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.

Physiology of rumen bacteria associated with low methane emitting sheep

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

Doctor of Philosophy

in

Microbiology and Genetics

at

Massey University, Palmerston North,

New Zealand

Sandeep Kumar

2017

Abstract

The fermentation of feed and formation of methane (CH_4) by ruminant animals occur in the rumen, and both are microbial processes. There is a natural variation in CH_4 emissions among sheep, and this variation is heritable. Therefore, breeding for sheep that naturally produce less CH_4 is a viable strategy to reduce anthropogenic greenhouse gas emissions. Rumen bacteria play a major role in feed fermentation and in the formation of hydrogen (H_2) or formate, which are converted to CH_4 by other rumen microbes called methanogens. It has been shown that rumen bacterial community compositions in low CH_4 emitting sheep differ to those in high CH_4 emitting sheep. This led to the hypothesis that the metabolism of dominant rumen bacteria associated with low CH_4 emitting sheep should explain the lower CH_4 yield, for example by producing less H_2 or formate than bacteria associated with high CH_4 emitting sheep. In this project, the diversity and physiology of members of the bacterial genera *Quinella*, *Sharpea* and *Kandleria*, which are major bacterial groups associated with low-CH_4 emitting sheep, were investigated. It appeared that the genus *Quinella* is more diverse than previously suspected, and might contain at least eight potential species, although to date none have been maintained in laboratory culture. *Sharpea* and *Kandleria* contain two and one species respectively. Experiments with *Sharpea* and *Kandleria* showed that these behave like classical lactic acid bacteria that produce lactate as their major end product and did not change their fermentation pattern to produce more H_2 or formate when grown in the presence of methanogens. This strengthens a previous hypothesis that sought to explain low CH_4 emissions from sheep with *Sharpea* and *Kandleria* in their rumens, in which this invariant production of lactate was a key assumption. *Quinella* is another bacterium found in larger numbers in the rumen of some low CH_4 sheep. Virtually nothing is known about its metabolism. FISH probes and cell concentration methods were developed which helped in its identification and resulted in construction of four genome bins of *Quinella* that were more than 90% complete with as little as 0.20% contaminated. Bioinformatic analyses of the proteins encoded by these genomes showed that *Quinella* has the enzymes for lactate formation and for the randomising pathway of propionate formation. This indicated that lactate and propionate might be major fermentation end products of *Quinella*. Additionally, the presence of an uptake hydrogenase in the *Quinella* genomes opens up the new possibility that *Quinella* might even use free H_2 in the rumen. In all these possible pathways, little or no H_2 would be produced, explaining why an increased abundance of *Quinella* in the rumen would lead to lower CH_4 emissions from those sheep with high abundances of this bacterium.

Acknowledgements

If I had to make a list of the great things that happened during my PhD, then certainly getting Dr Peter Janssen as a supervisor would go on the top of this list. I sincerely thank you for your wisdom, encouragement and valuable advice throughout my PhD. You surely have made this steep learning curve easier for me. I gratefully acknowledge Dr Gemma Henderson and Dr Sandra Kittelmann for helping me to explore ARB and QIIME and thereafter, providing valuable feedback on the project. Your contributions have made a significant difference to my thesis.

I greatly appreciate Dr Mark Patchett for being supportive and providing consistent feedback. I thank Dr Graeme Attwood for invaluable suggestions and constructive criticism throughout the project. I thank Dr Sinead Leahy for providing me with much needed support in metagenomic analysis and Dr Eric Altermann for helping me with genome annotations. I thank Dr Dragana Gagic and Associate Professor Jasna Rakonjac for attending my monthly supervisor meetings and Dr Sinead Waters for being my Teagasc supervisor. I thank Dr Christina Moon, Dr Ron Ronimus, Dr Janine Kamke, Priya Soni, Faith Cox and all of the rumen microbiology team for helping and supporting me in, and outside of the lab. I thank Dr Arjan Jonker and all animal technicians for helping me with sample collection and sheep handling and Bryan Treloar for HPLC analysis. I thank Dr Ruy Jauregui for bioinformatics support and Dr Siva Ganesh and Catherine McKenzie for statistical support.

I would like to thank Teagasc, Ireland for awarding me a Walsh Fellowship, and the New Zealand Agricultural Greenhouse Gas Research Centre **(**NZAGRC**)** for providing financial support to carry out my PhD project.

I would like to thank all the fellow students in the student room for providing a great social environment. Great thanks to my diverse flatmates covering eleven countries and five continents over the years, especially Aniek, Ermanno, Stefi, Daniel, Taina, Stephen, Maarten, Judith, Savannah and Linda. Thanks are also due to my adopted, extended New Zealand family; Dr Helal Ansari, Dr Wajid Hussain, Dr Sandeep Gupta, Dr Tanu Gupta, Dr Preeti Raju, Agneta Ghose, Lovepreet Kaur, Jaspreet Singh, Debjit Dey and Shampa Dey for making this journey joyful.

And finally, to my family, mother Deo Pari Devi, father Shambhu Sharan Prasad, brother Santosh Kumar and sister Sarika Anjan for showing faith in me and providing much needed love and encouragement. Cherished memories of my grandmother Sita Devi were a source of great comfort during my present academic journey.

Dedication

To the most impressive and beautiful woman of my life, my mother, the late **Deo Pari Devi**. Like any other child, I also ran out of the words to show my gratitude and emotions for all the love and sacrifices you have made to make me a better human being. I wish you could be here to share the joy of my accomplishment, which belongs more to you than me. But I am sure that you are watching me from heaven and pouring your blessing on me as always. I dedicate not only this thesis, but all my future achievements to you.

Table of contents

List of tables

List of figures

Abbreviations

2GenRFV	Double general substrate-Rumen Fluid-Vitamin mix
AA	Auxiliary Activities
aa	Amino acid(s)
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BES	2-Bromoethanesulfonic acid
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUbstitution Matrix
BSA	Bovine serum albumin
CAZy	Carbohydrate-Active enZYmes
CAI	Codon adaptation index
CDS	Coding DNA sequence
cfu	Colony-forming units
CH_4	Methane
CO_2	Carbon dioxide
CoA	Coenzyme A
COGs	Clusters of Orthologous Groups
CRISPR	Clustered regularly interspaced short palindromic repeat
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FGD	Functional Genome Distribution
FISH	Fluorescence *in situ* hybridisation
GHG(s)	Greenhouse gas(es)
GIT	Gastrointestinal
HMM	Hidden Markov Model
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Lysogeny broth
mRNA	Messenger RNA

NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NoSubRFV	Rumen fluid vitamin mix with no added growth substrates
NZ	New Zealand
O_2	Oxygen
ORF	Open reading frame
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
QIIME	Quantitative Insights Into Microbial Ecology
RNA	Ribonucleic acid
TAE	Tris acetate EDTA
TE	Tris EDTA
TEM	Transmission electron micrograph/microscopy
TMH	Transmembrane helix
tRNA	Transfer RNA
UV	Ultra violet
VFA	Volatile fatty acid
v:v	Volume to volume
v:v:v	Volume to volume to volume
v/v	Volume/volume
w/v	Weight/volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Measurement Units:

°C	Degrees Celsius
µg	microgram
µL	microlitre
µm	micrometre
µM	micromolar
bp	Base pair

h	Hour
kcal	kilocalorie
kb	kilobase pairs
kDa	kilodaltons
kPa	kilopascal
kV	kilovolts
L	Litre
M	Molar
Mb	megabase pairs
mg	milligram
min	minutes
mL	millilitre
mM	millimolar
ng	nanogram
nm	nanometer
ppm	Parts per million
rpm	revolutions per minute
s	seconds

Chapter 1 Introduction and literature review

1.1 Introduction

The world's ruminant population is increasing in response to an increasing global demand for ruminant products such as milk and meat, and they also provide wool and leather. Ruminant-derived products play a major role in the economy of many countries. In 2016 New Zealand (NZ) exported NZ$17.5 billion worth of ruminant products (dairy, meat and wool), which represents 36.2% of the total value of NZ's exports (Statistics, 2016 NZ). This production is based on rumen microbial fermentation, and as the total rumen fermentation has increased, so has concern about the associated environmental load associated with this fermentation. One major concern is the methane (CH_4) gas emitted as a result of ruminant enteric fermentation, which contributes significantly to NZ's total anthropogenic CH_4 emissions. CH_4 accounts for about 31.5% of NZ's total greenhouse gas (GHG) emission, almost six times the global average for CH_4 emissions from ruminants (Clark, 2009). The amount of CH_4 gas emitted into the atmosphere is increasing constantly and recently reached a new high of more than 1800 ppb (Figure 1.1).

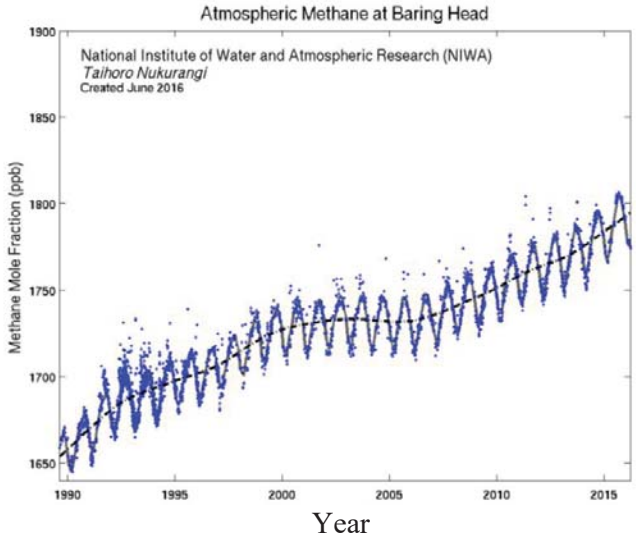


Figure 1.1 Atmospheric CH_4 at Baring Head (New Zealand). Reproduced from (NIWA, 2016) with permission.

1.2 The rumen

Nearly all the anthropogenic CH_4 formed in NZ's agriculture is derived from enteric fermentation in the rumens of farmed cattle, sheep, deer and goats. The rumen, or reticulo-rumen, is the name commonly applied to the first three of the four stomach compartments (Figure 1.2) (rumen, reticulum, omasum and abomasum) of ruminants (Hobson and Stewart, 1997). It is one of the largest organs of ruminants, and occupies 80% of the space of the abdomen. The size of the rumen varies from animal to animal even in the same species (Goopy et al., 2014) and it is approximately 70-100 litres in cattle and 5-10 litres in sheep. Ruminants typically eat grasses and low-growing plants, and also plant products as prepared feeds for farmed ruminants. Ingested feed is chewed and further broken down by rumination, a process in which the animal regurgitates the cud from the rumen to the mouth continuously to break the partly-fermented feed down into smaller pieces. This process accelerates the rate of fermentation as feed particles are continuously mixed with large volumes of saliva which helps in breakdown of short chain triglycerides by salivary lipase (Mandel, 1987). Saliva also helps to maintain the pH of the digesta in a range optimal for digestion by microbes in the rumen (Vaughan et al., 2011). The rumen can be considered as an environment maintained by the ruminant host that provides a habitat for a large number of microorganisms that ferment feed ingested by the animal. The rumen-inhabiting microorganisms breakdown the feed components and the host animal can use these products as an energy source. Overall, the rumen works like a continuous fermenter where ingested plant feed is used as a substrate and digested by the resident microorganisms, although the inputs and outputs are semi-continuous (Hungate, 1966; Hristov et al., 2003; Krause et al., 2013). The relationship of the host ruminant and the microorganisms residing in the rumen can be seen as a symbiotic relationship in which microorganisms get the habitat and substrate supply (ingested plant material) from the host and in return they ferment the feed and provide energy and other important products to the animal in the form of microbial protein and volatile fatty acids (Van Nevel and Demeyer, 1997; Calsamiglia et al., 2007; Mitsumori and Sun, 2008).

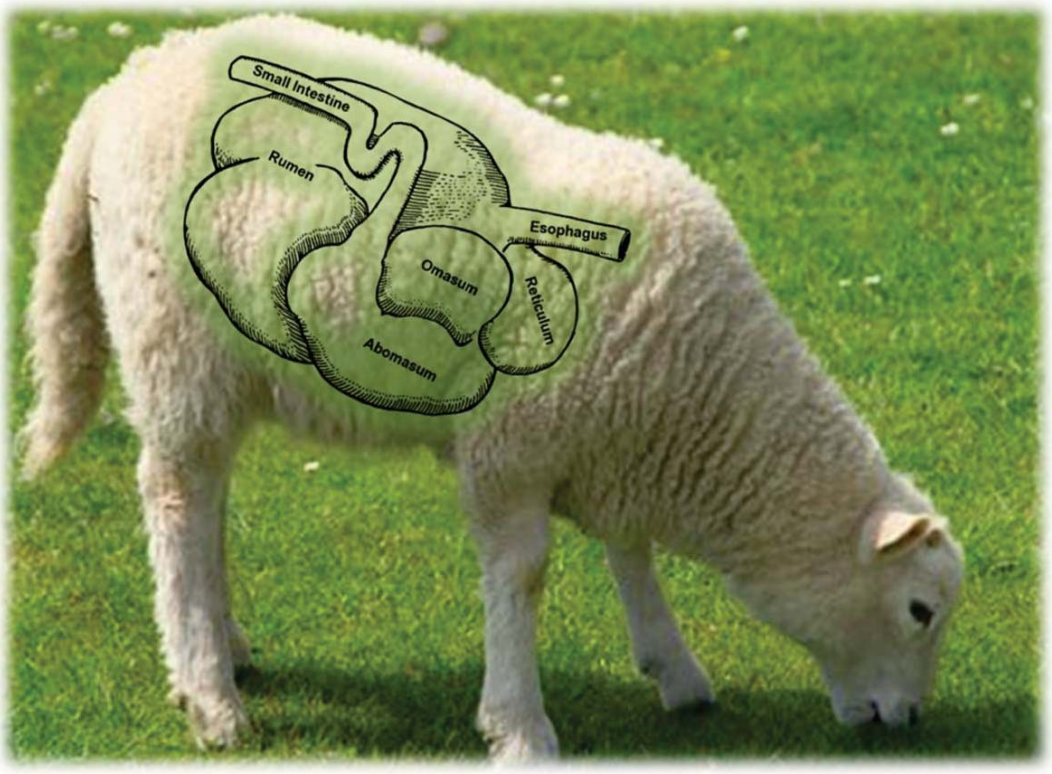


Figure 1.2 Four stomach compartments of a ruminant (rumen image was adapted from one available Wikimedia Commons, originally contributed by Pearson Scott Foresman).

1.3 Microbial groups in the rumen

The rumen microbiome has been known since ancient times, although its function and relationship with the host ruminant was unclear until the 19th century when research in rumen microbiology started to unravel the relationships between the rumen fluid, bacteria, and plant fibre degradation (Krause et al., 2013). The relationship between host and bacterial groups is not always beneficial. Sometimes it causes ruminant health problems like bloat, nitrite poisoning, and acidosis. However, most of the time rumen microbes benefit the host by degrading plant material that is otherwise poorly usable as an energy source by the animal. Before the development of anaerobic media in combination with reducing agents, the microbial population in the rumen was estimated to be around 10^6 cfu/mL (Johnson et al., 1944). A series of experiments followed to determine the growth requirements and conditions for ruminal microbes (Huhtanen and Gall, 1953a, b). A successful anaerobic technique in combination with reducing agents was developed by Robert E. Hungate to culture cellulose-digesting bacteria from the rumen of cattle (Hungate, 1950; Hungate et al., 1952). Since then, significant advances

have been made in the understanding of this complex microbial system. The rumen microbiome consists of bacteria, archaea, fungi, protozoa and phage. The rumen bacteria are present at about 8-10×10^{10} cells/mL of rumen fluid, followed by archaea (10^7-10^9 cfu/mL), protozoa (10^4-10^6 cfu/mL), fungi (10^4-10^5 cfu/mL) and phage (>10^9 particles/mL) (Klieve and Swain, 1993; Morvan et al., 1996; Waghorn and Dewhurst, 2007; Noel, 2013). Different species have different inter-dependent roles which are essential for sustaining the microbial community and its collective activity. Although there is a large amount of variation and diversity in different species of ruminants feeding on different diets, a core microbiome is found in all rumens (Henderson et al., 2015).

Bacteria are vital for ruminant digestion, and there are many different taxa present in the rumen. The majority of the rumen bacteria are strict anaerobes and culture-based studies showed that these anaerobes collectively out-number aerobic bacteria by a factor of 1000 to 1 (Hungate, 1975). Culture independent studies have estimated different numbers of bacterial species in the rumen. For example, Kim et al. (2011) reported 5271 species level OTUs (operational taxonomic units) based on 16S rRNA gene sequence catalogues from multiple studies, whereas sequence analysis conducted by (Hess et al., 2011) found nearly 1000 bacterial species in the rumen of cattle on just one diet. Bacterial community composition depends on several factors including diet (Pitta et al., 2009), ruminant species (An et al., 2005) and growth stage (Welkie et al., 2010). The domain Bacteria can further be divided into different phyla. In the rumen *Firmicutes*, *Bacteroidetes* and *Proteobacteria* are the dominant phyla (Henderson et al., 2015; Mao et al., 2015). Bacteria play a major role in the rumen fermentation, especially in fibre degradation, and the individual species vary greatly in substrate utilisation and end products produced. This differentiation of physiologies allows the numerous diverse species to co-exist in the rumen, occupying different niches. Some bacteria, like *Butyrivibrio fibrisolvens,* can hydrolyse a broad range of substrates, whereas others are more specialised, such as *Fibrobacter succinogenes*, which prefers cellulose and its breakdown products (Hobson and Stewart, 1997). Based on a global rumen census study (Henderson et al., 2015), seven most abundant bacterial groups were identified: the genera *Prevotella, Butyrivibrio,* and *Ruminococcus,* and unclassified groups in the families *Lachnospiraceae* and *Ruminococcaceae* and in the orders *Bacteroidales* and *Clostridiales*. For most of these, their biological significance in the rumen fermentation is to be yet unravelled.

Archaea present in the rumen are represented by strictly anaerobic methanogens belonging to phylum *Euryarchaeota*. Estimates of abundance based on phylogenetic markers, the 16S and 18S rRNA, suggest that 0.3 to 3.3% of the total microbial rRNA in the rumen belongs to archaea (Sharp et al., 1998; Ziemer et al., 2000). Members of orders *Methanobacteriales*, *Methanomicrobiales*, *Methanosarcinales* and *Methanomassiliicoccales* are found commonly (Janssen and Kirs, 2008; Seedorf et al., 2014). The most abundant hydrogenotrophic methanogens in the rumen use hydrogen (H_2) or formate produced by microbial fermentation as their energy source. They use the electrons released from H_2 or formate to reduce CO_2 to CH_4 (Janssen, 2010). Members of *Methanosphaera* (in the order *Methanobacteriales*) and *Methanomassiliicoccales* use H_2 to reduce methanol and other methyl compounds to produce CH_4 (Wolin, 1976; Miller and Wolin, 1985; Poulsen et al., 2013).

