	602	hypothetical protein
463	333113	333562	449	hypothetical protein
465	333621	334448	827	hypothetical protein
466	334467	335102	635	hypothetical protein
467	335108	335614	506	hypothetical protein
468	335629	336249	620	SAM-dependent methyltransferase
470	336253	336846	593	SAM-dependent methyltransferase
471	337007	337279	272	ferric uptake regulator family protein
472	337711	338667	956	succinylglutamate desuccinylase/aspartoacylase family protein
473	338686	339438	752	hypothetical protein
474	339447	340208	761	SAM-dependent methyltransferase
475	340247	340987	740	peptide/nickel ABC transporter ATP-binding protein
476	340989	341711	722	peptide/nickel ABC transporter ATP-binding protein
477	341715	342533	818	peptide/nickel ABC transporter permease
479	342533	343456	923	peptide/nickel ABC transporter permease
481	343449	345059	1610	peptide/nickel ABC transporter periplasmic protein
482	345186	345992	806	hypothetical protein
483	346122	347324	1202	MFS transporter
485	347417	347971	554	hypothetical protein
486	347995	348243	248	GNAT family acetyltransferase
487	348320	348460	140	GNAT family acetyltransferase
488	348465	348908	443	MarR family transcriptional regulator
490	349270	350838	1568	lysyl-tRNA synthetase lysS
492	350958	351443	485	transcription elongation factor GreA/GreB
493	351588	352565	977	tRNA dihydrouridine synthase A DusA
494	352566	353048	482	hypothetical protein
496	353067	353882	815	Baf family transcriptional regulator
498	353895	354698	803	biotin-[acetyl-CoA-carboxylase] ligase birA
499	354700	355692	992	ornithine carbamoyltransferase argF
500	355818	356654	836	cell wall hydrolase
501	357076	357870	794	pyrroline-5-carboxylate reductase proC
502	357872	358675	803	rRNA methylase SpoU family protein
503	358795	359763	968	6-phosphofructokinase pfkA
504	359925	363464	3539	DNA polymerase III E DnaE
505	363741	364931	1190	S-adenosylmethionine synthetase MetK
506	365254	366723	1469	xylulokinase xylB
507	366792	367370	578	hypothetical protein
508	367367	369202	1835	hypothetical protein
509	369361	372033	2672	S1 RNA binding domain-containing protein
510	372115	372654	539	hypothetical protein
511	372664	373176	512	hypothetical protein
513	373202	373930	728	hypothetical protein
515	373946	375871	1925	glycoside hydrolase family 2 Gh2B
517	375951	377321	1370	MatE efflux family protein
519	377494	378432	938	carbamate kinase arcC
520	378486	380015	1529	ATP-dependent Clp protease ATP-binding subunit ClpX
521	380034	381455	1421	serine protease HtrA family protein
523	381707	381967	260	hypothetical protein
524	382054	382866	812	histidinol phosphate phosphatase HisJ family
526	382914	383930	1016	UDP-glucose 4-epimerase galE
527	384072	385343	1271	UTP-glucose-1-phosphate uridylyltransferase GalU
528	385422	387107	1685	lysozyme Lyc
530	387254	388732	1478	L-fucose isomerase related protein
531	388921	391044	2123	beta-xylosidase Xyl
533	391373	393085	1712	sugar ABC transporter substrate-binding protein
534	393150	394022	872	sugar ABC transporter permease
535	394038	394874	836	sugar ABC transporter permease
536	395023	397866	2843	alpha-glucuronidase Gh115
538	398069	398938	869	AraC family transcriptional regulator
540	398948	400936	1988	alpha-D-glucuronidase Agu67A
542	401017	401901	884	AraC family transcriptional regulator
544	401911	403791	1880	acetyl-xylan esterase
545	403871	405886	2015	beta-glucosidase Bgl3E
546	405906	407633	1727	hypothetical protein
547	407699	409177	1478	sugar ABC transporter substrate-binding protein
548	409246	410256	1010	sugar ABC transporter permease
549	410249	411157	908	sugar ABC transporter permease
551	411172	413538	2366	hypothetical protein
553	413764	414423	659	hypothetical protein
554	414428	415954	1526	NHL repeat-containing protein
555	415975	416850	875	sugar ABC transporter permease
556	416840	417778	938	sugar ABC transporter permease
558	417795	420710	2915	sugar ABC transporter substrate-binding protein
559	420728	422083	1355	endo-1,4-beta-xylanase Xyn
562	422448	423389	941	AraC family transcriptional regulator
564	423404	424852	1448	xylosidase/arabinofuranosidase Xsa43