Protozoa are comparatively few in number but, due to their larger size in comparison to other microorganisms in the rumen, can account for up to 50% of the biomass in rumen (Newbold et al., 2015). Protozoa found in the rumen are ciliates and may be associated with other microbial members of the community, such as bacteria and methanogens. Lately, the association of protozoa and methanogens has regained attention, resulting in more detailed studies (Tymensen et al., 2012; Tymensen and McAllister, 2012; Ng et al., 2016). They are involved in the breakdown of major components of feeds (Coleman, 1985), which also facilitates bacterial action in further ruminal fermentation processes (Veira, 1986). In addition to a role in fermentation of the feed, protozoa also act as predator of other microorganisms.

Rumen fungi constitute almost 10% of the rumen microbial biomass, and were the most recently discovered members of the rumen microbial community (Orpin, 1975; Yokoyama and Johnson, 1988; Krause et al., 2013). They degrade plant cell wall carbohydrates by producing a wide range of enzymes that hydrolyse various glycosidic bonds for growth on a number of polysaccharides (Orpin, 1975).

Viruses (bacteriophages and archaeophages) are postulated to play a key role in rumen. They control the bacterial and methanogen populations in rumen by lysing them. This process also releases bacterial and methanogenic proteins which are used by the host animal as a source of amino acids (Klieve and Swain, 1993).

1.4 Pathways of fermentation in the rumen

Ruminants themselves cannot digest a large part of the ingested plant matter as they lack cellulose degrading enzymes in their digestive tract. The rumen microbes hydrolyse plant constituents such as cellulose, hemicellulose, and starch, to simple sugars which are then fermented to various end products (Figure 1.3). Proteins are also hydrolysed and the resulting amino acids used by the rumen microbes or fermented to produce a range of different products. Except for the produced gases (CH_4, CO_2), the major end products are volatile fatty acids (mainly acetate, propionate and butyrate) that are absorbed across the rumen wall into the animal's bloodstream and used as an energy and carbon source by the animal.

The end products of the rumen fermentation and their proportions depend on various factors including type of feed, passage rate through the rumen and microbial community composition (Hobson and Stewart, 1997). The fermentation pathways are also influenced by the presence of other microorganisms. In the feed fermentation process, different microorganisms cooperate in a syntrophic manner (Krause et al., 2013), where the products of one species can be the energy source for another. This includes the fermentation and degradation of products like succinate, lactate, butyrate, propionate, ethanol, formate and H_2. One of the key products that defines interactions between rumen microbes is H_2. This H_2 is used by methanogenic archaea to form CH_4. Formate produced during fermentation can also be used by methanogens to form CH_4 but H_2 is quantitatively more important as a CH_4 precursor (Hungate et al., 1970). The amount of CH_4 formed is therefore highly dependent on the amount of H_2 produced in the fermentation process. Furthermore, H_2 production is also influenced by effectiveness of H_2 removal. If the produced H_2 is not used or removed, then re-oxidation of NADH to NAD^+ is restricted, which subsequently inhibits carbohydrate degradation, ATP production, microbial growth and VFA production (Hungate et al., 1970; Wolin, 1979; Cottle et al., 2011). Importantly, different pathways are affected to differing degrees by H_2, and pathways that produce more H_2 per unit of fermented substrate are more strongly inhibited by H_2 build-up. That means that changes in the H_2 use by methanogens will influence which fermentation pathways are favoured and so change the mix of microbial species present (Janssen, 2010).

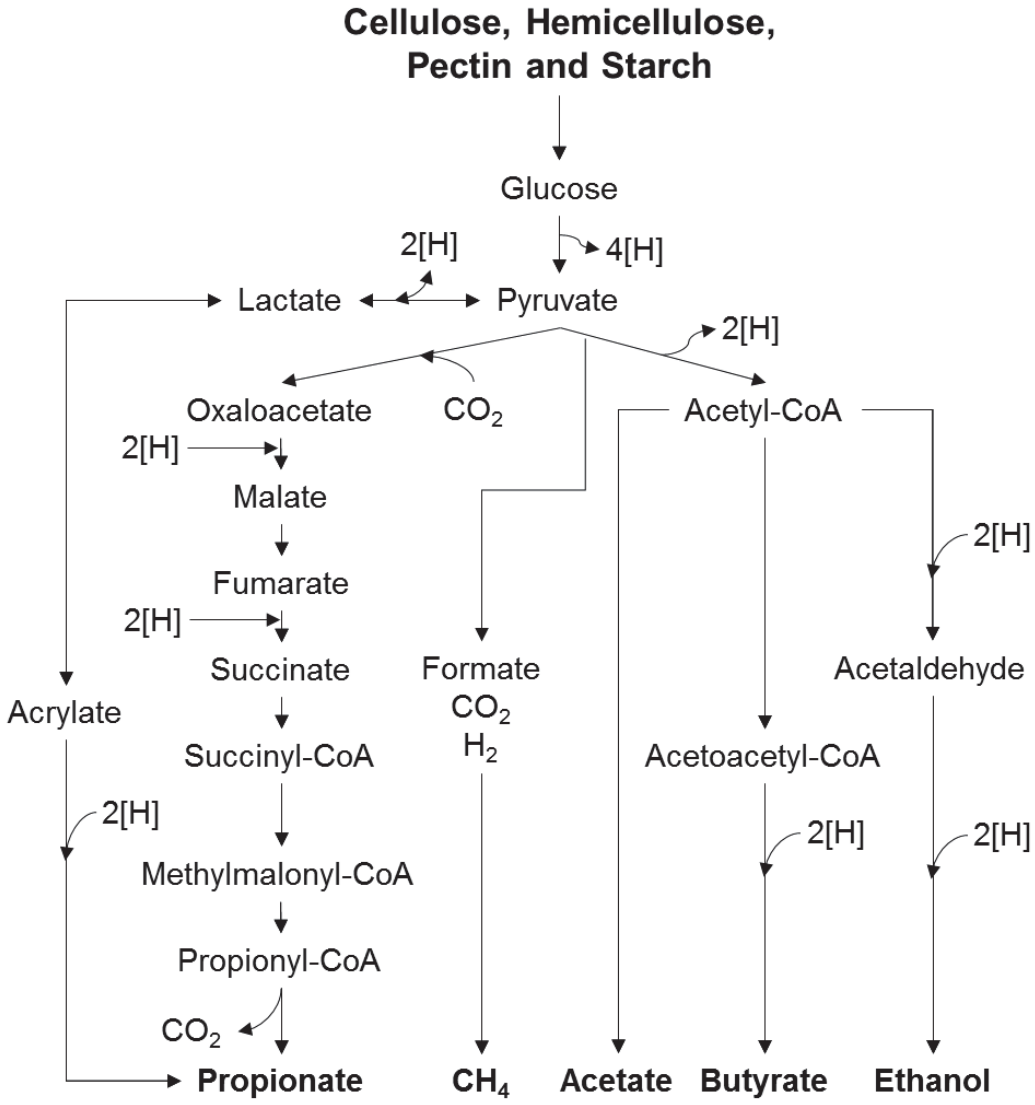


Figure 1.3 Simplified scheme of pathways of fermentation of complex plant polysaccharides, using glucose as an example of a monosaccharide released from polysaccharide breakdown (adapted from (Van Soest, 1994)). [H] represents electrons.

1.5 Mitigation of ruminant CH_4 emissions

Researchers around the globe are working on strategies to reduce ruminant CH_4 emissions. Supplemental plant feeds that contain compounds such as tannins, saponins, phenolic monomers and oils have been shown in some instances to inhibit the growth of some rumen microorganisms and help to reduce CH_4 formation in the rumen (Kamra, 2005; Patra and Yu, 2012). Furthermore, changes in diet, use of bacteriocins, defaunation of the rumen, immunisation against methanogens, and using homoacetogens and probiotics as feed supplements have been reported or suggested as ways to mitigate CH_4 emissions (Callaway and Martin, 1997; Joblin, 1999; Moss et al., 2000; Wright et al., 2004; Martin et al., 2010; Niderkorn et al., 2011; Kumar et al., 2014; Hristov et al., 2015). Many other strategies (Hristov et al., 2013; Kumar et al., 2014) for CH_4 mitigation are being investigated to find effective solutions to reduce CH_4 emission from ruminants. There is also a significant amount of work going on in New Zealand to breed productive low-CH_4 ruminants, test high sugar grasses and feed brassicas for their CH_4 mitigation potential, develop an anti-methanogen vaccine, and identify and test small molecules inhibitors of methanogens (Pinares-Patio et al., 2003; Wedlock et al., 2010; Buddle et al., 2011; Pinares-Patiño et al., 2011a; Pinares-Patiño et al., 2011b; Pinares-Patiño et al., 2013a; Pinares-Patiño et al., 2013b; Shi et al., 2014; Henderson et al., 2016).

1.6 Low CH_4 emitting animals

There are several factors which can affect ruminant CH_4 yield. The most important is the amount of feed consumed by ruminants. However, there are variations in CH_4 emissions between animals fed controlled amounts of the same feed. A study conducted by Pinares-Patiño et al. (2011a); Pinares-Patiño et al. (2013b) showed that gross CH_4 emissions (g/day) and CH_4 yield (g/kg DMI) both are heritable and repeatable traits. Furthermore, progeny from low CH_4 selection line sheep also have shown to produce less CH_4 even after shifting from lucerne pellets diet to grazing pasture (Jonker et al., 2017a). This means that selective breeding of ruminants for a low CH_4 trait could be used as a strategy to reduce enteric CH_4 emission (Hegarty et al., 2010; Martin et al., 2010; Wall et al., 2010; Buddle et al., 2011). There are several factors related to the animal that could potentially explain regulation and retention of low or high CH_4 emission traits. These include the size of the rumen and volume and passage rate of feed. A smaller rumen has low feed retention time, which is correlated with low CH_4 emissions (Goopy et al., 2014).

This likely a reflection, of changes in the passage kinetics leading to less H_2 and formate being produced through fermentation in favour of other products (Janssen, 2010).

The amount of CH_4 formed appears to be influenced by the rumen microbial community structure (Ross et al., 2012; Kittelmann et al., 2014; Rooke et al., 2014; Wallace et al., 2015; Kamke et al., 2016; Tapio et al., 2017). Variations in the composition of the microbial community inside the rumen appear to affect the amount of H_2 formed and hence CH_4 yields. Understanding this requires knowledge of the bacteria involved in H_2 production and how they are associated with low and high CH_4 yields from ruminants as lower amounts of H_2 would lead to less CH_4 production.

In addition to the influence that heritable anatomical characteristics of the animal have on the rumen microbial diversity and abundance, which again are correlated with difference in CH_4 emissions, there have been numerous studies conducted to show direct control of microbial community structure by the host (Weimer et al., 2010; King et al., 2011; Malmuthuge et al., 2012; Roehe et al., 2016). Host immune responses against members of the microbiome can also influence microbial diversity (Yáñez-Ruiz et al., 2015), which could lead to similar results.

1.7 Difference in rumen microbial communities between high and low CH_4 animals

Microbial community analysis using amplicon sequencing of phylogenetic markers to predict the relative abundance of bacteria, archaea, and eukaryotes is a commonly used method in microbial ecology, particularly for uncultivable microorganisms. The 16S rRNA gene is ubiquitous in prokaryotes. Its sequence varies between different species and can be therefore used to generate a library to catalogue species present in the microbiome. The library is prepared by amplifying the 16S rRNA gene from a mixture of "total" DNA extracted from a sample from the microbiome of interest. The library is then sequenced and the sequences can be clustered into similar sequence types on the basis of sequence similarity. Long-length sequences of nearly full-length 16S rRNA genes (>1300 bp) are valuable for preparing curated taxonomic frameworks and quality databases (Cole et al., 2003; McDonald et al., 2012; Quast et al., 2013), whereas, shorter sequences of defined regions are used to survey communities using high-throughput next generation sequencing approaches that allow only 200 to 400 bp of gene sequence to be easily sequenced (libraries of amplicons). It is anticipated that, new sequencing technologies

that allow longer sequencing reads will use more commonly for phylogenetic analysis in the future. Homology searches of the sequences against the 16S rRNA databases can then identify the bacteria at various taxonomic levels followed by analyses of their relative abundance in the samples. This approach allows multiple samples to be analysed in parallel, using bar-coded primers to distinguish the samples in a sequencing run (Caporaso et al., 2010; Smith et al., 2010). A recent study by Kittelmann et al. (2014) reported on the association of microbial taxa with low and high CH_4 sheep rumen using high-throughput barcoded 454 Titanium amplicon sequencing of bacterial, archaeal, and eukaryotic marker genes. In community profiling of 236 high and low CH_4 emitting sheep they did not find any eukaryotic (protozoa and fungi) associations with the CH_4 emission trait. However, two different bacterial community types (Q and S) were found to be associated with low CH_4 emitting sheep. The Q-type community was characterised by high relative abundance of *Quinella ovalis* while the S-type community was characterised by *Sharpea azabuensis*, *Fibrobacter* spp., *Kandleria vitulina*, *Olsenella* spp., and *Prevotella bryantii*. The species characteristics of the Q- and S-type communities were much less abundant in the rumen of high CH_4 emitting sheep. There was no obvious archaeal community structure differences between these two low CH_4 community types, but *Methanosphaera* spp. were relatively more abundant in low CH_4 emitting sheep. A similar study was conducted by Wallace et al. (2015) in cattle by selecting four pairs of extreme high and extreme low CH_4 emitting beef cattle from a group of 72 animals. In contrast to the study on sheep, members of the family *Succinivibrionaceae* were found to be the most abundant bacterial family in all samples, but were slightly more abundant in low CH_4 emitters, which corroborates the recent findings by Danielsson et al. (2017) conducted on cattle. The genera *Mitsuokella* and *Dialister* were more abundant in the low CH_4 emitting beef cattle than in the high emitters. Notably, members of the family *Veillonellaceae* were reported to be relatively more abundant in low CH_4 emitting cattle but their identity was not resolved to the genus level. Members of the phylum *Synergistetes* and the genera *Desulfovibrio*, *Megasphaera, Mogibacterium*, and *Pyramidobacter* were found to be more abundant in high CH_4 emitting beef cattle (Wallace et al., 2015). Many of these important bacteria groups associated with high and low CH_4 animal have not been characterised in detail and warrant further investigation using different approaches like metagenomic, metatranscriptomics and *in vitro* fermentation studies.

1.8 "Omics" approaches to understand community structure and physiology

Omics technologies mainly refer to detection of genes, mRNA, proteins and metabolites in biological samples using genomics, transcriptomics, proteomics and metabolomics datasets respectively (Horgan and Kenny, 2011).

Metagenomics refer to the direct study of genetic material extracted from environmental samples. It is mainly use as a culture-independent approach to understand the complex microbial communities. In contrast to amplicon sequencing of the 16S rRNA gene, this method relies on the extraction of the DNA from microbiome samples which is sequenced using shotgun sequencing platforms to produces millions of sequence reads including 16S rRNA gene sequences. Generally all of the data are used to gain insight into the metabolic potential inherent in the microbiome. Advancement in high-throughput sequencing technologies and sequencing data analysis has given an incredible insight into the microbial composition and to some extent the function of different environments (Di Bella et al., 2013) including the rumen. Sequences generated using high-throughput next generation sequencing platforms (i.e., Illumina, Roche 454, Ion Torrent and PacBio) can be used for different purposes including taxonomic profiling or community based analysis, and gene abundance and metabolic pathway prediction. This method provides vital information about a microbiome containing cultured and yet to be cultured member of a microbial population. It is used for overall prediction of microbiome function by identifying key genes and their abundance. The number of metagenomics studies of the rumen microbiome has been constantly increasing since a study by Brulc et al. (2009) used comparative metagenomics to understand the metabolic potential of bovine rumen targeting the glycoside hydrolase and cellulosome functional genes. They found that the rumen microbiome is diet driven, as those fibre degrading microorganisms capable of using readily available polysaccharides were the first to colonise plant particles, followed by the cellulose- and xylan-degrading microorganisms. Similarly, Hess et al. (2011) studied biomass-degrading genes from the cow rumen and found 27,755 putative carbohydrate-active genes from which 57% were identified as having cellulose degrading capacity. They also identified novel glycosyl hydrolyses with potential application in the biofuel industry. To understand plant biomass degradation in the rumen, Svartstrom et al. (2017) used 99 *de novo*-assembled genomes from the rumen microbiome of the moose. Many of these genomes were affiliated with *Bacteroidetes* (68), *Firmicutes* (12) and *Spirochaetes* (4), and included a wide range of carbohydrate-active enzymes (55800

glycosyl transferases, 3409 polysaccharide lyases, 5320 carbohydrate esterases and 5793 carbohydrate-binding modules) such as starch-degrading enzymes, pectin-depolymerising hydrolases, pectin- and arabinoxylan-degradation enzymes, enzymes involve in glycan utilisation, and other polysaccharides-degrading enzymes. Recently metagenomics approaches have been used to investigate aspect of enteric CH_4 emission. Ross et al. (2013) used deep parallel sequencing to investigate the effect of two CH_4 mitigating diets on rumen microbiome community profiles. They found that regardless of the additives used, changes in diets altered the rumen microbiome in a similar manner and concluded that these alterations in microbiome community profile corresponded with changes in CH_4 emissions. Wallace et al. (2015) compared the gene abundance between the rumen microbiomes of high and low CH_4 emitting cattle. Eight out of nine significantly associated genes were found to be involved in CH_4 metabolism. Genes involved in acetate formation and pyruvate metabolism were comparatively lower in low CH_4 emitters. A study conducted by (Shabat et al., 2016) suggested that genes involved in propionate formation *via* the randomising pathway were much more abundant in feed-efficient cattle. Similarly, a metagenomic study of the goat rumen showed that the abundance of genes involved in propionate formation *via* the randomising pathway increased two fold when bromochloromethane was used as an inhibiter of methanogens (Denman et al., 2015). They hypothesised that administration of bromochloromethane resulted in a shift in bacterial population abundance towards *Prevotella* and *Selenomonas* sp. that promoted propionate production by utilising H_2 via the randomising pathway. However, the same effect would have been seen if electrons had flowed directly to propionate without H_2 acting as an intermediate (Janssen, 2010). Furthermore, the rumen of low CH_4 emitting sheep as compared to high CH_4 emitting sheep had relatively higher gene abundances of genes involved in lactate, propionate and butyrate formation (Kamke et al., 2016), which was linked to a higher relative abundance of lactate-producing *Sharpea* spp., and lactate-using *Megasphaera elsdenii* (converting lactate to propionate and butyrate). Overall, metagenomics has been proven as a successful tool to reveal the rumen microbiome metabolic potential in general and helped in understanding the fermentation pattern in high and low CH_4 emitting animals. However, metagenomics only informs about the metabolic potential of the rumen microbiome on the basis of presence or absence of genes.