566	425565	426353	788	hypothetical protein
567	426433	427686	1253	endo-1,4-beta-glucanase/xylanase Cel5
569	427707	430484	2777	xylosidase/arabinofuranosidase and esterase Xsa43
571	430842	431270	428	ATP synthase F1 epsilon subunit AtpC
572	431276	432670	1394	ATP synthase F1, beta subunit atpD
574	432848	433708	860	ATP synthase F1 gamma subunit atpG
575	433726	435246	1520	ATP synthase F1 alpha subunit atpA
576	435269	435808	539	ATP synthase F1 delta subunit atpH
577	435969	436394	425	ATP synthase F0 B subunit atpF
578	436525	436806	281	ATP synthase F0 C subunit atpE
579	436858	437628	770	ATP synthase F0 A subunit atpB
580	437639	438049	410	hypothetical protein
582	438073	438351	278	hypothetical protein
583	438613	440643	2030	methyl-accepting chemotaxis protein McpC
584	440875	442647	1772	aspartyl-tRNA synthetase aspS
585	442800	443444	644	hypothetical protein
586	443551	444006	455	LysM domain-containing protein
587	444069	445055	986	hypothetical protein
588	445170	446576	1406	PTS system N- acetylglucosamine-specific IIBC component NagE
589	446664	447164	500	PTS system N- acetylglucosamine-specific IIA component
591	447279	449027	1748	carbon starvation protein CstA
592	449024	449596	572	hypothetical protein
594	449937	451301	1364	magnesium transporter mgfE
596	451345	453222	1877	oligopeptide ABC transporter substrate-binding protein
597	453255	454607	1352	GGDEF domain-containing protein
598	454811	456298	1487	glutamate synthase small subunit GltD
599	456341	460858	4517	glutamate synthase large subunit GltB
601	460889	462997	2108	glutamine synthetase type III GlnA
602	463037	464563	1526	asparagine synthase glutamine-hydrolyzing AsnB
604	464647	466335	1688	ammonium transporter amt
605	466710	467873	1163	hypothetical protein
607	467906	469072	1166	hypothetical protein
608	469091	470821	1730	hypothetical protein
609	470922	471368	446	hypothetical protein
610	471341	471700	359	hypothetical protein
612	471804	472580	776	hypothetical protein
614	472556	473902	1346	MATE efflux family protein matE
615	474087	474908	821	rRNA methylase SpoU family protein
617	474911	477430	2519	adenosine deaminase add
618	477462	478199	737	amino acid ABC transporter ATP-binding protein
620	478186	478872	686	amino acid ABC transporter permease
621	478883	479599	716	amino acid ABC transporter substrate-binding protein
622	479823	482087	2264	pyruvate formate lyase Pfl
623	482258	482992	734	pyruvate formate-lyase 1-activating enzyme pfl
624	483010	483774	764	cNMP binding domain-containing protein
625	483805	484365	560	NUDIX domain-containing protein
626	484368	485522	1154	beta-lactamase family protein
627	485545	487338	1793	Na/Pi-cotransporter II-related protein
628	487448	488362	914	GNAT family acetyltransferase
629	488649	488909	260	phosphocarrier HPr family protein
630	488922	489845	923	hypothetical protein
631	489861	490679	818	P-loop-containing ATPase
632	490952	491749	797	UDP-N-acetylenolpyruvoylglucosamine reductase murB
633	491889	492824	935	glucokinase Glk
634	492849	493796	947	Hpr kinase/phosphatase HprK
636	493852	495720	1868	excinuclease ABC subunit C uvr
638	495785	497983	2198	ATP-dependent metallopeptidase Hfl
639	498227	499837	1610	CTP synthase pyrG
640	499912	500637	725	glucosamine-6-phosphate isomerase nag
641	500634	501512	878	carbohydrate kinase ROK family protein
642	501652	503232	1580	sugar ABC transporter substrate-binding protein
644	503422	504288	866	sugar ABC transporter permease
645	504288	505289	1001	sugar ABC transporter permease
646	505464	506471	1007	oligopeptide ABC transporter ATP-binding protein OppF
647	506458	507585	1127	oligopeptide ABC transporter ATP-binding protein OppD
648	507563	508624	1061	oligopeptide ABC transporter permease OppC
650	508681	509781	1100	oligopeptide ABC transporter permease OppB
652	509874	511919	2045	oligopeptide ABC transporter substrate-binding protein OppA
653	512081	513235	1154	iron-containing alcohol dehydrogenase
654	513250	514260	1010	myo-inositol 2-dehydrogenase
655	514288	515124	836	Myo-inosose-2 dehydratase
656	515217	515996	779	myo-inositol catabolism protein
658	516025	517956	1931	myo-inositol catabolism protein
660	518138	519154	1016	myo-inositol catabolism protein
661	519201	520865	1664	hypothetical protein
662	520882	521808	926	carbohydrate kinase ROK family protein
663	521825	522916	1091	unsaturated rhamnogalacturonyl hydrolase Gh105
665	523099	523857	758	AraC family transcriptional regulator
668	523844	525532	1688	PHP domain-containing protein
669	525705	526496	791	hypothetical protein
670	526560	529721	3161	alpha-mannosidase Man38
671	529851	531218	1367	sugar transporter GPH family protein
673	531269	531925	656	hypothetical protein

675	531988	533601	1613	phosphoenolpyruvate carboxykinase Pck
676	533771	535099	1328	glycoside hydrolase family 27 Gh27
678	535102	535953	851	tetrapyrrole methylase family protein
679	535953	536696	743	SAM-dependent methyltransferase
680	536683	537621	938	PSP1 domain-containing protein
681	537618	538664	1046	DNA polymerase III delta subunit hol
682	538675	539115	440	hypothetical protein
683	539167	539760	593	guanylate kinase gmk
684	539769	541178	1409	pyruvate kinase pyk
685	541160	542593	1433	Orn/Lys/Arg decarboxylase
686	542586	543179	593	HD domain-containing protein

References

- Akin, D.E. (1980) Evaluation by electron microscopy and anaerobic culture of types of rumen bacteria associated with digestion of forage cell walls. *Applied and Environmental Microbiology* 39, 242-252.
- Altermann, E. and Klaenhammer, T.R. (2003) GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *Omics A Journal of Integrative Biology* 7, 161-169.
- Anderson, K. and Blair, B., 1996, Regulation of the cellulolytic activity of *Eubacterium cellulosolvens* 5494: a review. *SAAS bulletin, biochemistry and biotechnology* 9, 57.
- Armstrong, D.G. and Gilbert, H.J., 1985, Biotechnology and the rumen: a mini review. *Journal of the Science of Food and Agriculture* 36, 1039-1046.
- Aspinall, G. and Preiss, J. (1980) Chemistry of cell wall polysaccharides. *The biochemistry of plants. A comprehensive treatise. Volume 3. Carbohydrates: structure and function.*, 473-500.
- Attwood, G.T., Kelly, W.J., Altermann, E.H., Moon, C.D., Leahy, S. and Cookson, A.L. (2008) Application of rumen microbial genome information to livestock systems in the postgenomic era. *Australian Journal of Experimental Agriculture* 48, 695-700.
- Attwood, G.T. and Reilly, K. (1995) Identification of proteolytic rumen bacteria isolated from New Zealand cattle. *Journal of Applied Bacteriology* 79, 22-29.
- Attwood, G.T., Klieve, A.V., Ouwerkerk, D. and Patel, B.K.C., 1998, Ammonia-hyperproducing bacteria from New Zealand ruminants. *Applied and Environmental Microbiology* 64, 1796.
- Attwood, G.T., Reilly, K. and Patel, B.K.C. (1996) *Clostridium proteoclasticum* sp nov, a novel proteolytic bacterium from the bovine rumen. *International Journal of Systematic Bacteriology* 46, 753-758.
- Ayers, W.A., 1958, Nutrition and physiology of *Ruminococcus flavefaciens*. *Journal of Bacteriology* 76, 504.
- Badger, J.H. and Olsen, G.J. (1999) CRITICA: coding region identification tool invoking comparative analysis. *Molecular Biology and Evolution* 16, 512.
- Baskin, T.I., 2001, On the alignment of cellulose microfibrils by cortical microtubules: a review and a model. *Protoplasma* 215, 150-171.

- Beard, C.E., Hefford, M.A., Forster, R.J., Sontakke, S., Teather, R.M. and Gregg, K. (1995) A stable and efficient transformation system for *Butyrivibrio fibrisolvens* OB156. *Curr Microbiol* 30, 105-109.
- Black, G.W., Rixon, J.E., Clarke, J.H., Hazlewood, G.P., Theodorou, M.K., Morris, P. and Gilbert, H.J. (1996) Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. *Biochemical Journal* 319, 515-520.
- Bonfield, J.K., Beal, K.F., Betts, M.J. and Staden, R. (2002) Trev: a DNA trace editor and viewer. *Bioinformatics* 18, 194.
- Bonhomme, A. (1990) Rumen ciliates - Their metabolism and relationships with bacteria and their hosts. *Animal Feed Science and Technology* 30, 203-266.
- Borneman, W.S., Hartley, R.D., Himmelsbach, D.S. and Ljungdahl, L.G. (1990) Assay for trans-p-coumaroyl esterase using a specific substrate from plant cell walls. *Analytical Biochemistry* 190, 129-133.
- Bozzola, J.J. and Russell, L.D. 1999. *Electron microscopy: principles and techniques for biologists*. Jones & Bartlett Learning.
- Brown, R.M. and Saxena, I.M. (2000) Cellulose biosynthesis: A model for understanding the assembly of biopolymers. *Plant Physiology and Biochemistry* 38, 57-67.
- Campbell, M.M. and Sederoff, R.R. (1996) Variation in lignin content and composition - Mechanism of control and implications for the genetic improvement of plants. *Plant Physiology* 110, 3-13.
- Carpita, N.C. (1989) Pectic polysaccharides of maize coleoptiles and proso millet cells in liquid culture. *Phytochemistry* 28, 121-125.
- Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annual Review of Plant Physiology and Plant Molecular Biology* 47, 445-476.
- Carpita, N.C. and Gibeaut, D.M. (1993) Structural models of primary-cell walls in flowering plants - Consistency of molecular-structure with the physical properties of the walls during growth. *Plant Journal* 3, 1-30.
- Carver, T., Thomson, N., Bleasby, A., Berriman, M. and Parkhill, J. (2009) DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119.
- Cheng, K.J., Forsberg, C.W., Minato, H. and Costerton, J.W. (1991) Microbial ecology and physiology of feed degradation within the rumen. In: T. Tsuda, Y. Sasaki

- and R. Kawashima (Eds), *Physiological Aspects of Digestion and Metabolism in Ruminants*. Academic Press, Toronto, ON, p. 595-624.
- Chesson, A. and Forsberg, C.W. (1997) Polysaccharide degradation by rumen microorganisms. In: P.N. Hobson and C.S. Stewart (Eds), *The Rumen Microbial Ecosystem*. Blackie Academic & Professional, London, p. 329-381.
- Cho, S.J., Cho, K.M., Shin, E.C., Lim, W.J., Hong, S.Y., Choi, B.R., Kang, J.M., Lee, S.M., Kim, Y.H., Kim, H. and Yun, H.D. (2006) 16S rDNA analysis of bacterial diversity in three fractions of cow rumen. *Journal of Microbiology and Biotechnology* 16, 92-101.
- Coen, J.A. and Dehority, B.A. (1970) Degradation and utilisation of hemicellulose from intact forages by pure cultures of rumen bacteria. *Applied Microbiology* 20, 362-365.
- Coenen, G.J., Bakx, E.J., Verhoef, R.P., Schols, H.A. and Voragen, A.G.J. (2007) Identification of the connecting linkage between homo- or xylogalacturonan and rhamnogalacturonan type-I. *Carbohydrate Polymers* 70, 224-235.
- Cole, J., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R., Kulam-Syed-Mohideen, A., McGarrell, D., Marsh, T. and Garrity, G., 2009, The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* 37, D141.
- Cosgrove, D.J. (2000) Expansive growth of plant cell walls. *Plant Physiology and Biochemistry* 38, 109-124.
- Costerton, W. (1992) Pivotal role of biofilms in the focused attack of bacteria on insoluble substrates. *International Biodeterioration & Biodegradation* 30, 123-133.
- Cotta, M. and Forster, R., 2006, The family Lachnospiraceae, including the genera *Butyrivibrio*, *Lachnospira* and *Roseburia*. *Prokaryotes* 4, 1002-1021.
- Cotta, M.A. and Hespell, R.B. (1986) Proteolytic activity of the ruminal bacterium *Butyrivibrio fibrisolvens*. *Applied and Environmental Microbiology* 52, 51-58.
- Craig, W.M., Broderick, G.A. and Ricker, D.B. (1987) Quantitation of microorganisms associated with the particulate phase of ruminal ingesta. *Journal of Nutrition* 117, 56-62.
- Delcher, A.L., Harmon, D., Kasif, S., White, O. and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* 27, 4636.