Metatranscriptomics is a functional omics approach where extracted RNAs from a microbiome sample are used for shotgun sequencing to generate millions of sequences with an aim to understand gene expression. While metagenomics can answer questions about the diversity of microorganisms in the microbiome and their metabolic potential, metatranscriptomics reveals the functional potential of the microbiome by analysing expressed genes. Similar to rumen metagenomics, metatranscriptomics studies using rumen samples were also initially focused on identification of expressed glycosyl hydrolases and lignocellulolytic genes. A study by Qi et al. (2011) identified glycosyl hydrolases and lignocellulolytic genes in the rumen of muskoxen were expressed at nearly nine-fold higher levels within the total carbohydrate active enzymes compared to previous rumen metagenomes studies. As these authors used polyA enrichment in the isolation of mRNA for sequencing, these novel expressed genes were proposed to originate from the eukaryotic community within the muskoxen rumen. Li and Guan (2017), conducted metatranscriptomic analysis of efficient (low residual feed intake) or low CH_4-emitting beef cattle and inefficient (high residual feed intake) or high CH_4 emitting beef cattle to find the relationship between the active rumen microbiome and feed efficiency. They found that the relative abundance of members of the families *Lachnospiraceae*, *Lactobacillaceae*, and *Veillonellaceae* was two-fold higher in efficient cattle than in inefficient cattle. Using Spearman's rank correlation analysis, these bacterial families were found to be significantly associated with methane metabolism, glyoxylate and dicarboxylate metabolism, tryptophan and valine, leucine and isoleucine degradation pathways. This finding contradicts Wallace et al. (2015) where members of the family *Veillonellaceae* were relatively more abundant in low CH_4 emitting cattle. It would be interesting to resolve the members of this family down to the genus level to identify the genera that are associated with high and low CH_4 emitting animals.

A study which combined metagenomics and metatranscriptomics was conducted in New Zealand on a cohort of sheep, aiming to understand the difference in the rumen microbiome of low and high CH_4 emitting animals (Shi et al., 2014). This demonstrated that the differences in CH_4 emission was not associated with an increase in relative abundance of methanogens and methanogenesis pathway genes. However, the expression of genes coding for enzymes in the methanogenesis pathway was significantly higher in high-CH_4-yielding sheep. Similarly, an extension of that study, conducted by Kamke et al. (2016), found that the rumen microbiomes of low CH_4 yield sheep contained relatively

higher abundances of gene transcripts associated with lactate and butyrate formation. These sheep were abundant in *Sharpea* (S- type community sheep) which was associated with less methane being emitted. On the other hand *Quinella*, the diagnostic bacterium in the Q- type low-CH_4 community, has not been study in detailed to understand its association with low CH_4 yield sheep.

1.8.1 Use of genomics to understand rumen bacterial physiology

One approach to understand the bacterial physiology would be by using pure and mixed-culture *in vitro* experiments. However, the unavailability of pure rumen isolates limit this opportunity in many cases. Robert E. Hungate started the first successful initiative to isolate pure bacterial cultures from the rumen (Hungate, 1947), and since then, with slight modifications in the culture media and conditions, many different isolation experiments have been conducted (Bladen et al., 1961; Caldwell and Bryant, 1966; Koike et al., 2010; Kenters et al., 2011; Noel, 2013; Nyonyo et al., 2013), resulting in an increase in the number of rumen bacteria isolated in pure culture. The Hungate1000 project has already started a large initiative to sequence 1000 genomes from rumen microbes(Creevey et al., 2014). Its aim is to provide a reference set of bacterial, archaeal, phage, fungal and ciliate protozoa genomes which can be used to understand the rumen microbiome, aiding research to improve feed degradation and lower enteric CH_4 emissions. So far, The Hungate1000 project has successfully sequenced 417 genomes (404 bacteria, 6 archaea, and 7 bacteriophage) (personal communication from Sinead Leahy; 15 December 2016). Combined with other projects, the rumen bacterial and archaeal genome dataset is now composed of 501 genomes (480 bacteria, 21 archaea). However, many important bacterial group remains uncultured as pointed out in a recent study on rumen microbial diversity (Henderson et al., 2015).

Use of metagenomic assemblies to construct genomes have already proven effective in uncovering the so called microbial “dark matter”. Hug et al. (2016) constructed a genome tree using 3,083 bacterial and archaeal genomes, and concluded that 59.7% of bacterial phyla and 50% of archaeal phyla still have no cultivated isolates. Advancements in metagenomic assembly and development of binning programmes like SPAdes (Bankevich et al., 2012), Velvet (Zerbino and Birney, 2008), MEGAHIT (Li et al., 2015), MetaBAT (Kang et al., 2015), CONCOT (Alneberg et al., 2014), GroopM (Imelfort et al., 2014) and MyCC (Lin and Liao, 2016) have already proven to be valuable tools to

reconstruct draft genomes from metagenomic assemblies. Such genomes can be called metagenome-assembled genomes (MAGs). CheckM (Parks et al., 2015) enables estimates to be made of the level of completeness, contamination and strain heterogeneity and so act as a quality control tool. Using these programmes, large metagenomic fragment can be binned into genomic bins on the basis of coverage, GC content and tetranucleotide frequency. This enables computational reconstruction of the genomes of microbes that are as-yet-uncultured. Assembled genome may help to understand the physiology of such yet-to-be cultured bacteria. However, genomes assembled from metagenomes should always be interpreted with caution as errors may occur in differentiating the genes amongst closely related species or strains. The success of assembling whole genomes also depends on several factors like the abundance of the target bacterial population in samples, DNA extraction methods sequencing and computational techniques.

To date only three published studies have attempted to reconstruct genomes from rumen metagenomic assemblies (Hess et al., 2011; Solden et al., 2017; Svartstrom et al., 2017). The first successful genome assemblies from rumen metagenomic samples were conducted by Hess et al. (2011) using samples from the rumen of cow. They reconstructed 15 bacterial genomes (2 *Spirochaetales*, 7 *Clostridiales*, 5 *Bacteroidales*, and 1 *Myxococcales*) with differing levels of completeness (60 to 93%). Solden et al. (2017) conducted a metagenomic study on rumen samples from the Alaskan moose, to understand members of the uncultured *Bacteroidales* family BS11, one of the seven dominant rumen bacterial groups (Henderson et al., 2015). This study revealed many potential pathways of hemicellulose sugar fermentation of *Bacteroidales* family BS11 using four nearly complete genome sequences. Similarly, Svartstrom et al. (2017) constructed 99 genomes from metagenomic assemblies from the moose rumen microbiome. This study generated 68 *Bacteroidetes* genomes and uncovered many new *Bacteroidetes* clades with fibrolytic potential.

Genomes assembled from metagenome can also reveal why particular bacteria are not culturable. Knowledge about genomes can guide cultivation strategies to aide in the predictions of the physiology of particular bacteria. For example; a metagenomic study of the candidate phylum *Saccharibacteria* suggested that member were host specific and capable of fermenting sugars in anaerobic environments. Using these findings He et al. (2015) isolated *Saccharibacteria* in the presence of *Actinomyces* in media that mimicked oral cavity conditions with a high sucrose concentration. A similar approach was used to

isolate a member of the *Succinovibrionaceae* implicated in low methane emissions from the wallaby foregut, based on genomic insights into the physiology and nutritional requirements of this bacterium (Pope et al., 2011).

1.9 Research aims

Most CH_4 formed in the rumen is generated using H_2 produced during the fermentation of feed (Hungate et al., 1970; Henderson et al., 2015), with H_2 being incorporated stoichiometrically into CH_4 ($4\ H_2 + CO_2 \rightarrow CH_4 + 2\ H_2O$). Rumen fermentation is mediated by the diverse bacterial community in the rumen. A study conducted by Pinares-Patiño et al. (2013b) showed that there is natural variation in CH_4 yields among sheep, expressed as the amount of CH_4 formed per unit of feed consumed. Some sheep are naturally high CH_4 emitters while others are low CH_4 emitters, and these are characterised by different bacteria (Kittelmann et al., 2014). It can therefore be hypothesised that bacteria associated with low CH_4 emitting animals should ferment the feed in such a way that less H_2 is formed, which would ultimately leads to less CH_4 emission.

The overarching aim of this study was to provide insights into bacteria that were differentially-associated with CH_4 yields in sheep, using culture dependent and culture independent approaches to study their fermentation products and understand how they contribute to differences in CH_4 yields. The objectives of this PhD project were to:

- Confirm the identity of bacterial taxa associated with high and low CH_4 yields.
- Refine the taxonomy of key rumen bacteria that are differentially-associated with high and low CH_4 yields.
- Understand the physiology of differentially-associated bacteria with respect to their fermentation products and the implications for CH_4 formation in the rumen.

Chapter 2 Materials and methods

2.1 Animal use and ethics approvals

The use of fistulated cows for the collection of rumen contents for culture media preparation was approved by the AgResearch Grasslands Animal Ethics Committee under animal ethics approval AE13398. The collection of rumen samples from sheep for metagenomic analyses and microscopy was approved by the AgResearch Grasslands Animal Ethics Committee under animal ethics approval AE13282.

2.2 Rumen fluid collection for media preparation

Cows fed meadow hay were used for rumen content collections. The day before collection, feed was withheld from 4 pm and rumen contents were collected the following morning at 9 - 10 am. The rumen contents were filtered through a double layer of cheese cloth with the mesh size of approximately 1 mm (Stockinette Cirtex Industries Ltd., Thames, New Zealand). Filtered rumen fluid was centrifuged at 10,000 *g* for 20 min (Sorvall Evolution RC, Thermo Fisher Scientific Inc., Waltham, MA, USA) at 4 °C to remove fine particulate material. The supernatant (rumen fluid) was decanted and stored at –20 °C until further use.

2.3 Autoclaving

Unless noted otherwise, all autoclaving was carried out at 121 °C for 20 min at 15 psi overpressure.

2.4 Media solutions and additives

2.4.1 Trace element solution SL10

The following were added in the order given to one litre of distilled water: 10 ml HCl (25%), 1.5 g $FeCl_2.4H_2O$, 0.19 g $CoCl_2.6H_2O$, 0.1 g $MnCl_2.4H_2O$, 70 mg $ZnCl_2$, 6 mg H_3BO_3, 36 mg $Na_2MoO_4.2H_2O$, 24 mg $NiCl_2.6H_2O$ and 2 mg $CuCl_2.2H_2O$ (Widdel et al., 1983). The solution was autoclaved in aliquots of 200 ml in glass bottles.

2.4.2 Selenite/tungstate solution

The selenite/tungstate solution was prepared by dissolving the following in one litre of distilled water: 0.5 g NaOH, 3 mg $Na_2SeO_3.5H_2O$ and 4 mg $Na_2WO_4.2H_2O$. The solution was autoclaved in aliquots of 200 ml in glass bottles (Tschech and Pfennig, 1984).

2.4.3 Vitamin 10 concentrate solution

To prepare the vitamin 10 concentrate solution, the following components were dissolved in one litre of anaerobic distilled water: 4-aminobenzoate (0.04 g), D-(+)-biotin (0.010 g), nicotinic acid (0.1 g), hemicalcium D-(+)-pantothenate (0.05 g), pyridoxamine hydrochloride (0.15 g), thiamine hydrochloride (0.1 g), cyanocobalamin (0.05 g), D,L-6,8-thioctic acid (0.03 g), riboflavin (0.03 g) and folic acid (0.01 g) (Kenters et al., 2011). The solution was then bubbled with O_2-free N_2 for 20 min. This anaerobic vitamin 10 concentrate solution was passed through a 0.2-µm pore size Millex GP sterile syringe filter (Millipore) using a sterile syringe and needle into a sterile (autoclaved) serum vial that had been flushed with nitrogen gas and sealed with a butyl rubber stopper.

2.4.4 Clarified rumen fluid preparation

Frozen rumen fluid (section 2.2) was thawed at room temperature, then centrifuged at 20,000 g for 15 min at 4 °C. The supernatant were collected in serum bottles (120 ml; BellCo Glass, Vineland, NJ, USA) and bubbled with O_2-free N_2 for 15 min. The bottles were then sealed with a butyl rubber stopper (BellCo Glass) and aluminium crimp cap (BellCo Glass). Sealed serum bottles were then autoclaved for 15 min to inactivate viruses. Bottles were uncapped after cooling, and 1.63 g $MgCl_2.6H_2O$ and 1.18 g $CaCl_2.2H_2O$ were added per 100 ml of rumen fluid while stirring on a magnetic stirrer. The heavy precipitate that formed was removed by centrifuging the solution at 30,000 g for 60 min at 4 °C.

2.4.5 No Substrate Rumen Fluid Vitamin mix (NoSubRFV)

NoSubRFV was prepared by dissolving 2 g of yeast extract (BD, Claix, France) in 100 ml of clarified rumen fluid. The mixture was bubbled with O_2-free N_2 for 20 min then passed through a 0.2-µm pore size Millex GP sterile syringe filter (Millipore) using a sterile syringe and needle into a sterile (autoclaved) serum vial that had been flushed with nitrogen gas and sealed with a butyl rubber stopper. Two millilitres of vitamin 10 concentrate (section 2.4.3) was added anaerobically per 100 ml of this preparation using a sterile syringe and needle and the mix was stored at 4 °C (Kenters et al., 2011). For use, 0.5 ml of NoSubRFV was added to 9 ml of growth medium.

2.4.6 General substrate-Rumen Fluid-Vitamin mix (2GenRFV)

2GenRFV was prepared by dissolving the following component in clarified rumen fluid (section 24.4) to generate a final volume of 100 ml: 0.72 g D-glucose, 0.68 g D-cellobiose, 0.60 g D-xylose, 0.60 g L-arabinose, 1.76 ml Na L-lactate syrup (50%), 4 g casamino acids, 4 g Bacto-Peptone and 4 g yeast extract. The mixture was bubbled with O_2-free N_2 for 20 min then passed through a 0.2 μm-pore size Millex GP sterile syringe filter (Millipore) using a sterile syringe and needle into a sterile (autoclaved) serum vial that had been flushed with nitrogen gas and sealed with a butyl rubber stopper. Two millilitres of vitamin 10 concentrate (section 2.4.3) was added anaerobically per 100 ml of this preparation using a sterile syringe and needle and the mix was stored at 4°C. For use, 0.5 ml of 2GenRFV was added to 9 ml of growth medium.

2.4.7 Sugar, amino acid and substrate solutions

Anaerobic sugar and amino acid solutions were prepared by dissolving the required weight of chemical (D-glucose, D-fructose, D-galactose, D-glucuronic acid, D-galacturonate, cellobiose, sucrose, lactose, raffinose, D-xylose, L-arabinose, L-rhamnose, glycerol, D-mannitol, citrate, L-glutamate, L-alanine, L-aspartate, fumarate, succinate, L-lactate) per 100 ml of distilled water to generate stock solutions of between 0.2 M and 1 M. Each solution was bubbled with O_2-free N_2 for 20 min then passed through a 0.2 μm pore size Millex GP sterile syringe filter (Millipore) using a sterile syringe and needle into a sterile (autoclaved) serum vial that had been flushed with nitrogen gas and sealed with a butyl rubber stopper. These substrate solutions were used in substrate testing experiments (chapter 4). Different amounts of substrates were added to culture media to give the desired final concentration for determining the conditions for optimal growth of strains of *Sharpea* and *Kandleria*.

2.4.8 Pectin (10% w/v)

Ten grams of pectin was dissolved in distilled water to a final total volume of 100 ml then bubbled with O_2-free N_2 for 30 min in a serum bottle. The bottle was then sealed with a butyl rubber stopper and aluminium crimp cap and autoclaved for 15 minutes. The pectin solution was stored at 4 °C until use. For use, 1 ml of 10% pectin was added to 9 ml RM02 medium to achieve a final concentration of 1% (w/v).

2.4.9 3M sodium formate and 0.5 M sodium acetate for methanogen growth

Each solutions were prepared separately. For 40 ml preparation, 5.44 g of sodium formate and 8.16 g sodium acetate were dissolved separately in 30 ml of distilled water. Once dissolved, the solutions were made up to the final volume of 40 ml with distilled water and bubbled with O_2-free N_2 for 30 min. Solutions were then passed through 0.2 μm pore size Millex GP sterile syringe filters (Millipore) using sterile syringes and needles into a sterile (autoclaved) serum vial that had been flushed with nitrogen gas and sealed with butyl rubber stoppers. This solution was used in co-culture experiments (chapter 4).

2.5 Culture experiments

2.5.1 Substrate utilisation test

Substrate utilisation tests were performed using strains of *Sharpea* and *Kandleria*. The cultures were grown in Hungate tubes. First, anaerobic substrate solutions were prepared as described in section 2.4.6. Stock cultures (frozen at −80 °C; AgResearch Rumen Microbiology collection; see section 4.2 for details) of four strains each of *Sharpea* and *Kandleria* were revived in RM02 medium containing 2GenRFV followed by two rounds of passaging every 24 hours. For the actual utilisation tests, NoSubRFV (0.5 mL) was added to Hungate tubes containing 9 mL of RM02 medium, then single substrates were added into those tubes to achieve the desired concentration (see Table 4.1 for details). To inoculate these, 0.5 ml of a culture of *Sharpea* or *Kandleria* was added into tubes. Each substrate treatment for each strain was conducted in triplicate with one blank uninoculated control tube and triplicate cultures with NoSubRFV but no added substrate. All tubes were incubated by shaking at 39 °C for 5 days and the optical density at 600 nm was recorded every 24 h for 4 days by inserting the tubes directly into a Spectronic 200 spectrophotometer (Thermo-Fisher Scientific Inc.). The spectrophotometer was set to zero absorbance using the blank uninoculated tube.

2.5.2 Co-culture experiment

A co-culture experiment was carried out by growing *Methanobrevibacter olleyae* (a hydrogen- and formate-utilising methanogen) in combination with two strains each of *Sharpea* or *Kandleria* as hydrogen and/or formate producer (see Table 4.4 for strain details). *Methanobrevibacter olleyae* strain 1H5-1P (DSM 16632) was purchased from the DSMZ, Braunschweig, Germany. This experiment used serum vials in two separate

sub-experiments. In the first experiment, *M. olleyae* was added to fully-grown cultures of *Sharpea* or *Kandleria* and in the second experiment *Sharpea* or *Kandleria* was added to fully grown *M. olleyae* cultures. In the first experiment *Sharpea* or *Kandleria* strains were grown for three days in 45 ml RM02 medium containing 10 mM fructose and 2.5 ml NoSubRFV (section 2.4.6), then *M. olleyae* (2.5 ml of a grown culture) was added. Samples for volatile fatty acid analysis were collected after *Sharpea*/*Kandleria* inoculation and growth, then again when *M. olleyae* was added and after one week of *M. olleyae* growth. In the second experiment, *M. olleyae* was grown for 4 days and then *Sharpea* or *Kandleria* (0.5 ml of a grown culture) was added to the vials along with fructose to a concentration of 10 mM. Samples were collected after *Sharpea*/*Kandleria* inoculation and after 72 h of growth of *Sharpea*/*Kandleria* with *M. olleyae*.

For these experiments, inocula of *Sharpea* and *Kandleria* were grown in RM02 media supplemented with NoSubRFV and 10 mM fructose. Inocula of *M. olleyae* were grown on RM02 media supplemented with 60 mM sodium formate and 2 mM sodium acetate. All incubations were at 39 °C, and the vials were shaken on an Orbitek XL orbiting platform (Infors HT, Basel, Switzerland) at 50 rpm.