- Delcher, A.L., Salzberg, S.L. and Phillippy, A.M. (2003) Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics* 10.3.1-10.3.18
- Diez-Gonzalez, F., Bond, D.R., Jennings, E. and Russell, J.B. (1999) Alternative schemes of butyrate production in *Butyrivibrio fibrisolvens* and their relationship to acetate utilisation, lactate production, and phylogeny. *Archives of Microbiology* 171, 324-330.
- Dyrlov Bendtsen, J., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340, 783-795.
- Ebringerova, A., Hromadkova, Z. and Heinze, T. (2005) Hemicellulose. *Polysaccharides 1: Structure, Characterization and Use. Advances in Polymer Science* 186, 1-67.
- Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* 14, 755.
- Eddy, S.R. (2002) A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinformatics* 3, 18.
- Edwards, J.E., Huws, S.A., Kim, E.J., Lee, M.R.F., Kingston-Smith, A.H. and Scollan, N.D. (2008) Advances in microbial ecosystem concepts and their consequences for ruminant agriculture. *Animal* 2, 653-660.
- Edwards, J.E., McEwan, N.R., Travis, A.J. and Wallace, R.J. (2004) 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* 86, 263-281.
- Ehrlich, G.G., Goerlitz, D.F., Bourell, J.H., Eisen, G.V. and Godsy, E.M. (1981) Liquid chromatographic procedure for fermentation product analysis in the identification of anaerobic bacteria. *Applied and Environmental Microbiology* 42, 878.
- Escalante, A., Wachter, C. and Farres, A., 2001, Lactic acid bacterial diversity in the traditional Mexican fermented dough pozol as determined by 16S rDNA sequence analysis. *International Journal of Food Microbiology* 64, 21-31.
- Ewing, B., Hillier, L.D., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research* 8, 175.
- FAOSTAT. (2007) Food and agriculture organization of the united nations. <http://faostat.fao.org/>.

- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.
- Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R. and White, B.A. (2008) Polysaccharide utilisation by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology* 6, 121-131.
- Forster, R.J., Gong, J.H. and Teather, R.M. (1997) Group-specific 16S rRNA hybridization probes for determinative and community structure studies of *Butyrivibrio fibrisolvens* in the rumen. *Applied and Environmental Microbiology* 63, 1256-1260.
- Gardiner, K., 1991, Pulsed field gel electrophoresis. *Analytical chemistry* 63, 658-665.
- Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S. and Eddy, S.R. (2009) Rfam: updates to the RNA families database. *Nucleic Acids Research* 37, D136.
- Gregg, K., Hamdorf, B., Henderson, K., Kopečný, J. and Wong, C. (1998) Genetically modified ruminal bacteria protect sheep from fluoroacetate poisoning. *Applied and Environmental Microbiology* 64, 3496.
- Grabber, J.H., 2005, How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. *Crop Sci* 45, 820-831.
- Gong, J., Egbosimba, E.E. and Forsberg, C.W. (1996) Cellulose-binding proteins of *Fibrobacter succinogenes* and the possible role of a 180-kDa cellulose-binding glycoprotein in adhesion to cellulose. *Canadian Journal of Microbiology* 42, 453-460.
- Hackmann, T. and Spain, J. (2010) Invited review: Ruminant ecology and evolution: Perspectives useful to ruminant livestock research and production. *Journal of Dairy Science* 93, 1320-1334.
- Haft, D.H., Selengut, J.D. and White, O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Research* 31, 371.
- Hamada, S. and Fujita, S., 1983, DAPI staining improved for quantitative cytofluorometry. *Histochemistry and Cell Biology* 79, 219-226.
- Harrison, P.W., Lower, R.P.J., Kim, N.K.D. and Young, J.P.W. (2010) Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in Microbiology* 18, 141-148.
- Hashimoto, W., He, J., Wada, Y., Nankai, H., Mikami, B. and Murata, K. (2005a) Proteomics-based identification of outer-membrane proteins responsible for