2.5.3 End product analysis using gas chromatography

Samples (2 ml) were collected from substrate utilisation and co-culture experiments for volatile fatty acid analysis. Samples were centrifuged at 5000 g for 5 min and supernatant fractions were collected and filtered through cellulose-free sterile syringe filters (0.22 μm pore size; Millipore) and stored at –20 °C until used. Substrate and product concentrations were measured using high-performance liquid chromatography (HPLC; LC10AVP, Shimadzu Scientific Instruments, Columbia, MD, USA). The column was an Aminex HPX-87H column (dimensions 300 × 7.8 mm; Bio-Rad, Miami, FL, USA), the temperature of 45°C, 5 mM sulfuric acid as the mobile phase at a flow rate of 0.8 mL per minute and coupled to quantification using a RID 10A refractive index detector (Shimadzu Scientific Instruments). The injection volume was 50 μL. Standards containing fructose, lactate, acetate, formate, propionate, butyrate and ethanol were prepared over a concentration range of 2.5 mM to 20 mM, and used to prepare standard calibrations for converting the detector outputs to concentrations. The calibrations showed linear relationships from 0 to 20 mM.

2.5.4. Enzymatic determination of D- and L-lactate

The D-Lactic Acid Assay Kit and L-Lactic Acid Assay Kit (Megazyme Inc., Bray, Ireland) were used for measurements of D- and L-lactate concentrations, respectively. All samples were diluted to yield a lactic acid concentration up to 0.30 g/L, the linear range of the assay. The procedure was that described by the manufacturer for the microplate assay procedure with a 224 µL reaction volume.

2.6 Growth media

Anaerobic media were prepared either in Hungate tubes or in serum bottles whereas aerobic media were prepared in Schott bottles.

2.6.1 RM02 growth media

RM02 was normally used as the base medium to grow most rumen bacteria, as it partially mimics the rumen conditions (Kenters et al., 2011). The following components were dissolved in 950 ml of distilled water; 1.4 g K_2HPO_4, 0.6 g $(NH_4)_2SO_4$, 1.5 g KCl, 1 ml selenite/tungstate solution (section 2.4.2), 1 ml trace element solution SL10 (section 2.4.1) and 4 drops of resazurin solution (0.1% w/v in distilled water). The solution was boiled and then cooled in an ice bath, while being bubbled with CO_2 (O_2-free). Once at room temperature, 4.2 g $NaHCO_3$ and 0.5 g of L-cysteine·HCl·H_2O was added to the medium which was then dispensed into Hungate tubes (9 mL/tube) while being gassed with CO_2. The tubes were sealed with butyl rubber stoppers and plastic screw caps, and sterilised by autoclaving. Autoclaved medium was stored in the dark for at least 24 h before use. Additives including substrates were added to the medium as required. The complete medium was then incubated at 37 °C for at least 12 h. If medium was prepared in serum bottles then 45 ml medium was dispensed while being gassed with CO_2 and sealed with a butyl rubber stoppers and aluminium crimp caps.

For preparation of anaerobic agar medium, 15 g agar were added to this medium, which was then prepared in serum bottles. Once autoclaved, these were poured into 90-mm diameter polystyrene Petri dishes in an anaerobic chamber (Coy Laboratory Products, Inc., Grass Lake, MI, USA), then transferred to an Oxoid Anaerobic Jar (Thermo Fisher Scientific) for incubation at 39 °C. The gas phase in the anaerobic chamber and anaerobic jars was 97% CO_2 and 3% H_2 (v:v).

2.6.2 LB (lysogeny broth) liquid and solid medium

LB medium was prepared aerobically and used for cultivation of *E. coli* strains for DNA cloning purposes. It was prepared by dissolving 10 g Bacto-tryptone, 5 g yeast extract and 10 g NaCl in 900 ml of distilled water. The pH was adjusted to 7.0 (if needed) and the solution made up to 1 litre with distilled water and, then autoclaved in Schott bottles. To prepared LB agar, 1.5% (w/v) bacteriological agar was added before autoclaving. After autoclaving, LB medium was cooled to 55°C in a water bath before adding any additional components (antibiotics, X-gal [5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside] and IPTG [isopropyl β-D-1-thiogalactopyranoside]), then poured into Petri dishes and allowed to set.

2.7 DNA cloning methods

2.7.1 Reagents preparation

2.7.1.1 Phosphate buffered saline (PBS)

For 1 L PBS, 8 g NaCl, 0.2 g KCl, 1.44 g $Na_2HPO_4 \cdot 2H_2O$, 0.24 g KH_2PO_4 were dissolved individually in distilled water. Once dissolved, all solutions were added together and pH was adjusted to 7.4 with NaOH or HCl if necessary and the solution was made up to 1 litre with distilled water, then autoclaved in Schott bottles.

2.7.1.2 50 × Tris-acetate-EDTA (TAE) buffer

50 × TAE buffer contained 2 M Tris, 1 M glacial acetic acid and 50 mM EDTA were dissolved in distilled water, and the pH was adjusted to 8.0 with 1 M NaOH. The solution was made up to 1 litre with distilled water, then autoclaved in Schott bottles.

2.7.2 DNA extraction from various sample types

Three different DNA extraction methods were used to extract DNA from rumen fluid, concentrated *Quinella* cells and cultures.

1. This DNA extraction method was used in combination with the QIAquick PCR purification kit (Qiagen, Venlo, The Netherlands). Thirty mg of freeze-dried rumen fluid samples were ground to a fine powder (using a coffee grinder), then 30 mg of the powder was added to a bead beating vial (1.8 ml capacity, closed with a screw cap), containing 0.4 g zirconium beads (0.3 g of 0.1 mm and 0.1 g

of 0.5 mm diameter) that had been sterilised by autoclaving. 200 µl of 20% SDS (sodium dodecyl sulfate), 282 µl of buffer A (200 mM NaCl, 200 mM Tris (pH 8.0) and 20 mM EDTA), 268 µl of PB buffer (QIAquick PCR purification kit) and 550 µL of phenol:chloroform:isoamylalcohol (25:24:1, v:v:v) were added to the sample. Bead beating was performed in a Biospec mini bead beater (Bartlesville, OK, USA) for 4 min at maximum speed. The mixture was then centrifuged at 13, 000 g at 4 °C. 350 µl of supernatant was carefully pipetted into a sterile 1.5 ml plastic tube and mixed with 650 µl of PB buffer. 800 µl of the mixture was added to a QIAquick PCR purification column and then vacuum was applied to draw the liquid through the column. This step was repeated with the remainder of the sample. The column containing the bound DNA was washed three times with 900 µl of PE buffer (QIAquick PCR purification kit) and then drained on a paper towel for 10 minutes. DNA was then eluted from the column into collection tubes by adding 80 µl of EB buffer (QIAquick PCR purification kit) while applying a vacuum for 5-7 minutes.

2. Genomic-tips (Qiagen) were also used for extraction of genomic DNA from concentrated *Quinella* cells. *Quinella* cells, suspended in PBS (section 2.12.3), were first pelleted by centrifugation at 5,000 g for 7 minutes then snap-freeze-grinding of the cell pellet was done in liquid nitrogen using a mortar and pestle. The snap-freeze grinding step was repeated for 8 rounds to ensure cell lysis and then the final ground material was collected from the mortar using 11 ml of buffer B1 (containing 22 µl of 1 mg/ml RNase A). Subsequent DNA extraction was done according the manufacturer's instructions for bacteria.
3. DNA extraction from pure cultures of *Sharpea* and *Kandleria* was performed using the InstaGene matrix extraction procedure (Bio-Rad Laboratories Inc., Hercules, CA, USA). First 1.8 ml of culture was pelleted by centrifugation at 14,000 *g* for 1 min in a sterile 2 ml microfuge tube. 200 µl of InstaGene matrix was added to the pellet and incubated for 30 min at 56 °C, then mixed by vortexing at high speed for 10 s. After centrifugation at 12,000 *g* for 5 min to remove cell debris, the supernatant containing DNA was collected and stored at −20 °C for further use.

2.7.3 DNA quantification

DNA quantity and quality was checked using two different methods: 1) the Qubit fluorometer (Invitrogen, Carlsbad, OR, USA) using a Quant-iT dsDNA High sensitivity (HS) assay kit (Invitrogen), and (2) the NanoDrop ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), each following manufacturer's instructions.

2.7.4 Polymerase chain reaction (PCR)

A Mastercycler pro S (Eppendorf AG, Hamburg, Germany) was used for all PCR amplifications. Oligonucleotide primers were ordered from either IDT (Custom Science, Auckland, New Zealand) or Macrogen (Macrogen Inc. Seoul, Korea). PCR details, reagents and primers used are described in Table 2.1.

2.7.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used routinely to check PCR products. Agarose (Sigma-Aldrich, St Louis, MO, USA) was dissolved in 1 × TAE buffer to a final concentration of 1% (w/v) by boiling. Once the mixture was slightly cooled (around 55 °C), SYBR safe nucleic acid dye (Life Technologies, Carlsbad, OR, USA) was added as recommended by the manufacturer to the dissolved agarose and the mixture was poured into gel tray and left to set for 45 min. The gel was then transferred into a Wide Mini-Sub Cell GT electrophoresis system (Bio-Rad Laboratories Inc., Miami, CA, USA) or an Owl A2 Large Gel System (Thermo Fisher Scientific) containing 1 ×TAE buffer (section 2.7.1.2). DNA samples were first mixed with Orange G loading buffer (Sigma-Aldrich) at a final concentration of 20% (v/v) and then loaded into precast wells in the gel. Electrophoresis was performed at 60-100 V for 45-60 min. The bands were then visualised using UV trans-illumination and images recorded using a Gel Logic 200 imaging system (Eastman Kodak, New York, NY, USA).

2.7.6 Clone library construction

Clone libraries were constructed to generate long-length 16S rRNA gene sequences of *Quinella*, *Sharpea* and *Kandleria*. These libraries were also used to estimate the relative abundance of *Quinella* in samples prepared for metagenomics experiments, and to generate 16S rRNA gene-containing DNA fragments from samples prepared for metagenomics experiments. DNA was extracted using one of the methods described in

section 2.7.2. Specific regions of 16S rRNA genes were amplified using the PCR conditions listed in Table 2.1. PCR products were checked by agarose gel electrophoresis for quantity and specificity and then purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) following the manufacturer's instructions. Purified PCR products (3-4 µl) were ligated into the pCR 2.1 TOPO cloning vector (TOPO-TA cloning kit, Life Technologies), following the manufacturer's instructions. Ligated plasmids were then chemically transformed into One Shot TOP10 competent cells (Life Technologies) following the manufacturer's instructions. Transformed cells were plated onto LB agar plates containing ampicillin (50 µg/ml), 40 µl of X-Gal (40 mg/ml) and 40 µl of 100 mM IPTG for blue-white colony selection. White colonies were picked as positive and streaked on to LB agar plates containing ampicillin (50 µg/ml).

2.7.6.1 Colony PCR and sequencing

Colony PCR was performed to test whether the transformed cells contained a plasmid with the expected insert size. For colony PCR, a streaked colony from an LB-ampicillin plate was used as DNA template and transferred to a PCR master mix (contained all essential PCR components except primers and DNA template) using a sterile toothpick. GEM2987f (10 pmol/µL) and TOP168r (10 pmol/µL) primers were used to amplify the cloned fragment (Table 2.1). PCR products were then checked by agarose gel electrophoresis for quantity and size. Clones with the expected inserts size were grown in LB medium containing 50 µg/ml of ampicillin. These clones were then stored at −80 °C in LB medium supplemented with sterile glycerol at final concentration of 50% for future use. Colony PCR products of the expected size were sequenced using the universal bacterial primer 514r (Table 2.1). Full length sequences of cloned inserts were obtained using region-specific universal bacterial primers (514r, 518f, 800r and 968f; see Table 2.1) targeting different locations of 16S rRNA gene. All sequencing was carried out either at the Massey Genome Sequencing Service (Massey University, Palmerston North, New Zealand) or at Macrogen Inc. Seoul, Republic of Korea. Sequencing samples were prepared as mixes of purified colony PCR product:sequencing primer (50 ng:10 pmol in 20 µl reaction volumes) for the samples that were sequenced at the Massey Genome Sequencing Service, whereas 55 µl colony PCR samples were sent by courier when sequenced at Macrogen. Resulting sequences were then aligned using Geneious 8.1 software (Biomatters Ltd.) to assemble the full length 16S rRNA gene sequence for a particular clone.

Table 2.1 Primers used for amplification and sequencing of DNA and plasmid fragments.

Purpose	Primers	Primer sequence (5' - 3')	Target gene	Fragment size	PCR conditions
Amplification of bacterial 16S rRNA gene	27f[a]	GAGTTTGATCMTGGCTCAG	Partial 16S rRNA	~1465 bp	94 °C for 4 min 94 °C for 1 min 55 °C for 1 min 72 °C for 1 min (25 cycles) 72 °C for 10 min
	1492r[a]	GGYTACCTTGTTACGACTT			
Amplification of insert from TOPO 2.1 plasmid vector	GEM2987f[b]	CCCAGTCACGACGTTGTAAAACG	TOPO 2.1 vector and insert sequence	varied	94 °C for 4 min 94 °C for 15 s 55 °C for 30 s 72 °C for 1 min (30 cycles) 72 °C for 7 min
	TOP168r[b]	ATGTTGTGTGGAATTGTGAGCGG			
Sequencing partial bacterial 16S rRNA genes	514r	CCG CGG CKG CTG GCA C	Partial 16S rRNA	not applicable	–[c]
	518f	CCAGCAGCCGCGGTAATACG			
	800r	TACCAGGGTATCTAATCC			
	968f	AACGCGAAGAACCTTAC			
	1100r	GGGTTGCGCTCGTTG			
Sequencing inserts in from TOPO 2.1 plasmid vector	M13f	GTAAAACGACGGCCAGT	TOPO 2.1 vector and insert sequence	not applicable	–[c]
	M13r	GCGGATAACAATTTCACACAGG			

[a, b]Used as pairs for amplification of defined products.

–[c], used for sequencing only.

2.7.6.2 Clone analysis

DNA sequences were processed by trimming vector sequences and bad quality sequences (with ambiguous electropherogram base calls from the 5' and 3' ends). Resulting sequences were used to query an in-house refined 16S rRNA bacterial database (Henderson et al., 2017) using the BLAST sequence similarity algorithm in QIIME (Caporaso et al., 2010) to find their closest relative. Sequences of interest were then aligned to the whole database using the SINA aligner (Pruesse et al., 2007) and imported into ARB (Ludwig et al., 2004) using the ARB parsimony (quick add mark) insertion function for further analysis.

2.8 Pyrosequencing analysis to assess bacterial community composition

The study presented in chapter 3 used the dataset of partial bacterial 16S rRNA gene sequences generated from 236 rumen samples using 454 Titanium pyrosequencing by Kittelmann et al. (2014). These sequence data have been submitted to the EMBL database under the study accession number ERP003779. QIIME (Caporaso *et al.*, 2010) pipeline v1.4.0 and v1.5.0 were used for pyrosequencing analysis. Pyrosequencing reads were first checked for quality using FastaQC and assigned to their respective biological samples using nucleotide barcodes. Only sequences >400 bp with a quality score over 27 (sliding window 50 bases) along the whole sequence were included for analysis. The pyrosequencing dataset was denoised using Acacia (Bragg *et al.*, 2012). Denoised sequences were grouped into OTUs (Operational Taxonomic Units) and from each OTU, one representative sequence was selected and designated the repset sequence for that OTU. These repset sequences were compared using BLAST against reference sequences in an improved bacterial taxonomic framework (Henderson et al., 2017), which consist of, a refined database of publicly available 16S rRNA gene sequences from rumen bacteria. In this step, each repset sequence was assigned to its closest relative in the taxonomy framework. In this way, the sequences in each OTU were assigned to a bacterial taxon, with multiple OTUs with highly similar repset sequences being assigned to the same taxon. The bacterial genus level was chosen as the taxonomic rank to summarise the repset data. Beta-diversity matrices were calculated to compare the communities-types based on the taxonomic and phylogenetic assignments. The community differences among the samples were visualised using Principal Coordinate

Analysis (PCoA) plots. For further details see 454 Overview Tutorial: de novo OTU picking and diversity analyses using 454 data (http://qiime.org/tutorials/tutorial.html).

2.9 Identification of taxa associated with samples from high and low CH_4 sheep

To analyse the bacteria and bacterial communities in rumen samples from high and low methane sheep, different approaches were used. First, Principal Coordinate Analysis (PCoA) was performed using the χ^2-distance metric to understand the sample distribution pattern on the basis of taxonomically assigned OTUs and sample category (high and low methane). Later, low abundance taxa were also identified and eliminated from the dataset to concentrate on key bacterial taxa (described in section 3.2.3.1). Statistical tests were performed to identify the taxa and OTUs that are significantly associated with high and low methane sheep samples and bacterial community types (see chapter 3):

A. Spearman's rank correlations were used to understand the significance of the correlation of taxon abundance with methane yield.

B. Kruskal-Wallis tests (Kruskal and Wallis, 1952) were performed to make significant comparisons between the different bacterial community types. A p-value <0.05 was considered significant.

C. Sparse Partial Least Squares – Discriminant Analysis (sPLS-DA) was performed in R (Cao et al., 2016) to identify diagnostic OTUs associated with low methane community types. A subset of discriminating OTUs was identified by selecting the 200 most discriminating OTUs from each of dimensions 1 and 2, using the "Variable Importance in the Projection" (VIP) function. These were combined to give 255 unique OTUs. Assessment of sPLS-DA via cross-validation methodology was carried out using the "leave-one-out" accuracy method. All of these analyses were performed using the mixomics package (Gonzalez et al., 2012).

2.10 Phylogenetic analysis of *Quinella*, *Sharpea* and *Kandleria*

For phylogenetic analysis, clone libraries of full-length 16S rRNA gene sequences of *Quinella*, *Sharpea* and *Kandleria* were prepared (method described in section 2.7.6) and the sequences used in combination with pre-existing sequences from SILVA database version 119 (Quast et al., 2013). First, high *Quinella*, *Sharpea* and *Kandleria*-abundant rumen content samples were identified from the stored samples from the study of Kittelmann et al. (2014), based on the abundances predicted from the pyrosequencing

data. DNA was isolated from these samples and PCR amplification of bacterial 16S rRNA gene sequences was conducted using the primers 27f and 1492r to amplify almost full length bacterial 16S rRNA gene sequences (Table 2.1), prior to their cloning. Clones were sequenced for identification using the 514r bacterial primer, and full-length sequences of selected clones were generated by assembling reads that were generated using various universal bacterial and plasmid primers (27f, 514r, 1100r, 1492r and M13r). The full-length sequences were first checked for chimeras using Bellerophon (Huber et al., 2004) and UCHIME (Edgar et al., 2011), and the fractional treeing method. In the fractional treeing method, 450 bp of sequence from each end (3′ and 5′) of the 16S rRNA gene was used to generate two trees, one using sequences from the 5′ end and another using 3′ end sequences. The sequences with conserved positions in both trees were taken as non-chimeric sequences. These full-length clone library sequences together with pre-existing nonchimeric sequences of *Quinella*, *Sharpea* and *Kandleria* were used for taxonomic refinement. Sequences were first aligned to entries in the refined SILVA database (Henderson et al., 2017) using the SINA aligner and then imported into ARB. In ARB, different treeing algorithms were used to build a stable phylogeny for *Quinella*, *Sharpea* and *Kandleria.* Additionally, a maximum likelihood phylogeny was implemented in RAxML version 8 (Stamatakis, 2014) using the GTRGAMMA nucleotide substitution model with rapid bootstrap analysis, to confirm the position of each sequences in the tree. Trees were rooted using the 16S rRNA gene sequence of *Fibrobacter succinogenes* (FibSuc43, GenBank accession CP002158). The resulting RAxML tree was imported into ARB (Ludwig et al., 2004). Clusters of sequences that generally had bootstrap support greater than 70% were identified and defined at species and genus levels based on average sequence identities within each cluster (>97% for species; >93% for genera). More details on the definition of genera and species are given in chapter 3. The final sequences used in the tree were deposited in NCBI with GenBank accession numbers MF184869 to MF184922.

2.11 Fluorescence *in situ* hybridisation (FISH)

2.11.1 Rumen samples

The sheep used were part of a long-term experiment to identify and breed low-CH_4 emitting sheep (Pinares-Patiño et al., 2013b). They were fed lucerne pellets twice daily and rumen samples were collected by stomach tubing 2 h after the morning feeding.

2.11.2 Rumen fluid fixation and preservation for FISH

Rumen samples were kept on ice immediately after collection and transported to the laboratory. The samples were fixed in two different ways: one in 4% paraformaldehyde (PFA) (w/v) solution:sample (1:3), and second in absolute ethanol:sample (1:1). PFA-fixed samples were used in FISH experiments to identify *Quinella* in rumen samples by microscopy, while ethanol-fixed samples were used for FISH combined with single cell sorting of *Quinella*. Post sampling processing steps for both types of fixation were the same except that the ethanol-fixed samples were processed immediately on arrival into the laboratory and PFA-fixed samples were kept at 4 °C for 2 h. Samples were centrifuged for 8,000 g for 5 min. Supernatants were discarded and the pellets were washed twice in PBS buffer by repeating the centrifugation step. Finally, pellets were diluted in 750 µl of PBS and pure ethanol (1:1) and stored at –20 °C.

2.11.3 FISH probe design

Sequences from the SILVA database (refined by Henderson et al. (2017) in combination with 16S rRNA clone library sequences of *Quinella* were used as reference sequences to design FISH probes. Initially the ARB software Probe Design Tool was used to find *Quinella* specific probes but this was unsuccessful. Then, 16S rRNA gene sequences from *Quinella* were inspected manually in MEGA 7 (Kumar et al., 2016) for signature conserved 18-20 bp length sequences. Shortlisted probe sequences were tested using the ARB Probe Match Tool to confirm specificity. Specific probes with exact matches to *Quinella* sequences were synthesised by IDT (Custom Science, Auckland, New Zealand) with three different 5' fluorochromes (Cy3, Cy5 and Alexa488). Details are given in Table 5.1. A bacterial-specific domain-level probe EUB338 (5´ GCTGCCTCCCGTAGGAGT; target site *E. coli* position. 338-355), and a nonsense probe nonEUB338 (the reverse and complement of EUB338) were used as positive and negative controls respectively. Probes were aliquoted (2.5 µg) in 200 µl microcentrifuge tubes and stored at –20 °C. Before use, probes were resuspended in 50 µL of sterile nuclease-free water to make a working concentration of 50 ng/µL.

2.11.4 FISH probe testing

Newly designed probes were tested using PFA-fixed rumen fluid samples. Probe stringencies (conditions at which probes specifically hybridised only to *Quinella*-like

cells) were optimised by varying formamide concentration, NaCl concentration and hybridisation temperature. The rest of the procedure was as described by Hugenholtz et al. (2001) with slight modifications. Briefly, 3-5 µL (depending upon cell concentration) samples of fixed cells were applied on 10-well FISH slides and air dried for at least 3 h or overnight. Slides were then dehydrated using a series of ethanol washes starting at 50%, 80% and finally 100% ethanol for 3 min each and finally air dried. A hybridisation oven was pre-warmed to the desired hybridisation temperature, e.g. 46 °C, 50 °C or 52 °C. Hybridisation buffers (8 µL; see Table 2.2) with various formamide and NaCl concentrations were applied to different slides in duplicate. Probes (1 µL/well of 50 ng/µL) were applied either singly or in combination to different wells and mixed gently with a micropipette tip without touching the well surface. Each slide was then transferred carefully into a 50 mL Falcon tube containing moistened paper towels, capped firmly and placed horizontally in the hybridisation oven for 2 h. After hybridisation, slides were well rinsed with the same pre-warmed wash buffer (Table 2.2) to remove unhybridised probes. Slides were then transferred into Falcon tubes containing wash buffer (Table 2.2) and placed in a waterbath (at a temperature 2 °C higher than the hybridisation temperature) for 15 min. Slides were then rinsed briefly with ice-cold distilled water and dried immediately either with compressed air or in a 39 °C oven. Antifade mounting solution (VECTASHIELD; Vector Laboratories Ltd., Burlingame, CA, USA) was applied carefully on the slides and covered with a coverslip without trapping air bubbles in the wells. Slides were then visualised by epifluorescence microscopy using the appropriate filters (for each probe's fluorochrome).

2.11.5 Clone FISH

Clone FISH (Schramm et al., 2002) was performed to test the newly-designed FISH probes. 16S rRNA genes clones (i.e. *E. coli* containing pCR2.1 plasmid with a *Quinella* 16S rRNA gene insert) with exact sequence matches were used as positive controls whereas cloned sequences with one or more mis-matches were used as negative controls. These were clones generated as described in section 2.7.6. First, the inserts and vectors were sequenced to ensure they had the correct sequence orientation (5'-3') using primer 514R that targets a region flanking the cloning site of the vector (Table 2.1), so that the RNA generated contained the correct probe binding sites. Clones with the wrong orientation (5'-3') were sub-cloned to get the correct orientation. Briefly, plasmids were extracted from *Quinella* clones (i.e., *E. coli* containing pCR2.1 plasmid with *Quinella*

16S rRNA gene inserts) and 16S rRNA gene sequences were amplified using 27f and 1492r primers (Table 2.1). Amplified sequences were then ligated into the plasmid (pCR2.1) and transformed in *E. coli* JM109 (DE3). Clones were then screened again by sequencing for examples with the desired orientation. Clones with the correct orientation were then grown in LB medium containing ampicillin, until they achieved an optical density of 0.4 (600 nm), then 100 µl of 100 mM IPTG were added to 10 ml of grown cells and incubated at 37 °C for 1 h. 50 µl (34 mg/ml) of chloramphenicol was then added to the cells which were further incubated for 3 h at 37 °C. This process was used to increase the copy number of transcribed 16S rRNA to increase the FISH probe targets so as to enhance hybridisation signal intensity (Schramm et al., 2002). Methods described in section 2.11.4 were followed to determine conditions for probe stringency.

2.11.6 Liquid FISH

A liquid FISH method was developed using concentrated *Quinella* samples (section 2.12.3) and probe Quin1231. FISH probe stringency was tested using the protocol described in section 2.11.4 with some changes. Here all steps were conducted in 1.5 mL microfuge tubes. First, 800 µL of TE buffer (10 mM Tris and 1 mM EDTA, pH 7.5 with HCl) was added to 200-500 µL (depending on cell density) of ethanol-fixed concentrated *Quinella* cells and centrifuged at 8,000 *g* for 5 min at 20 °C. The supernatant was discarded and the pellet was washed twice with 1 mL TE buffer by re-suspending it and then centrifuging it at 8,000 *g* for 5 min at 20 °C and discarding the supernatant. The pellet was then re-suspended in 1 mL of 0.22 µm-filtered TE-lysozyme (1 mg/mL) solution and incubated for 10 min at room temperature. The mixture was centrifuged at 8,000 *g* for 3 min at 4 °C and the supernatant was removed. The pellet was then washed again by adding 1 mL of PBS and centrifuging at 8,000 *g* for 3 min at 4 °C and discarding the supernatant. The pellet was then re-suspended in 500 µL of hybridisation buffer by gentle mixing. 70 µL aliquots of this homogenised suspension were transferred to 1.5 mL plastic tubes and 7 µL of the desired probe (50 ng/µL) was added into each tube. Hybridisation was carried out at the desired temperature for 2 h. Remaining hybridisation buffer was kept at the same temperature whereas the wash buffer at 2 °C warmer. After hybridisation, 150 µL of the pre-warmed hybridisation buffer was added to each tube and centrifuged at 8,000 *g* for 3 mins at 20 °C. The supernatant was removed and 200 µl of the pre-warmed wash buffer was added. The pellet was re-suspended by pipetting and then incubated at the wash buffer temperature for 20 min. Wash buffer was then removed

by centrifuging at 8,000 *g* for 3 min at 20 °C. Finally, the pellet was re-suspended in 100 µL of PBS and hybridisation with the probe was checked by epifluorescence microscopy using various filters (appropriate for each probe's fluorochrome). A hybridisation temperature of 46 °C and a wash temperature of 48 °C was found to result in optimal *Quinella*-specific probe hybridisation to large *Quinella*-like cells, without binding to other cell types.

Table 2.2 Hybridisation and washing buffer preparation for FISH.

	Formamide concentration (v/v, %)				
	0	20	40	60	80
Hybridisation buffer	Volume (µL)				
5 M NaCl	360	360	360	360	360
1 M Tris-HCl, pH 8.0	40	40	40	40	40
100% formamide	0	400	800	1200	1598
Sterile Milli-Q water	1598	1198	798	398	0
10% (w/v) SDS	2	2	2	2	2
Total volume	2000	2000	2000	2000	2000
Wash buffer	Volume (µL)				
5 M NaCl	9000	2150	460	40	0
1 M Tris-HCl, pH 8.0	1000	1000	1000	1000	1000
0.5 M EDTA	0	500	500	500	175
Sterile Milli-Q water	39950	46300	47990	48410	48775
10% (w/v) SDS	50	50	50	50	50
Total volume	50000	50000	50000	50000	50000

2.12 Experiments to provide better understanding of *Quinella* morphology and physiology

2.12.1 Sample collection

Three rounds of rumen fluid sampling were conducted from sheep fed on lucerne pellets. The samples were taken by stomach tubing. Sample 1 was made by pooling rumen fluid samples from 12 sheep with high relative abundances (as assessed by phase contrast microscopy) of *Quinella*-like cells. Sample 2 was pooled rumen fluid samples (two sampling days apart) from one sheep with a high relative abundance of *Quinella*-like cells. Sample 3 was collected from a single sheep with a high relative abundance of *Quinella*-like cells. All above samples were processed as described in section 2.12.3. Sample 4 was prepared from sample 1 by sorting single cells (section 2.14) of probable *Quinella* into batches of 50 cells using the fluorescence signal from probe Quin1231 after hybridisation (section 2.11.6).

2.12.2 Isolation of *Quinella*-like cells from sheep rumen samples

Isolation of *Quinella* was attempted from rumen samples identified as having high *Quinella* abundance (by examining fresh samples by phase contrast microscopy and by FISH). Rumen fluid containing *Quinella* cells was directly collected into the RM02 tubes while being bubbled with CO_2 (O_2-free). Tubes were then transported to the laboratory and kept at 39 °C until they were used within 1.5 h. Hungate tubes containing anaerobically prepared RM02 medium containing 2GenRFV plus (final concentrations) 20 mM mannitol, 10 mM salicin and 0.08 % [w/v] pectin was used for first inoculation. The dilution to extinction method (Kenters et al., 2011) was used to isolate *Quinella* from rumen samples. Briefly, collected samples were diluted and inoculated in two sets so that in set one each tube received an estimated 10 bacterial cells and in set two each received an estimated 40 bacterial cells, based on estimates of 10^9 bacterial cells per ml of rumen contents.

Tubes that showed any turbidity were examined microscopically and sub-cultured into fresh media. Tubes that were identified as having large cells were used for spread plating onto anaerobic agar media containing similar additives as before, plus 1.5 g bacteriological agar per 100 ml of media. Agar plates were checked every 24 h for colonies. Colonies were examined microscopically, and those with larger cells were used

for FISH and 16S rRNA gene sequence-based identification (sections 2.11.4 and 2.7.6) to identify *Quinella*.

2.12.3 Method development for *Quinella* concentration from rumen fluid

All rumen fluid samples (section 2.12.1) were transported to the laboratory on ice and squeezed through 300 µm nylon mesh. The filtrate was diluted 1:1 with PBS buffer and left at room temperature for 10 min. The diluted filtrate was then filtered through 23 µm nylon mesh into clean sterile 50 mL falcon tubes (40 mL sample/tube) and centrifuged at 100 *g* for 5 min at room temperature to pellet protozoa and particulate materials. The supernatant was decanted, filtered again through 23 µm nylon mesh and centrifuged at 800 *g* for 5 min. The supernatant was discarded and the pellet (containing potential *Quinella* cells) was washed three times by re-suspending in 40 mL PBS buffer and then filtered through 23 µm nylon mesh and centrifuged at 800 *g* for 5 min. *Quinella* abundance was monitored at each step by visualising samples using light microscopy (with a 100 × oil immersion lens) for large oval *Quinella*-like cells. Finally, cells were re-suspended in 1x PBS and stored at –20 °C until used.

2.13 Electron microscopy of *Quinella*

Sample 3 (section 2.12) was used for *Quinella* cell concentration by the method described in section 2.12.1. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) analyses of these concentrated cells were carried out at the Manawatu Microscopy and Imaging Centre (MMIC), Massey University, Palmerston North, New Zealand.

2.13.1 Scanning electron microscopy (SEM)

The cells were fixed in modified Karnovsky's fixative (3% [v/v] glutaraldehyde, 2% [v/v] formaldehyde in 0.1 M sodium phosphate buffer, pH 7.2) in a 1:6 ratio (sample:fixative) for at least 8 h at room temperature. Samples were centrifuged at 1,600 *g* for 4 min and a drop of pelleted sample re-suspended in a small amount of fixative solution was sandwiched between two membrane filters (0.4 µm, Isopore, Merck Millipore Ltd, Billerica, MA, USA) in an aluminium clamp and washed three times in 0.1 M sodium phosphate buffer (pH 7.2, 15 min each). The sample was then dehydrated in a series of ethanol solutions (25%, 50%, 75%, 95%, and 100% [v/v]) in water for 15 min each, with and a final 100% ethanol dehydration for 1 h. The samples were then dried to critical

point using liquid CO_2 as the critical point drying fluid and 100% ethanol as the intermediary (Polaron E3000 series II critical point drying apparatus; Quorum Technologies Ltd., Lewes, UK). The samples were mounted onto aluminium stubs and sputter coated with approximately 100 nm of gold (BAL-TEC SCD 005 sputter coater; Angstrom Engineering Inc., Kitchener, Ontario, Canada) and viewed with a FEI Quanta 200 scanning electron microscope (Philips Electron Optics, Eindhoven, The Netherlands) at an accelerating voltage of 25 kV.

2.13.2 Transmission electron microscopy (TEM)

The *Quinella* cells suspended in PBS (prepared as described in section 2.12.3) were fixed for at least 2 h in modified Karnovsky's fixative (section 2.13.1). Bovine serum albumin (BSA) was then added to the samples to a final concentration of 20% (w/v). Cells were harvested by centrifugation at 1,600 g for 4 min and excess fixative was removed by aspiration with an autopipetter. Three drops of 20% (w/v) BSA were added to the pellet and mixed well to prevent the pellet falling apart during the many changes of fluids through the processing. Samples were centrifuged again at 1,600 g for 4 min and the supernatant discarded. One drop of 25% (w/v) glutaraldehyde in water (Sigma-Aldrich) was added at the top of the pellet and left for about 5 min to coagulate the BSA. The pellet was then removed from tube and cut off the excess BSA gel. The pellet was sliced into thin pieces and returned to the first fixative to ensure all of the BSA was coagulated. The pellet slices were washed by resuspending in 0.1 M sodium phosphate buffer (pH 7.2) followed by centrifugation at 900 g for 10 min. Post fixation treatment was done in 1% (v/v) osmium tetroxide in phosphate buffer for 30 min at room temperature. The washing step was repeated again as above. Pellet slices were then dehydrated with a series of acetone solutions (25%, 50%, 75%, 95% and 100% [v/v]) in water for 10-15 min each followed by two changes of 100% acetone for one hour each. Samples were then transferred to resin:acetone (50:50 v/v) solution and agitated overnight on a PELCO R2 rotary mixer (Ted Pella Inc., Redding, CA, USA) then the resin:acetone mix was replaced with 100% resin for 8 h on the mixer and this step was repeated twice (overnight in 100% resin, then another 8 h in 100% resin). Samples were embed in moulds with fresh resin and cured in a 60 °C oven for 48 h. Light microscope sections were cut at 1 µm thickness using a glass knife on an ultramicrotome (EM UC7; Leica, Wetzlar, Germany) and heat fixed onto glass slides. These heat-fixed sections were stained with 0.05% (w/v) toluidine blue for approximately 12 s and viewed under a light microscope (Axioplan; Zeiss,

GmbH, Germany). The block was then trimmed down to the selected area and cut using a diamond knife (Diatome, Vienna, Austria) at 100 nm. These were stretched with chloroform and mounted on a grid using a Quick Coat G pen (Saiko, Tokyo, Japan). Grids were stained in saturated uranyl acetate in 50% (v/v) ethanol for 4 min, washed with 50% (v/v) ethanol then with Milli-Q water and then stained in lead citrate (Venable and Coggeshall, 1965) for a further 4 min. This was followed by a wash in Milli-Q water. Samples were then dried and viewed with transmission electron microscope (FEI Tecnai G2 Spirit BioTWIN; Brno-Černovice, Czech Republic).

2.14 Sample type 4 processing for metagenome sequencing

Concentrated *Quinella* cells obtained from sample 1 were first ethanol fixed and then hybridised with *Quinella* specific FISH probe (Quin1231) using the liquid FISH method described in section 2.11.6. Then *Quinella* cells were sorted using a MoFlo XDP high speed cell sorter (Beckman Coulter, Indianapolis, IN, USA) by selecting cell size and fluorescence signal as control parameters. Batches of 50 cells that both bound the Quin1231 probe and had cell sizes corresponding to the putative *Quinella* cells were sorted in a 384 well LightCycler® 480 Multiwell Plate, (Roche Molecular Systems, Inc., Pleasanton, CA, USA) and stored at −80 °C.

2.14.1 Whole genome amplification from sorted *Quinella* cells

This process was performed in a biosafety cabinet using UV-sterilised filter tips, multiwell plates, plate seals, and 1.5 mL plastic tubes, and a UV-sterilised autopipetter to eliminate potential contamination.

2.14.2 Cell lysis

Plates containing sorted *Quinella* cells were taken from the −80 °C freezer and subjected to five rounds of freeze-thaw cycles by moving plate from −80 °C for 20 min to 70 °C for 10 min. Plates were then centrifuged at 1000 *g* and 4 °C for 2 min to pellet the cells. 1 µl of lysis buffer D2; 3 µl DTT, and 33 µl buffer DLB (provided in the REPLI-g® Single Cell kit; Roche Molecular Systems, Inc.) were added to each sample before centrifugation at 1000 g and 4 °C for 2 min and incubation at 65 °C for 10 min. 1 µl of STOP buffer (from the REPLI-g® Single Cell kit) was added to the mixture which was then centrifuged at 1000 g and 4°C for 2 min. The plate was then placed on ice while preparing the master mix for the MDA (multiple displacement amplification) reaction.