- import of macromolecules in *Sphingomonas* sp. A1: alginate-binding flagellin on the cell surface. *Biochemistry* 44, 13783-13794.
- Hashimoto, W., Momma, K., Maruyama, Y., Yamasaki, M., Mikami, B. and Murata, K. (2005b) Structure and function of bacterial super-biosystem responsible for import and depolymerization of macromolecules. *Bioscience, Biotechnology, and Biochemistry* 69, 673-692.
- Hatfield, R.D., Wilson, J.R. and Mertens, D.R. (1999) Composition of cell walls isolated from cell types of grain sorghum stems. *Journal of the Science of Food and Agriculture* 79, 891-899.
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* 280, 309-316.
- Henrissat, B. (1998) Glycosidase families. *Biochemical Society Transactions* 26, 153-156.
- Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* 7, 637-644.
- Henrissat, B., Coutinho, P.M., Rancurel, C. and Lombard, V. (2011) CAZy (Carbohydrate-Active enZYmes) database. <http://www.cazy.org/>.
- Hespell, R.B. and Cotta, M.A. (1995) Degradation and utilisation by *Butyrivibrio fibrisolvens* H17c of xylans with different chemical and physical-properties. *Applied and Environmental Microbiology* 61, 3042-3050.
- Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119.
- Ioelovich, M., 2008, Cellulose as a nanostructured polymer: A short review. *BioResources* 3, 1403-1418.
- Ishii, T. (1997) Structure and functions of feruloylated polysaccharides. *Plant Science* 127, 111-127.
- Janecek, S., Svensson, B. and MacGregor, E.A. (2003) Relation between domain evolution, specificity, and taxonomy of the alpha-amylase family members containing a C-terminal starch-binding domain. *European Journal of Biochemistry* 270, 635-645.
- Jarvis, M.C. (1984) Structure and properties of pectin gels in plant cell-walls. *Plant Cell and Environment* 7, 153-164.

- Jun, H., Qi, M., Ha, J. and Forsberg, C. (2007a) *Fibrobacter succinogenes*, a dominant fibrolytic ruminal bacterium: transition to the post genomic era. Asian-Australasian Journal of Animal Sciences 20, 802-810.
- Jun, H.S., Qi, M., Gong, J., Egbosimba, E.E. and Forsberg, C.W. (2007b) Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. Journal of Bacteriology 189, 6806.
- Justé, A., Thomma, B. and Lievens, B., 2008, Recent advances in molecular techniques to study microbial communities in food-associated matrices and processes. Food Microbiology 25, 745-761.
- Kabel, M.A., van den Borne, H., Vincken, J.P., Voragen, A.G.J. and Schols, H.A. (2007) Structural differences of xylans affect their interaction with cellulose. Carbohydrate Polymers 69, 94-105.
- Kelly, W., Asmundson, R. and Hopcroft, D., 1987, Isolation and characterization of a strictly anaerobic, cellulolytic spore former: *Clostridium chartatabidum* sp. nov. Archives of Microbiology 147, 169-173.
- Kelly, W.J., Leahy, S.C., Altermann, E., Yeoman, C.J., Dunne, J.C., Kong, Z., Pacheco, D.M., Li, D., Noel, S.J. and Moon, C.D. (2010) The Glycobiome of the Rumen Bacterium *Butyrivibrio proteoclasticus* B316^T Highlights Adaptation to a Polysaccharide-Rich Environment. PLoS ONE 5, e11942.
- Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S. and Janssen, P.H., 2010, Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. Journal of Microbiological Methods.
- Kim, M., Morrison, M. and Yu, Z. (2011) Status of the phylogenetic diversity census of ruminal microbiomes. FEMS Microbiology Ecology 76, 49-63.
- Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16, 111-120.
- Kircher, M. and Kelso, J., 2010, High throughput DNA sequencing—concepts and limitations. Bioessays 32, 524-536.
- Koike, S., Pan, J., Kobayashi, Y. and Tanaka, K. (2003a) Kinetics of *in sacco* fiber attachment of representative ruminal cellulolytic bacteria monitored by competitive PCR. Journal of Dairy Science 86, 1429-1435.

- Koike, S., Yoshitani, S., Kobayashi, Y. and Tanaka, K. (2003b) Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiology Letters* 229, 23-30.
- Kondo, T., Mizuno, K. and Kato, T. (1990) Cell-wall bound para-coumaric and ferulic acids in Italian ryegrass. *Canadian Journal of Plant Science* 70, 495-499.
- Kong, Z. (2007) Microarray Analysis of *Clostridium Proteoclasticum* Genes Involved in Hemicellulose Degradation. MSc, University of Auckland.
- Kopecny, J., Logar, R.M. and Kobayashi, Y. (2001) Phenotypic and genetic data supporting reclassification of *Butyrivibrio fibrisolvens* isolates. *Folia Microbiologica* 46, 45-48.
- Kopecny, J., Zorec, M., Mrazek, J., Kobayashi, Y. and Marinsek-Logar, R. (2003) *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrvibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. *International Journal of Systematic and Evolutionary Microbiology* 53, 201.
- Krogh, A., Larsson, B.È., Von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* 305, 567-580.
- Kumar, S., Nei, M., Dudley, J. and Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9, 299.
- Lambrecht, F.L. (1966) Notes on the growth curve of *Trypanosoma cruzi* Chagas 1909 as determined by optical density. *Revista do Instituto de Medicina Tropical de Sao Paulo* 8, 249-54.
- Lang, R.(2002) Ghostscript, Ghostview and GSview. Retrieved September 19, 2002from <http://www.cs.wisc.edu/~ghost/index.html>.
- Larue, R., Yu, Z., Parisi, V.A., Egan, A.R. and Morrison, M. (2005) Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. *Environmental Microbiology* 7, 530-543.
- Leedle, J. and Hespell, R.B., 1980, Differential carbohydrate media and anaerobic replica plating techniques in delineating carbohydrate-utilizing subgroups in rumen bacterial populations. *Applied and Environmental Microbiology* 39, 709.