2.14.3 MDA

A 1.5ml plastic tube was used to prepare the MDA master mix by mixing 8.7 µl reaction buffer, 0.6 µl phi-29 polymerase, 0.3 µl Styo 13 working solution, and SCG-H_2O (single cell grade- H_2O) to make up total 15 µl. All these reagents were provided in the REPLi-g single cell kit (Qiagen). The master mix was added to the lysed cells and centrifuged at 1000 g and 4 °C for 2 min. The plate was sealed with sterile plate seal and incubated at 30 °C for 16 h, then transferred to a 65 °C oven and incubated for 10 min to deactivate the enzyme. A 1:10 dilution of MDA products was made in a sterile biosafety cabinet to a final volume of 10 µl in 8-strip PCR tubes with SCG-H_2O. Diluted MDA products were mixed thoroughly by pipetting up and down, and stored at −80 °C until sequenced. This material was designated sample 4.

2.14.4 Screening and taxonomic identification of MDA products.

MDA products were screened using 16S rRNA gene based PCR. The primers 27f and 1492r were used to generate PCR products for clone library preparation to determine whether the amplified MDA products contained only *Quinella* DNA or not.

2.15 Estimation of *Quinella* cell abundance and sample preparation for metagenome sequencing

The relative abundance of *Quinella* in the cell suspensions (samples 1, 2 and 3) produced by the enrichment method described in section 2.12.3 was estimated by preparing 16S rRNA gene clone libraries. Briefly, DNA was extracted from these samples following methods described in section 2.7.2 (method 2 for sample 1 and method 1 for samples 2 and 3) and quantified (section 2.7.3). Sample 4 (MDA-amplified DNA from single cell sorted *Quinella* cell obtained from sample 3) was processed as described in section 2.14. These DNA were used to generate clone libraries following the method described in section 2.5.5. Cloned products were sequenced and analysed as described in section 2.7.6.2 to estimate the *Quinella* abundance in each sample. After confirming the *Quinella* abundance, DNA extracted from these 4 samples was used to construct bar-coded genomic shotgun libraries and sequenced using a single lane of Illumina MiSeq 2 × 300 bp paired-end reads. The library preparation and sequencing was carried out by Eurofins Genomics GmbH (Ebersberg, Germany).

2.16 Metagenome sequences processing

The sequences generated by Illumina MiSeq 2 were first checked for quality using FastQC (Smith et al., 2010). Poor quality reads were trimmed, if needed, using Trimmomatic.

Command used:

- /trimmomatic-0.32.jar PE -threads 33 -trimlog ...R1_trimlog ...R1_001.fastq ...R2_001.fastq ...R1_001_tmm.fq ...R1_001_s_tmm.fq ...R2_001_tmm.fq ...R2_001_s_tmm.fq SLIDINGWINDOW:10:15 MINLEN:250 CROP:250
- For details see: http://www.usadellab.org/cms/?page=trimmomatic

The resulting sequencing reads were assembled using SPAdes Genome Assembler (Nurk et al., 2013).

Command used:

- bin/spades.py --pe1-1 …R2_001.fastq --pe1-2 ….R2_001.fastq --careful –t 40 -o <output>

  Note: different options were also tried to get better assemblies, but this was the one finally used for the data presented here. For details see http://bioinf.spbau.ru/spades

The Quality Assessment Tool for Genome Assemblies (QUAST) was used to evaluate the genome assemblies (% G+C content, number of contigs, contig length, N50 and N75) (Gurevich et al., 2013).

Command used:

- python quast.py <contig_file(s)> -o <output_dir>
  For details see: http://quast.sourceforge.net/quast

The assembled reads in contigs were binned into potential genome bins using MetaBAT, a tool that uses integrated empirical probabilistic distances of genome abundance and tetranucleotide frequency (Kang et al., 2015) to bin contigs. To run MetaBAT, first an index file was generated for each assembly which was later used for contig binning.

Command used to generate the index file:

- bowtie2-build contigs.fasta contigsname

- bowtie2 -p 10 -x contigsname -U reads_R1.fastq -S contigsname_output.sam > bowtie.out
- samtools_0.1.18 view -Sb contigsname_output.sam > contigsname_output.bam

Command used to run MetaBAT

- …/metabat/runMetaBat.sh <options> assembly.fasta contigsname_output.bam

Note: Different options were also tried to generate better bins. For example:

--verysensitive	for greater sensitivity (combine may other options)
--minContig	minimum contigs with particular size for analysis
--minClsSize	minimum bin size

For details see the https://bitbucket.org/berkeleylab/metabat

2.17 *Quinella* genomic bins analysis

2.17.1 Phylotyping of genomic bins

To understand the initial taxonomic composition of genome bins, AMPHORA2 software (Wu and Scott, 2012), which uses a set of 31 bacterial and 104 archaeal protein coding marker genes, was used on the web-based AmphoraNet application (Kerepesi et al., 2014). Results generated were visualised using AmphoraVizu (Kerepesi et al., 2014) to understand the taxonomic-level composition of genome bins.

2.17.2 Quality assessment of genomic bins

The genome bins generated were assessed for quality (completeness and contamination) using CheckM (Parks et al., 2015). This uses a broad set of markers genes specific to the position in the genome within a reference genome tree.

Command used:

- checkm lineage_wf -t 20 -x fa <path to bin_folder with .fa files> <output_directory>
- checkm taxonomy_wf family Veillonellaceae -x fa <path to bin_folder with .fa files> <output_directory/taxonomy_wf>

The taxonomic identification of each bin was confirmed by extracting 16S rRNA gene sequences and then comparing them by BLAST as a query against a refined 16S rRNA bacterial database (Henderson et al., 2017) using QIIME.

Command used to extract 16S rRNA gene sequences from particular bins:

- checkm ssu_finder -x fa <path to bin_folder with .fa files> <output_directory>

For details about CheckM commands see https://github.com/Ecogenomics/CheckM/wiki

2.17.3 Amplification of near full length 16S rRNA gene sequence of *Quinella* genome bins

To understand the *Quinella* strains captured in each genome bin using 16S rRNA gene sequences as a marker, 16S rRNA genes were amplified from the original samples (1, 2, and 3) targeting the 16S rRNA genes and flanking regions found in the bins. First, contigs containing 16S rRNA gene fragments from *Quinella* genome bins were annotated to identify other genes present on these fragments. Forward primers were designed to target a CDS region of genes adjacent to the 16S rRNA gene sequences. Then these primers were used in conjunction with a universal bacterial reverse primer (1492R) to amplify nearly complete 16S rRNA gene regions together with flanking region sequences that should correspond to the reference contigs. Each forward primer was designed to produced different size products for their targeted genome bin. The length of expected amplified products ranged from 1851 to 2390 bp (Table 2.3). Primer specificity was checked using the Primer-BLAST tool from NCBI (Ye et al., 2012). The resulting PCR products were cloned and the inserts were amplified by colony PCR (section 2.7.6). Resulting product were sequenced using the 514r universal bacterial primer for initial identification of clones then the full amplified product was sequenced using the plasmid-targeted primers M13f, M13r, and the 16S rRNA gene-targeted primer 1100r (Table 2.1). These fragment reads were assembled together in Geneious 8.1. Finally, consensus sequences were generated from each of these assemblies and aligned with the reference contigs from the genome bins. The 16S rRNA gene regions from the resulting sequences were also used for phylogenetic analysis (Figure 5.11) by using sequences from the refined *Quinella* tree (Figure 3.11) to find the phylogenetic position of the 16S rRNA genes corresponding to the *Quinella* genome bins. A maximum likelihood phylogeny was implemented in RAxML version 7.3.2 (Stamatakis, 2014), using the GTRGAMMA nucleotide substitution model with rapid bootstrap analysis. Bootstrap values were calculated from 500 replicates. *Selenomonas ruminantium* 16S rRNA gene sequence was used as an outgroup.

Table 2.3 Primers used to amplify 16S rRNA genes from DNA of *Quinella*-enriched samples.

Quinella genome bin and contig identifier (bin_contig)	Primer	Primer sequence (5'- 3')	Sequence start position	Sequence end position	Primer length (bp)	G+C (mol%)	Melting temperature (°C)	Self-complementarity (bp)	Predicted product size in combination with 1492r[a]
1Q5 _5057819	9,354f	CTCGACGTTCT TAATCTTCG	9,354	9,373	20	45	54.1	4	1851
1Q7 _5055757	19,382f	ACGACGATAA TCCTGTGG	19,382	19,399	18	50	53.8	5	2194
2Q5 _4645198	111,625f	GATACGTCAG GTCATAGC	111,625	111,642	18	50	51.7	4	2026
3Q1 _4968023	24,448f	TGAATCAGCG AATAGAGC	24,448	24,465	18	44.4	51.8	5	2390

[a]Bacterial 16S rRNA gene universal primer (1492r) was used as the reverse primer.

2.18 *Quinella* genome bin annotation and functional analysis

Genome annotation of shortlisted *Quinella* genome bins was performed using the GAMOLA2 annotation tool (Altermann et al., 2017) in combination with the Artemis software suite (Rutherford et al., 2000). Genes were predicted using Prodigal (Hyatt et al., 2010). Genes that play major role in metabolism were translated and the deduced proteins were manually checked and compared to reference proteins using different databases (non-redundant protein database provided by the National Centre for Biotechnology Information (NCBI) (Sayers et al., 2011), clusters of orthologous groups (COG) database (Tatusov et al., 2001b) (Tatusov et al., 2001b) Pfam (Finn et al., 2016) and TIGRFAM (Haft et al., 2003) databases and reviewed sequences from the Uniprotkb/swiss-prot database (Boutet et al., 2016).

2.18.1 Functional genome analysis

To understand the similarity between genome bins, functional genome distribution (FGD) analysis was conducted. This analysis does not represent the phylogenetic distance. Instead, overall genome bin similarities were calculated using ORFeome amino-acid sequences based on the assumption that highly similar proteins are indicative of functional similarities. The similarities and absence of predicted proteins are combined into a pairwise FGD dissimilarity matrix (Altermann, 2012). This was then used to generate a tree by an unweighted pair group method with arithmetic mean (UPGMA) method (Sneath and Sokal, 1962).

OrthoMCL analysis (Li et al., 2003) was also conducted to identify the orthologous gene families present across all the genome bins.

2.18.2 Metabolic pathway construction

The Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017) and Metacyc (Caspi et al., 2014) were used as a reference databases to construct metabolic pathways. Key genes involved in metabolic pathways were compared with their experimentally validated homologs using the BLOSUM62 (BLOcks SUbstitution Matrix) sequence alignment (Henikoff and Henikoff, 1992) option within Geneious (Kearse et al., 2012). CAZyme (carbohydrate active enzymes) and transporters were searches for in *Quinella* genome bins using dbCAN (Yin et al., 2012; Lombard et al., 2014) and TransportDB 2.0 (Elbourne et al., 2017) databases for reference.

Chapter 3 Rumen bacterial taxa associated with high and low methane emitting sheep

3.1 Introduction

There has been an increasing effort over the last decade to develop strategies to reduce methane (CH_4) emissions from farmed ruminants (Buddle et al., 2011; Moumen et al., 2016; de Haas et al., 2017). The possibility of breeding naturally low CH_4 emitting animals is one strategy that has gained much attention, especially for sheep (Pinares-Patiño et al., 2011a; Pinares-Patiño et al., 2013a; Goopy et al., 2016; Jonker et al., 2017b; Jonker et al., 2017a) . Cattle and sheep both show variations in the amount of CH_4 emitted from a specified amount of feed (Pinares-Patiño et al., 2013a; Hayes et al., 2016; Lassen et al., 2016). Experiments conducted by Pinares-Patiño and colleagues to measure the repeatability and heritability in methane emissions from sheep showed that low CH_4-ranked sheep had a lower CH_4 yield (g CH_4 per kg dry matter intake [DMI]) on both grass and pelleted diets (Pinares-Patiño et al., 2011a), and when a large scale experiment was performed using 1225 sheep, it was found that the CH_4 emission trait is heritable and repeatable (Pinares-Patiño et al., 2013b). Also, recent experiments conducted by Jonker et al. (2017a) showed that progeny of low CH_4 selection line sheep also produce less CH_4 even after shifting from a pelletized lucerne diet to grazing pasture. The amounts and types of volatile fatty acids (VFAs) formed during feed fermentation in the rumen are some of the major factors that affect the amount of methane formed (Rooke et al., 2014), and greater acetate:propionate ratios indicate more methane formation. Microbes present in the rumen, especially bacteria, play a major role in feed fermentation and regulate the amounts of the different VFAs formed in the rumen. During rumen fermentation, formation of different VFAs leads to the formation of different amounts of hydrogen (H_2) which is further converted to methane by methanogenic archaea (Janssen, 2010). So, it is logical to envisage that rumen bacterial community composition influences fermentation products and hence CH_4 emissions. It is hypothesised that bacteria present in low CH_4 emitting animals ferment the feed so that less H_2 is formed, which in turn leads to less CH_4 being formed than in high CH_4 emitting animals (Kittelmann et al., 2014; Shi et al., 2014; Kamke et al., 2016). It has been reported by Kittelmann et al. (2014) that there are (at least) two different rumen bacterial community types (Q and S) associated with low CH_4 emitting sheep, and (at least) one type (H) with high CH_4 emitting sheep. The relative abundance of *Quinella ovalis* was reported to be greater in the Q-community type, while

Sharpea azabuensis together with various other taxa, such as *Fibrobacter* spp., *Kandleria vitulina*, *Olsenella* spp., and *Prevotella bryantii*, were found to be relatively more abundant in the S-type community. The aim of the work described in this chapter was to perform a deeper examination of these two low CH_4-associated community types, and of their diagnostic species, *Quinella ovalis* and *Sharpea azabuensis*.

3.2 Results and Discussion

3.2.1 Rumen bacterial community structure in high and low CH_4 emitting sheep

The first step was to identify bacterial taxa associated with high and low CH_4 emitting sheep rumen samples. The data used in this study were previously generated from rumen samples collected from 118 sheep identified as high and low CH_4 emitters from a total of 340 individuals (Kittelmann et al., 2014). The 340 animals were studied in four cohorts because the CH_4 measurement facility could not accommodate all the animals at once. Methane emissions were measured and rumen samples were collected twice from each animal, in two measuring rounds about 3 weeks apart. CH_4 yields (g CH_4 per kg DMI) were calculated and 118 individuals were categorised as high or low CH_4 emitters based on the average CH_4 yield in the two measuring rounds (Pinares-Patiño et al., 2011a; Kittelmann et al., 2014). These were chosen for further study, giving a total of 236 rumen samples each with an associated CH_4 yield (Kittelmann et al., 2014). Figure 3.1 shows the CH_4 yields from the two measuring rounds from the 118 high and low CH_4 emitting sheep selected for rumen microbial community analysis. A "high CH_4 sample" refers to a sample collected from an individual categorised as a high CH_4 yield sheep and a "low CH_4 sample" refers to a sample collected from an individual categorised as a low CH_4 yield sheep. Methane yields from individual sheep can be variable between measuring rounds, with some samples from some high CH_4 emitters falling among the samples from low CH_4 emitters and vice versa (Figure 3.1).

In the study presented here, the dataset of partial bacterial 16S rRNA gene sequences generated from these 236 rumen samples using 454 Titanium pyrosequencing by Kittelmann et al. (2014) was reanalysed using an updated 16S rRNA gene based bacterial taxonomic framework developed by Henderson et al. (2017). This revised framework provides deeper taxonomic assignment (down to the genus level) for many groups of rumen bacteria previously only assigned to undefined orders or families. That refinement

was achieved by classifying bacterial 16S rRNA gene sequences from 684 rumen samples from a global study into known genera or newly defined provisional genus-level groupings (Henderson et al., 2017).

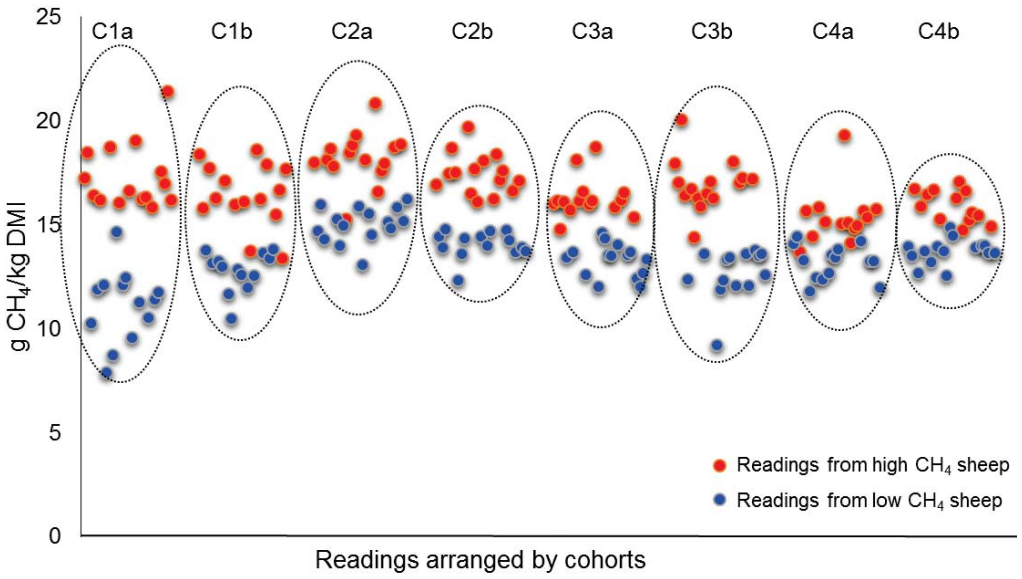


Figure 3.1 Methane yields associated with each of 236 rumen samples by cohort and measuring round (enclosed in dashed ovals). Each point represents one rumen sample with its associated CH_4 yield, and is coloured to indicate whether it came from a high (red) or low (blue) CH_4 emitting sheep. There were four cohorts of sheep (C1, C2, C3, and C4), and each was measured for CH_4 yield and had a rumen sample collected in each of the two measuring rounds (a and b).

The sequencing data consisted of 396,196 partial bacterial 16S rRNA sequence reads, with between 201 and 4073 sequences per sample.

The QIIME (Quantitative Insights Into Microbial Ecology) software package (Caporaso et al., 2010) was used for data analysis (for details see section 2.8). Sequencing data were denoised using Acacia (Bragg et al., 2012), which reduced the number of reads to 353,976 (174 to 3565 reads per sample). Figure 3.2A shows a rarefaction plot of OTUs in all the samples that contained greater than 700 reads (samples were grouped by high and low CH_4 yield). It seemed that the diversity of OTUs in samples from high CH_4-yield sheep was slightly greater than for low CH_4-yield sheep. Rarefaction plots were also drawn using the four samples with the greatest and the four samples with the smallest number of

reads (Figure 3.2B) to see whether there are any clear difference in the number of reads among samples. The general trend in the curves was similar regardless of the number of samples. The rarefaction curves showed that the number of sequence reads was not sufficient to cover all expected OTUs in the samples. However, it was expected that there would be relatively few genus-level taxa with many sequence reads, and that a small number of reads per sample would survey these dominant taxa. Therefore, to allow sufficient data for meaningful analyses, but at the same time avoid greatly reducing the number of samples, eight samples with reads less than 700 were excluded from analysis, reducing the number of usable samples to 228.

In total, 12150 operational taxonomic unit (OTUs), grouped at 97% sequence similarity, were found in the dataset. These were assigned to 269 bacterial taxa at the genus level, or as close as possible, and these genera belonged to 64 orders. Of these 64 orders, *Bacteroidales* and *Clostridiales* 1 constituted 39.7% and 32.4% of all sequences, respectively (Figure 3.3). Another 24.0% of sequences belonged to *Clostridiales* 2 (11.7%), *Erysipelotrichales* (4.7%), *Fibrobacterales* (4.1%), *Anaeroplasmatales* (2.3%), and *Rhodospirillales* (1.3%). The remaining 57 bacterial orders combined represented only 3.9% of the observed bacterial community (Figure 3.3). Overall taxonomic assignment at the order level was similar to that reported by Kittelmann et al. (2014). At the genus level, the resolution appeared to be finer, with 237 bacteria taxa found in the new analysis compared to 138 found by Kittelmann et al. (2014). However, the main findings remain the same, with the same major bacterial taxa found in these samples. These bacterial groups are very similar to those found in rumens from a wide range of different ruminant species (Henderson et al., 2015; Henderson et al., 2017).

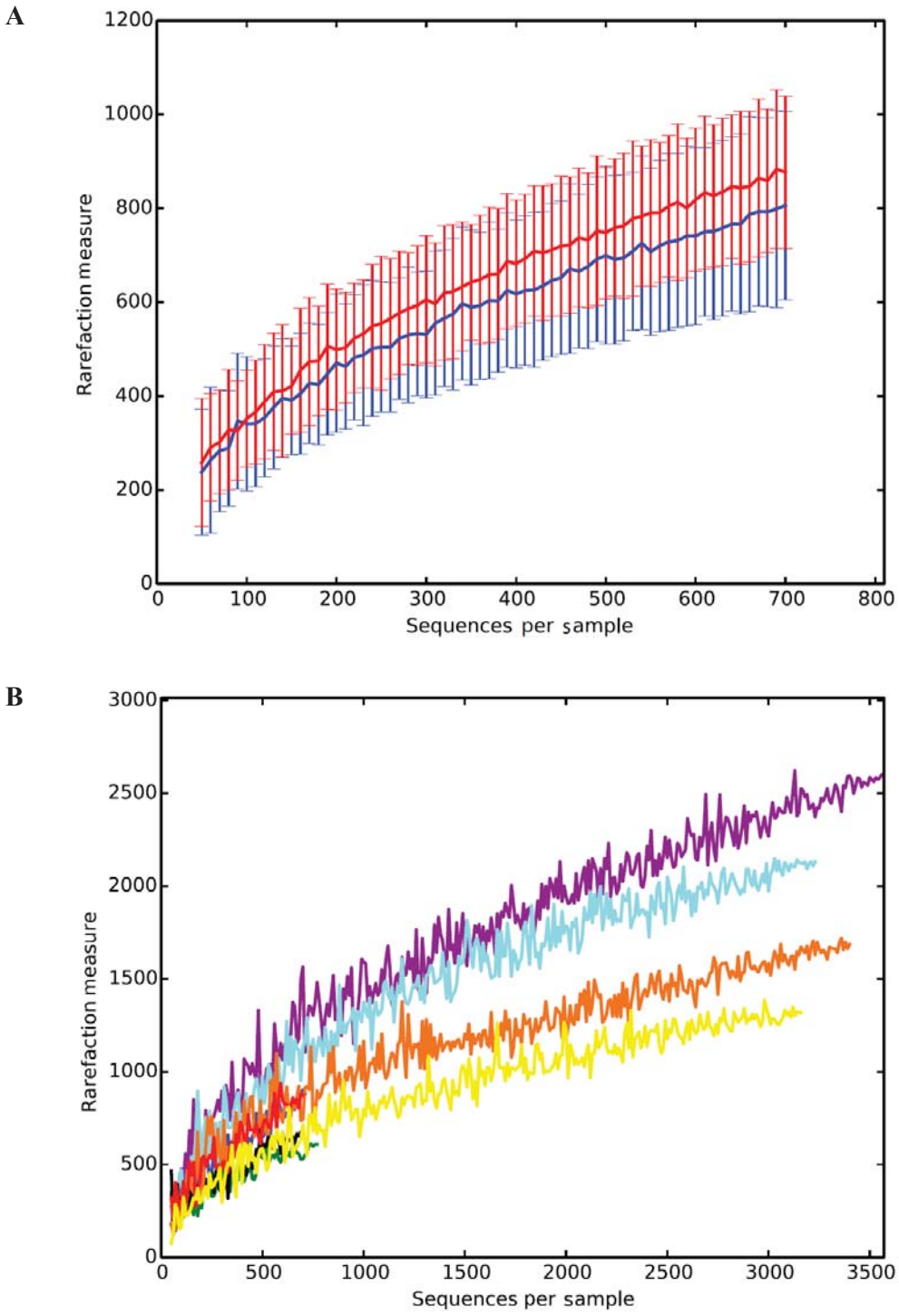


Figure 3.2 Rarefaction plots using chao1 matrix. Samples with less than 700 reads were excluded in this plot. A) Samples grouped by high (red line) and low (blue) CH_4 yield. B) Rarefaction plots for four samples (purple, light blue, orange and yellow coloured lines) with greatest number of reads and four samples (red, blue, black and green coloured lines) with the smallest number of reads.

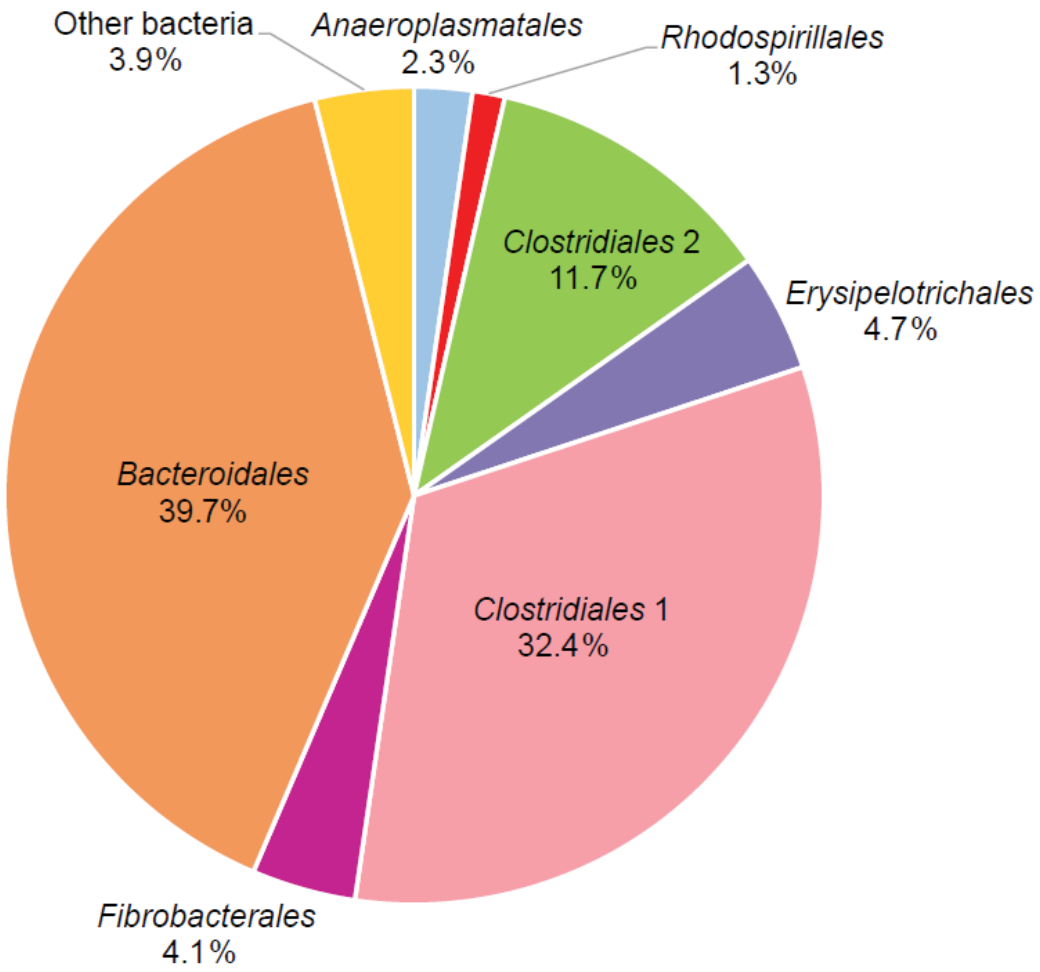


Figure 3.3 Contribution of various bacterial orders to overall bacterial community composition. Orders that contributed on average <1% are summarised as "Other bacteria".

3.2.2 Sample distribution by CH_4 and bacterial community composition

To determine if there was any association between the classification as a low or high CH_4 sample and the bacterial groups present in that rumen sample, a principal coordinate analysis (PCoA) was performed using the χ^2-distance metric (Figure 3.4). The 228 points in the graph, each representing one rumen sample, formed a broad tilted V. Both arms of the V-shape are dominated by low CH_4-associated samples, whereas the majority of high CH_4 samples lie at the junction of both arms. The two groups of samples associated with low CH_4 emitting sheep are separated, suggesting that two community types were associated with low CH_4 emitting sheep. There appears to be a continuum from each of the low CH_4-associated sample groups through to the high CH_4-associated samples. Some noise occurs throughout the graph, with some samples from high CH_4 sheep grouped at

the ends of the V and some samples from low CH_4 sheep grouped nearer the apex, and this is expected from the observed variation in CH_4 yields shown in Figure 3.1. However, overall the PCoA graph suggested that there were three broadly different types of bacterial communities: two associated with low and one associated with high CH_4 emissions. This agreed with the general pattern found by Kittelmann et al. (2014), using the same data but with a less well defined bacterial taxonomy.

The relative abundances of the different taxa were plotted to identify bacteria associated with the different low and high CH_4 communities (Figure 3.5). The samples were ranked by their position in the principal coordinate analysis following the direction of the arrow in Figure 3.4. The graph shows that the bacterial communities associated with the samples towards the extremes of the V in Figure 3.4 are different in composition to those associated with those at the apex. Samples from one arm of the V were characterised by a greater abundance of *Quinella* spp. and samples in the other arm by the presence of more sequences from *Sharpea* spp. (Figure 3.5). In contrast, there were no immediately obvious diagnostic taxa in samples that grouped near the apex of the V-shape (this is investigated in more detail below). The samples were generally categorised (Table 3.1) into three different bacterial community types: Q (*Quinella*-containing, n = 74), S (*Sharpea*-containing, n = 45) and H (high CH_4 mixed type, n = 109), based on the bacterial community dominance patterns (Figure 3.5) and CH_4 yield classification of the sheep (Figure 3.1). This agrees with and uses the community type designations introduced by Kittelmann et al. (2014). The association of samples from sheep classified as high and low CH_4 emitting with the 3 community types was not random (Table 3.1; $p < 0.0001$; χ^2-test).

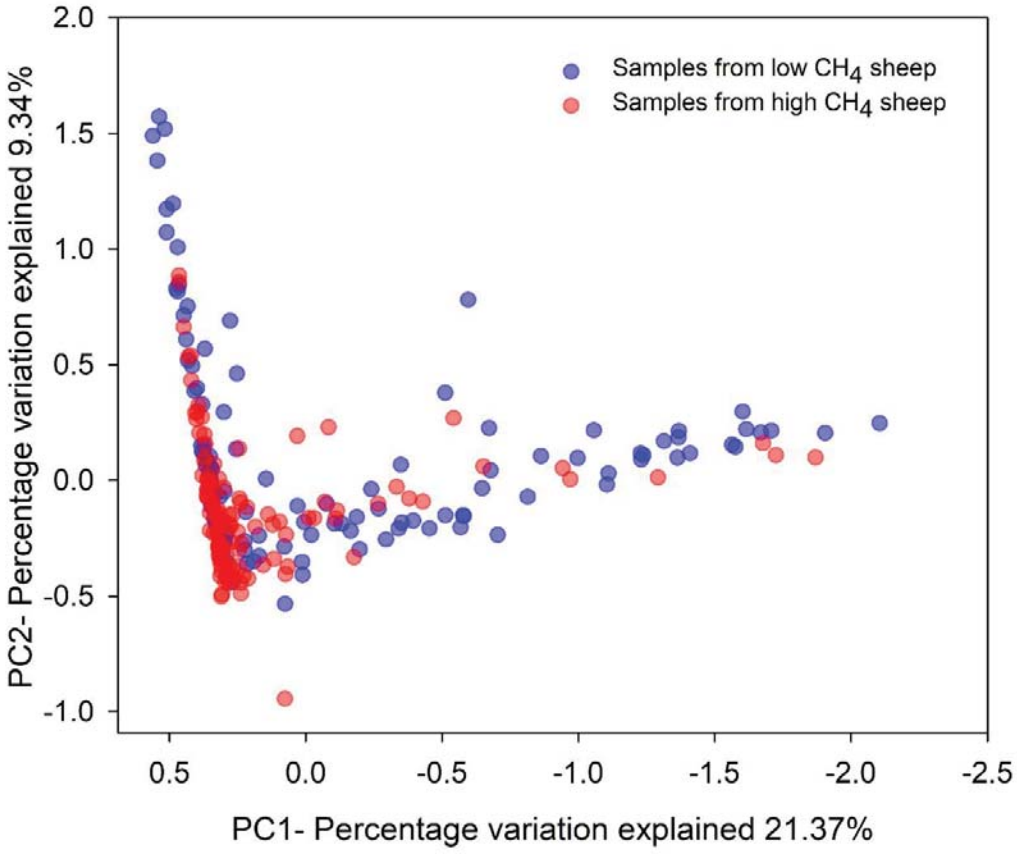


Figure 3.4 Principal coordinate analysis based on the χ^2-distance metric comparing the relative abundances of bacterial groups in each of 228 samples. Samples are coloured by CH_4 classification of the sheep: high CH_4 yield sheep (red), low CH_4 yield sheep (blue). The arrow shows the order in which the samples are arranged in Figure 3.4. The sum of the variation explained by the first two principal coordinates is the overall percentage variation explained by the two-dimensional graph (Ramette, 2007).

3.2.3 Bacterial taxa associated with the different community types

While *Quinella* and *Sharpea* appeared to be associated with the two low CH_4 bacterial community types (Figure 3.5), there may be other taxa that also differentiate the community types. A more detailed analysis was performed to identify these. To do that, a simplified dataset was generated by eliminating low-abundance taxa (section 3.2.3.1), and then a second dataset was generated from that by limiting it to samples associated with extremely high or low CH_4 yields (section 3.2.3.2). Using these, other taxa associated with the three different community types were identified (section 3.2.3.3).

Table 3.1 Distribution of community types in samples from high and low CH_4 sheep.

CH_4 classification	Community types			Total samples
	Q	H	S	
Samples from high CH_4 sheep	29.7% (22)[a]	75.0% (81)	31.1% (14)	117
Samples from low CH_4 sheep	70.3% (52)	25.0% (28)	68.9% (31)	111
Total samples	74	109	45	228

[a]Numbers in parentheses show the number of rumen samples in each classification.

3.2.3.1 Elimination of low-abundance taxa

To simplify the dataset, low-abundance taxa were eliminated, based on the understanding that carbon and energy flow in the rumen is the result of the major taxa present (Henderson et al., 2015). Taxa with relative abundances of less than certain cut-off (threshold) values were removed from the dataset. As the threshold value increased, the number of taxa, and sequence read coverage in the resultant simplified dataset decreased (Figure 3.6). However, by removing taxa that make very small contributions to the total, the number of taxa decreased rapidly while the number of sequences assigned to the remaining taxa remained large. A 1% cut-off was selected to produce a reduced dataset because it's application retained more than 96% of the observed bacterial community in terms of sequence reads (Figure 3.6), but reduced the number of taxa to 88 (i.e., under 37% of the total number of taxa). Again, PCoA analysis was used to find differences in the sample arrangement pattern when this 1% cut-off was used. By comparing this graph (Figure 3.7B) with the original PCoA graph (Figure 3.7A, which is the same as Figure 3.4), no obvious differences were seen in the sample distribution pattern, and the amount of variation described by the first two principal coordinates was similar (30.71% with all 269 taxa; 35.10% with the reduced dataset of 88 taxa). This indicated that the taxa eliminated with a 1% cut-off did not strongly influence the community composition of the low and high CH_4 samples. So, for further analysis, only the 88 taxa (Appendix A3.1) remaining after the 1% cut-off was applied were considered. This dataset is called the all samples (reduced taxa) dataset.