- Liab, K., Azadi, P., Collins, R., Tolan, J., Kim, J.S. and Eriksson, K.E.L. (2000) Relationships between activities of xylanases and xylan structures. *Enzyme and Microbial Technology* 27, 89-94.
- Linder, A., Bergman, R., Bodin, A. and Gatenholm, P. (2003) Mechanism of assembly of xylan onto cellulose surfaces. *Langmuir* 19, 5072-5077.
- Lu, G. and Moriyama, E.N.(2004) Vector NTI, a balanced all-in-one sequence analysis suite. *Briefings in bioinformatics* 5, 378.
- Mackenzie, C.R., Bilous, D., Schneider, H. and Johnson, K.G. (1987) Induction of cellulolytic and xylanolytic enzyme-systems in *Streptomyces* spp. *Applied and Environmental Microbiology* 53, 2835-2839.
- Martin, S.A. and Dean, R.G. (1989) Characterization of a plasmid from the ruminal bacterium *Selenomonas ruminantium*. *Applied and Environmental Microbiology* 55, 3035-8.
- Maynard, L., Loosli, J., Hintz, H. and Warner, R. 1979 *Animal Nutrition*. 7th. In: *The carbohydrates and their metabolism*. McGraw-Hill, Inc., New York, NY, p. 74-103
- McAllister, T.A., Bae, H.D., Jones, G.A. and Cheng, K.J. (1994) Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* 72, 3004-3018.
- Miller, M.E.B., Antonopoulos, D.A., Rincon, M.T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B. and Coutinho, P.M. (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS ONE* 4, e6650.
- Minato, H., Mitsumori, M. and Cheng, K.J. (1993) Attachment of microorganisms to solid substrate in the rumen. K.E.A. Shimada (Ed) *Genetics, Biochemistry and Ecology of Lignocellulose Degradation*. Uni Publishers, Tokyo., p. 139-145.
- Miron, J. and Benghedalia, D. (1993) Digestion of cell-wall monosaccharides of ryegrass and alfalfa hays by the ruminal bacteria *Fibrobacter succinogenes* and *Butyrivibrio fibrisolvens*. *Canadian Journal of Microbiology* 39, 780-786.
- Miron, J., Duncan, S.H. and Stewart, C.S. (1994) Interactions between rumen bacterial strains during the degradation and utilisation of the monosaccharides of barley straw cell-walls. *Journal of Applied Bacteriology* 76, 282-287.
- Mitchell, R.A.C., Dupree, P. and Shewry, P.R. (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiology* 144, 43-53.

- Moon, C., Gagic, D., Li, D., Sang, C., Altermann, E., Kelly, W., Leahy, S. and Attwood, G. (2010a) A metagenomic exploration of lignocellulose degradation in the bovine rumen. In: 5th JGI Annual User Meeting, Walnut Creek, CA, USA.
- Moon, C., Leahy, S., Kelly, W., Altermann, E., Li, D. and Attwood, G. (2010b) Diversity of plant polysaccharide degradation capabilities in rumen fibrolytic bacteria as revealed by genome and metagenome studies. In: Proceedings of ISME-13, Washington State Convention & Trade Center, Seattle, WA, USA.
- Moon, C.D., Pacheco, D.M., Kelly, W.J., Leahy, S.C., Li, D., Kopečný, J. and Attwood, G.T. (2008) Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate producing ruminal bacterium. International Journal of Systematic and Evolutionary Microbiology. 58, 2041-2045.
- Morgan, W. and Corbel, M. (1976) Recommendations for the description of species and biotypes of the genus *Brucella*. Developments in Biological Standardization 31, 27.
- Mrázek, J. and Kopečný, J., 2001, Development of competitive PCR for detection of *Butyrivibrio fibrisolvens* in the rumen. Folia Microbiologica 46, 63-65.
- Nakamura, A., Furuta, H., Maeda, H., Takao, T. and Nagamatsu, Y. (2002a) Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean polysaccharides. Bioscience Biotechnology and Biochemistry 66, 1155-1158.
- Nakamura, A., Furuta, H., Maeda, H., Takao, T. and Nagamatsu, Y. (2002b) Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan. Bioscience Biotechnology and Biochemistry 66, 1301-1313.
- Nawrocki, E.P. and Eddy, S.R., 2007, Query-dependent banding (QDB) for faster RNA similarity searches. PLoS Computational Biology 3, e56.
- Neinhuis, C. and Edelman, H., 1996, Methanol as a rapid fixative for the investigation of plant surfaces by SEM. Journal of Microscopy 184, 14-16.
- Nocek, J.E. and Russell, J.B. (1988) Protein and energy as an integrated system - Relationship of ruminal protein and carbohydrate availability to microbial synthesis and milk production. Journal of Dairy Science 71, 2070-2107.
- Noel, S *et al.* (Unpublished) AgResearch, Palmerston North, NZ.

- O'Neil, M.A. and York, W.S. (2003) The composition and structure of plant primary cell walls. In: J.K.C. Rose (Ed) *The Plant Cell Wall*. CRC Press, p. 1-54.
- Orpin, C. and Joblin, K. (1997) The rumen bacteria. P.N. Hobson and C.S. Stewart (Eds), *The Rumen Microbial Ecosystem*. Blackie Academic and Professional Publishers., London, p. 140-195.
- Orpin, C.G., Mathiesen, S.D., Greenwood, Y. and Blix, A.S. (1985) Seasonal changes in the ruminal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *Applied and Environmental Microbiology* 50, 144-151.
- O'Sullivan, A.C., 1997, Cellulose: the structure slowly unravels. *Cellulose* 4, 173-207.
- Paillard, D., McKain, N., Chaudhary, L.C., Walker, N.D., Pizette, F., Koppova, I., McEwan, N.R., Kopecny, J., Vercoe, P.E., Louis, P. and Wallace, R.J. (2007a) Relation between phylogenetic position, lipid metabolism and butyrate production by different *Butyrivibrio*-like bacteria from the rumen. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* 91, 417-422.
- Paillard, D., McKain, N., Rincon, M.T., Shingfield, K.J., Givens, D.I. and Wallace, R.J. (2007b) Quantification of ruminal *Clostridium proteoclasticum* by real-time PCR using a molecular beacon approach. *Journal of Applied Microbiology* 103, 1251-1261.
- Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E. and Blot, M. (1999) Genomic evolution during a 10,000-generation experiment with bacteria. *National Academy of Sciences USA* 96, 3807-12.
- Pauly, M., Albersheim, P., Darvill, A. and York, W.S. (1999) Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *The Plant Journal* 20, 629-639.
- Patterson, H., Irvin, R., Costerton, J. and Cheng, K., 1975, Ultrastructure and adhesion properties of *Ruminococcus albus*. *Journal of Bacteriology* 122, 278.
- Petersen, B.E., Wu, J., Cheng, L. and Zhang, D.Y. (2008) Instrumentation. *Molecular Genetic Pathology*, 365-392.
- Pizzi, A. and Eaton, N. (1985) The structure of cellulose by conformational analysis. 2. The cellulose polymer-chain. *Journal of Macromolecular Science-Chemistry* A22, 105-137.

- PP, D. and Rees, G., 1996, A rapid method for sequencing of rRNA gene (s) amplified by polymerase chain reaction using an automated DNA sequencer. *Indian journal of microbiology* 36, 9-72.
- Purushe, J., Fouts, D.E., Morrison, M., White, B.A., Mackie, R.I., Coutinho, P.M., Henrissat, B. and Nelson, K.E. (2010) Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their environmental niche. *Microbial Ecology*, 1-9.
- Qi, M., Nelson, K.E., Daugherty, S.C., Nelson, W.C., Hance, I.R., Morrison, M. and Forsberg, C.W., 2005, Novel molecular features of the fibrolytic intestinal bacterium *Fibrobacter intestinalis* not shared with *Fibrobacter succinogenes* as determined by suppressive subtractive hybridization. *Journal of Bacteriology* 187, 3739.
- Ramsahoye, B. (2002) Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. *Methods* 27, 156-161.
- Reilly, K. and Attwood, G.T. (1998) Detection of *Clostridium proteoclasticum* and closely related strains in the rumen by competitive PCR. *Applied and Environmental Microbiology* 64, 907-913.
- Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16, 276-277.
- Russell, J.B. and Rychlik, J.L. (2001) Factors that alter rumen microbial ecology. *Science* 292, 1119.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944.
- Saito, H. and Miura, K.I. (1963) Preparation of transforming deoxyribonucleic acid by phenol treatment. *Biochimica et Biophysica Acta* 72, 619-29.
- Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4, 406-425.
- Sakai, T., Sakamoto, T., Hallaert, J. and Vandamme, E.J. (1993) Pectin, pectinase, and protopectinase - production, properties, and applications. *Advances in Applied Microbiology* 39, 213-294.
- Sales, M., Lucas, F. and Blanchart, G. (2000) Effects of ammonia and amino acids on the growth and proteolytic activity of three species of rumen bacteria: *Prevotella*

- albensis*, *Butyrivibrio fibrisolvens*, and *Streptococcus bovis*. *Current Microbiology* 40, 380-386.
- Saulnier, L. and Thibault, J.F. (1999) Ferulic acid and diferulic acids as components of sugarbeet pectins and maize bran heteroxylans. *Journal of the Science of Food and Agriculture* 79, 396-402.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M. and Federhen, S. (2011) Database resources of the national center for biotechnology information. *Nucleic Acids Research* 39, D38.
- Scalbert, A., Monties, B., Lallemand, J.Y., Guittet, E. and Rolando, C. (1985) Ether linkage between phenolic-acids and lignin fractions from wheat straw. *Phytochemistry* 24, 1359-1362.
- Schoep, T.D. and Gregg, K., 2007, Isolation and characterization of putative *Pseudobutyrvibrio ruminis* promoters. *Microbiology* 153, 3071.
- Shevchik, V.E. and Hugouvieux-Cotte-Pattat, N. (2003) PaeX, a second pectin acetyltransferase of *Erwinia chrysanthemi* 3937. *Journal of Bacteriology* 185, 3091-3100.
- Shibuya, N. and Nakane, R. (1984) Pectic polysaccharides of rice endosperm cell-walls. *Phytochemistry* 23, 1425-1429.
- Shimoyama, T., Kato, S., Ishii, S. and Watanabe, K. (2009) Flagellum mediates symbiosis. *Science* 323, 1574.
- Shin, E.C., Cho, K.M., Lim, W.J., Hong, S.Y., An, C.L., Kim, E.J., Kim, Y.K., Choi, B.R., An, J.M., Kang, J.M., Kim, H. and Yun, H.D. (2004a) Phylogenetic analysis of protozoa in the rumen contents of cow based on the 18S rDNA sequences. *Journal of Applied Microbiology* 97, 378-383.
- Shin, E.C., Choi, B.R., Lim, W.J., Hong, S.Y., An, C.L., Cho, K.M., Kim, Y.K., An, J.M., Kang, J.M., Lee, S.S., Kim, H. and Yun, H.D. (2004b) Phylogenetic analysis of archaea in three fractions of cow rumen based on the 16S rDNA sequence. *Anaerobe* 10, 313-319.
- Shin, E.C., Kim, Y.K., Lim, W.J., Hong, S.Y., An, C.L., Kim, E.J., Cho, K.M., Choi, B.R., An, J.M., Kang, J.M., Jeong, Y.J.J., Kwon, E.J., Kim, H. and Yun, H.D. (2004c) Phylogenetic analysis of yeast in the rumen contents of cattle based on the 26S rDNA sequence. *Journal of Agricultural Science* 142, 603-611.

- Skillman, L.C., Evans, P.N., Strompl, C. and Joblin, K.N. (2006) 16S rDNA directed PCR primers and detection of methanogens in the bovine rumen. *Letters in Applied Microbiology* 42, 222-228.
- Staden, R., Beal, K.F. and Bonfield, J.K. (2000) The Staden package, 1998. *Methods in Molecular Biology* 132, 115-130.
- Stewart, C.S., Flint, H.J. and Bryant, M.P. (1997) The rumen bacteria. In: P.N. Hobson and C.S. Stewart (Eds), *The Rumen Microbial Ecosystem*. Blackie Academic and Professional Publishers., London, p. 10-72.
- Suen, G., Weimer, P.J., Stevenson, D.M., Aylward, F.O., Boyum, J., Deneke, J., Drinkwater, C., Ivanova, N.N., Mikhailova, N. and Chertkov, O. (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS ONE* 6, e18814.
- Tajima, K., Aminov, R.I., Nagamine, T., Ogata, K., Nakamura, M., Matsui, H. and Benno, Y. (1999) Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *FEMS Microbiology Ecology* 29, 159-169.
- Tajima, K., Arai, S., Ogata, K., Nagamine, T., Matsui, H., Nakamura, M., Aminov, R.I. and Benno, Y. (2000) Rumen bacterial community transition during adaptation to high-grain diet. *Anaerobe* 6, 273-284.
- Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24, 1596.
- Tardy, F., Nasser, W., Robert-Baudouy, J. and Hugouvieux-Cotte-Pattat, N. (1997) Comparative analysis of the five major *Erwinia chrysanthemi* pectate lyases: Enzyme characteristics and potential inhibitors. *Journal of Bacteriology* 179, 2503-2511.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A. and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* 28, 33.
- Teeri, T.T. (1997) Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends in Biotechnology* 15, 160-167.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25, 4876.

- Tittler, R.P. and Sandholzer, L.A., 1936, The use of semi-solid agar for the detection of bacterial motility. *Journal of Bacteriology* 31, 575.
- Tschech, A. and Pfennig, N., 1984, Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*. *Archives of Microbiology* 137, 163-167.
- Ulyatt, M.J., Beever, D.E., Thomson, D.J., Evans, R.T. and Haines, M.J. (1980) Measurement of nutrient supply at pasture. *Proceedings of the Nutrition Society* 39, A67-A67.
- Van Soest, P.J., Robertson, J. and Lewis, B., 1991, Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of Dairy Science* 74, 3583-3597.
- Varner, J.E. and Lin, L.-S., 1989, Plant cell wall architecture. *Cell* 56, 231-239.
- Vogel, J. (2008) Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* 11, 301-307.
- Wang, W., Yan, L., Cui, Z., Gao, Y., Wang, Y. and Jing, R., 2011, Characterization of a microbial consortium capable of degrading lignocellulose. *Bioresource Technology*.
- Ware, C.E., Bauchop, T., Hudman, J.F. and Gregg, K. (1992) Cryptic plasmid pBf1 from *Butyrivibrio fibrisolvens* AR10: Its use as a replicon for recombinant plasmids. *Current Microbiology* 24, 193-197.
- Weimer, P.J. (1998) Manipulating ruminal fermentation: a microbial ecological perspective. *Journal of Animal Science* 76, 3114.
- Weimer, P.J., Hatfield, R.D. and Buxton, D. (1993) Inhibition of ruminal cellulose fermentation by extracts of the perennial legume cicer milkvetch (*Astragalus cicer*). *Applied and Environmental Microbiology* 59, 405.
- Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J., 1991, 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology* 173, 697.
- Whitehead, T. (1992) Genetic transformation of the ruminal bacteria *Butyrivibrio fibrisolvens* and *Streptococcus bovis* by electroporation. *Letters in Applied Microbiology* 15, 186-189.
- Whitford, M.F., Forster, R.J., Beard, C.E., Gong, J. and Teather, R.M. (1998) Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. *Anaerobe* 4, 153-163.

- Widdel, F., Kohring, G.W. and Mayer, F., 1983, Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. *Archives of Microbiology* 134, 286-294.
- Willems, A., Amat-Marco, M. and Collins, M.D. (1996) Phylogenetic analysis of *Butyrivibrio* strains reveals three distinct groups of species within the *Clostridium subphylum* of the Gram-positive bacteria. *International Journal of Systematic Bacteriology* 46, 195-199.
- Wilson, D.B. (2009) Evidence for a novel mechanism of microbial cellulose degradation. *Cellulose* 16, 723-727.
- Woodfield, D. 1999 Genetic improvements in New Zealand forage cultivars. In, Vol. 61, p. 3-7.
- Wright, A.D.G., Auckland, C.H. and Lynn, D.H. (2007) Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. *Applied and Environmental Microbiology* 73, 4206-4210.
- Yeoman, C. (2009) The auxiliary replicons of *Butyrivibrio proteoclasticus*: a thesis presented in fulfilment of the Doctorate of Philosophy degree at Massey University, Palmerston North, New Zealand.
- Zablackis, E., Huang, J., Muller, B., Darvill, A.G. and Albersheim, P. (1995) Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. *Plant Physiology* 107, 1129-1138.