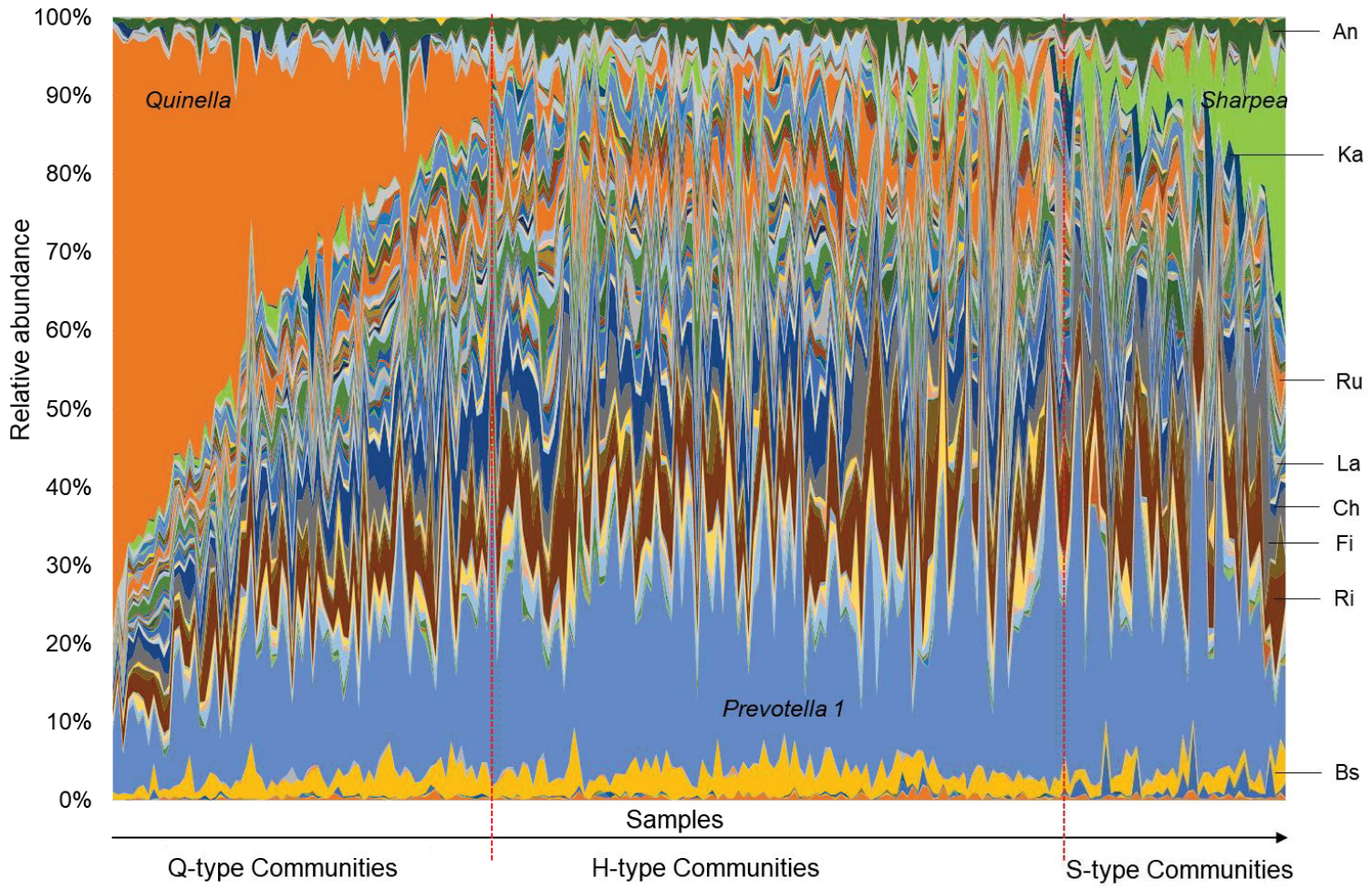


Figure 3.5 Relative abundance of bacterial taxa in 228 rumen samples. Samples are arranged in the order following the arrow shown in the principal coordinate graph (Figure 3.3). Taxa details: An = *Anaeroplasma*; Ka = *Kandleria*; Ru = *Ruminococcaceae* NK4A214 group; La = *Lachnospiraceae* NK3A20 group; Ch = *Christensenellaceae* R7 group; Fi = *Fibrobacter*; Ri = *Rikenellaceae* RC9 gut group Bs = *Bacteroidales* BS11 gut group.

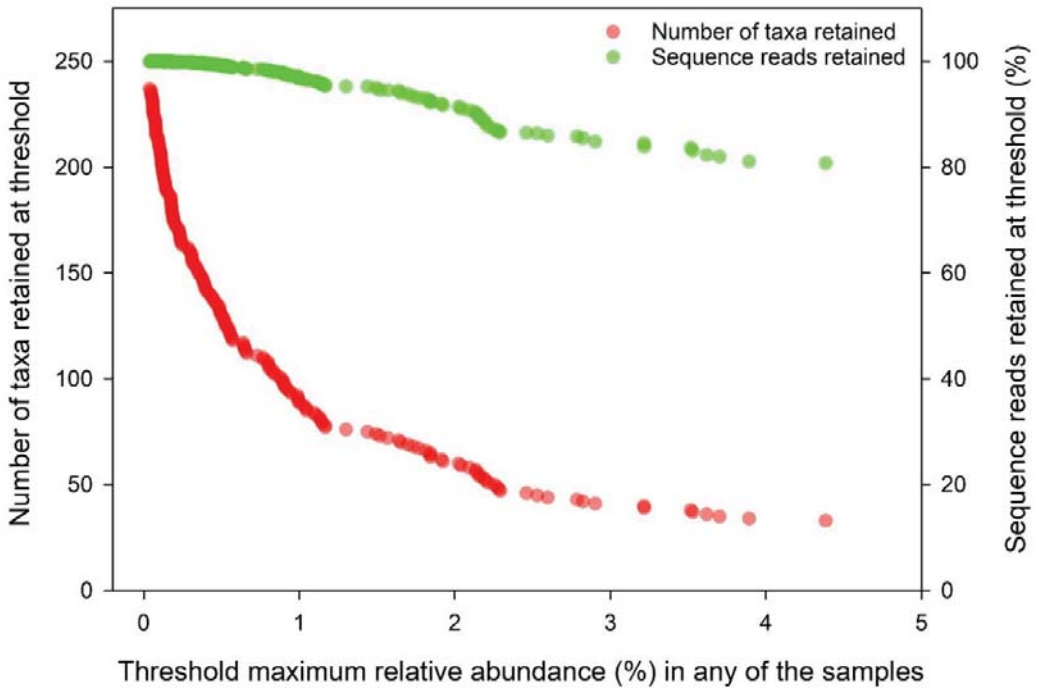


Figure 3.6 Number of taxa and sequence reads retained at different thresholds. The threshold is the maximum percentage relative abundance of any taxon in any one sample in the dataset. At each threshold, all taxa are discarded that are not present at threshold abundance in at least one sample.

3.2.3.2 Generating a dataset with only extreme CH_4 yield samples

The CH_4 yield associated with each of the samples varied from approximately 8 g CH_4 /kg DMI to 22 g CH_4 /kg DMI (Figure 3.8). A second dataset was generated from the all samples (reduced taxa) dataset that included only those samples that had associated CH_4 yields above 16 g CH_4 /kg DMI or below 14 g CH_4 /kg DMI (Figure 3.8B). This second dataset is referred to as the extreme samples dataset.

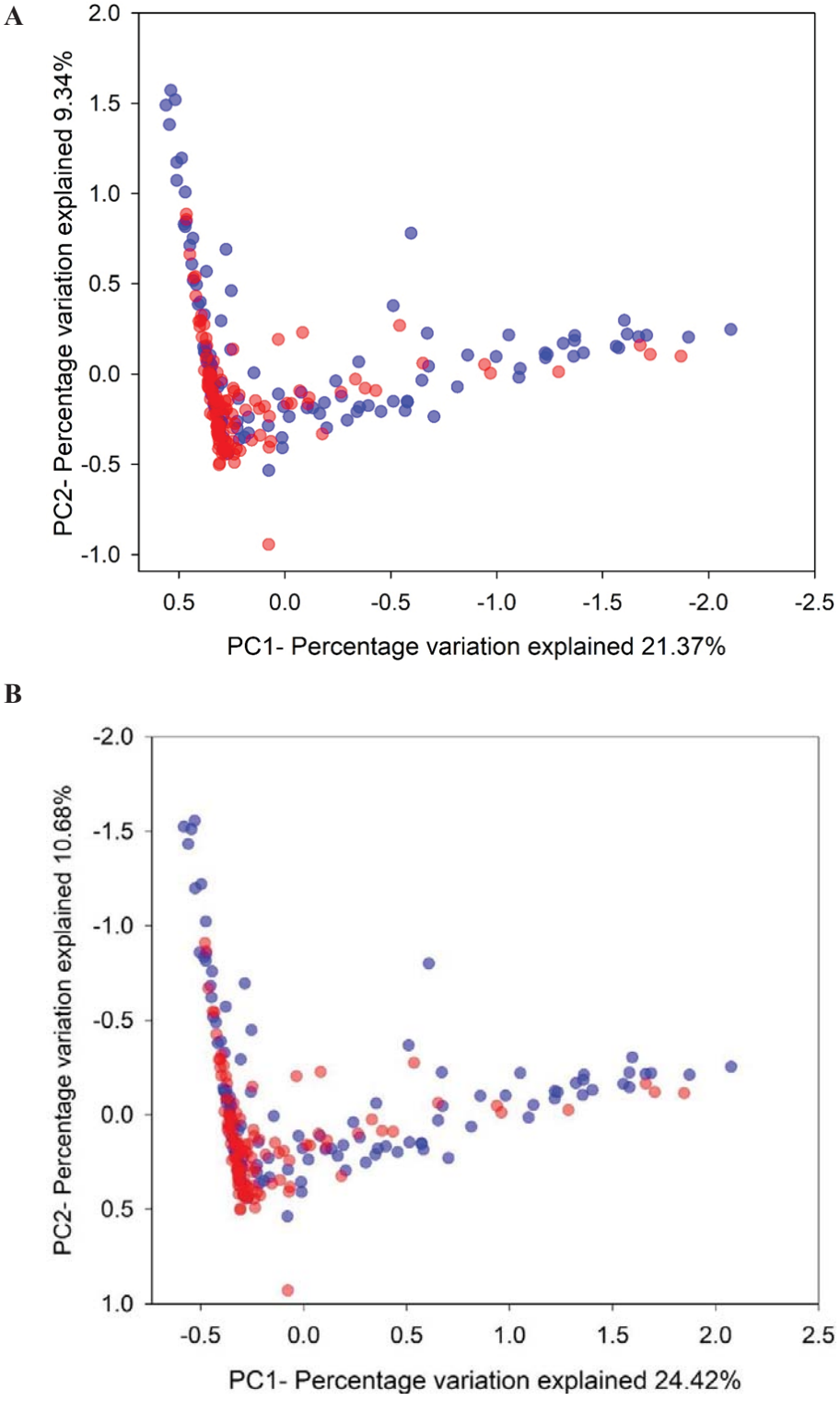


Figure 3.7 Principal coordinate analysis based on χ^2-distance metric comparing the relative abundances of bacterial groups in each of 228 samples A) including all taxa, and B) including only the 86 taxa remaining after a 1% cut-off was applied. Symbols: samples from high CH_4 yield sheep (red), samples from low CH_4 yield sheep (blue).

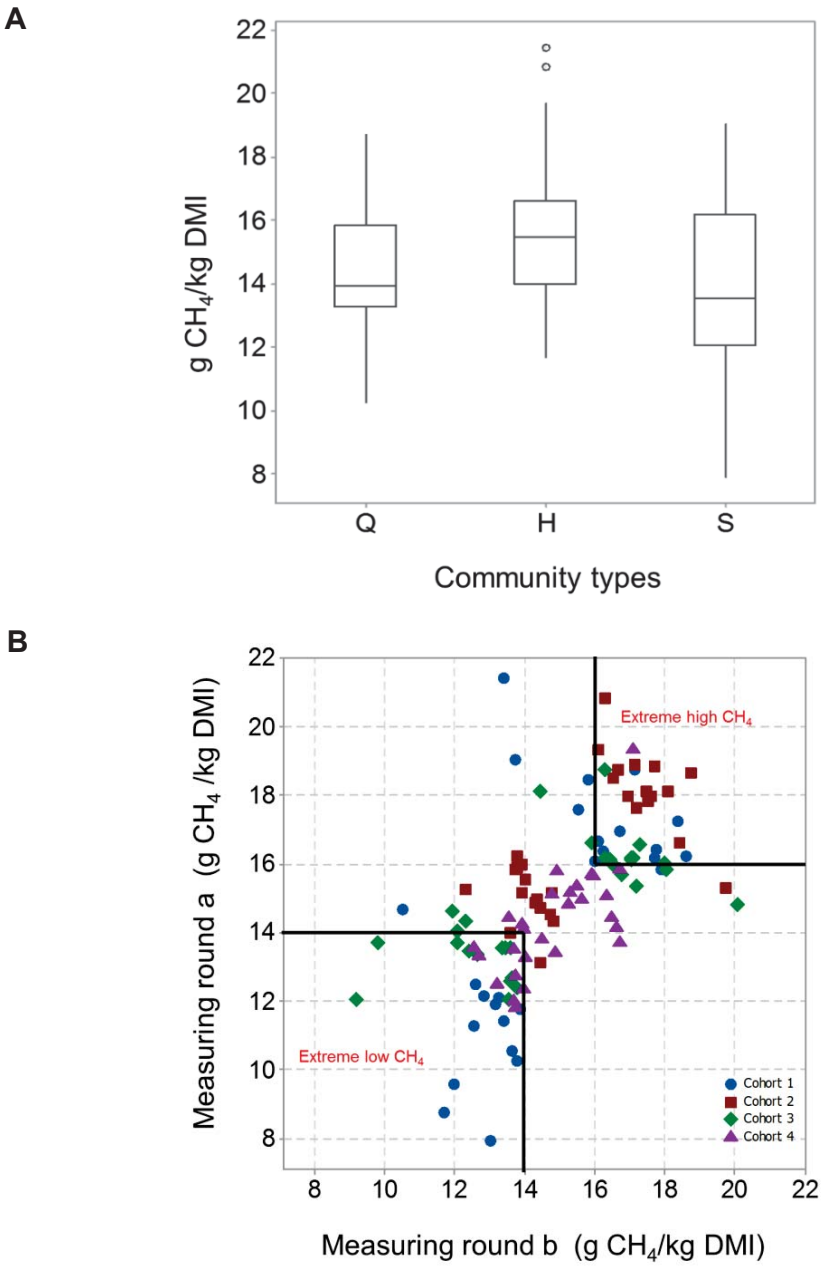


Figure 3.8 Sample distribution by CH_4 yield from sheep. A) Box plots show the variation in CH_4 yield associated with samples classified into the three different community types (Q, S and H). The horizontal lines in the boxes represent the means, and the upper and lower limits of the boxes represent one quartile (2nd and 3rd) either side of the mean. The tips of the vertical lines show the 1st and 4th quartiles, excluding two extreme samples that are shown as individual points. B) Sample distribution by cohort. Each point represents one sheep. Samples associated with CH_4 yields above 16 g CH_4/kg DMI and below 14 g CH_4 /kg DMI were selected as extreme high and low samples, respectively, and were used as the extreme samples dataset.

3.2.3.3 Taxa associated with Q, H and S-type communities

The two datasets (1, all samples [reduced taxa] and 2, extreme samples) were used to determine whether *Quinella* and *Sharpea* were significantly associated with low CH_4 yields, and to identify other potential high- and low-CH_4-associated taxa. The significance of the correlation of taxon abundance with CH_4 yield for each taxon was tested by Spearman's rank correlations, once using only the samples classified as Q and H-type, then in a second analysis using only samples classified as S and H community types, each time using the all sample (reduced taxa) dataset and the extreme samples dataset. The results for the 20 taxa with largest average relative abundance in each of the three community types, giving a total of 26 unique taxa, are shown in Table 3.2. *Quinella* was the most abundant taxon in the Q-type community, based on average abundances in samples assigned to that community type, and was significantly associated with the Q-type community in all tests (Table 3.2). In contrast, *Sharpea* and *Fibrobacter* were abundant taxa that were significantly more abundant in the S-type community. Other taxa were also differentially more abundant in the Q- and S-type communities and associated with low CH_4 yields, although less abundant based on 16S rRNA gene sequence reads. The most abundant of these were *Kandleria* and members of the *Coriobacteriaceae* UCG 001. Members of *Prevotella* 1, *Ruminococcaceae* UCG 005 and *Ruminococcaceae* NK4A214 group were significantly associated with the high CH_4 H-type community, but except *Prevotella* 1overall their relative abundances were not as high as the taxa that were associated with the two low CH_4 community types. Of the 26 taxa, only *Anaeroplasma* was not significantly different in abundance between community types, in any of the analyses using either of the datasets.

In addition, the Kruskal-Wallis test was used to identify taxa that were significantly associated with Q-, H- and S-type communities using the all samples (reduced taxa) and the extreme samples dataset (Figure 3.9). This analysis also showed that *Quinella* was significantly associated with the Q-type community, along with an uncultured genus of *Synergistaceae*. Multiple genera were associated with the S-type community, and most of these were also among the 26 most abundant taxa (Table 3.2). These included *Sharpea* and *Kandleria*.

Table 3.2 Average relative abundances (%) of taxa significantly associated with H- (high-CH_4) and Q- or S- (low-CH_4) type communities. The number of samples classified in each community type is indicated (n). Each statistical analysis was performed twice, once using all samples with the reduced taxa dataset, and once with only the extreme samples (Figure 3.8B). Additionally, the difference in relative abundance of each taxon was tested using the Kruskal-Wallis test and the significance of taxa differences are also listed (see Figure 3.9 for more details). The 20 taxa with greatest average relative abundance in any of the community type in either of the dataset are listed in the table, resulting in 26 unique taxa.

	Taxa	Average relative abundance in Q-type samples (n = 74)	Significance of correlation of abundance with CH_4 yield in Q and H-type samples	Average relative abundance in H-type samples (n = 109)	Significance of correlation of abundance with CH_4 yield in S and H-type samples	Average relative abundance in S-type samples (n = 45)	Significance of difference in relative abundance by Kruskal-Wallis test
Low-CH_4 associated taxa	*Kandleria*	0.29	b	0.18	AB	2.28	CD
	Quinella	31.90	AB	1.30	–	0.58	CD
	Fibrobacter	2.86	–	3.52	AB	7.53	CD
	Coriobacteriaceae UCG 001	0.07	–-	0.09	AB	1.18	CD
	Sharpea	0.77	b	1.35	b	9.17	CD
	Anaeroplasma	2.02	–	2.22	–	2.82	–
	Lachnospiraceae NK4A136 group	0.82	–	1.04	–	0.70	c
	Ruminococcus 1	1.81	–	2.34	–	1.88	D
	Saccharofermentans	0.77	–	0.97	–	0.77	c
	Syntrophococcus	0.35	–	0.39	–	0.76	CD

	Taxa	Average relative abundance in Q-type samples (n = 74)	Significance of correlation of abundance with CH_4 yield in Q and H-type samples	Average relative abundance in H-type samples (n = 109)	Significance of correlation of abundance with CH_4 yield in S and H-type samples	Average relative abundance in S-type samples (n = 45)	Significance of difference in relative abundance by Kruskal-Wallis test
High-CH_4 associated taxa	*Prevotellaceae* UCG 001	1.52	–	2.82	–	1.58	CD
	Dorea	0.27	–	0.72	–	1.10	Cd
	Bacteroidales S24_7	1.34	–	2.16	–	2.21	CD
	Lachnospiraceae NK3A20 group	1.80	AB	2.63	–	2.39	Cd
	Rikenellaceae RC9 gut group	5.29	b	8.63	–	8.45	CD
	Bacteroidales BS11 gut group	2.44	–	3.00	b	2.10	Cd
	Acetitomaculum	1.57	ab	2.33	–	2.47	cd
	Lachnospiraceae gauvreauii group	0.82	b	1.02	–	1.60	CD
	Thalassospira	0.86	–	1.67	b	0.49	CD
	Xylanibacter	1.21	Ab	2.20	ab	1.62	cd
	Prevotella 1	14.97	Ab	22.79	–	25.32	CD
	Lachnospiraceae AC2044 group	0.77	–	1.21	aB	0.44	CD
	Christensellaceae R7 group	4.38	–	5.49	aB	2.43	CD
	Cyanobacteria 4C0d 2	0.58	AB	0.87	AB	0.36	CD
	Ruminococcaceae NK4A214 group	1.99	AB	3.34	AB	1.92	CD
	Ruminococcaceae UCG 005	0.87	AB	1.50	AB	0.54	CD

[a]Significance of correlations in tests are indicated by capital letters where $0.05 \geq p \leq 0.001$ and lower case letters (where $p < 0.001$) : A and a, Spearman's rank across all samples; B and b, Spearman's rank across extreme samples; C and c, Kruskal-Wallis test across all samples; D and d, Kruskal-Wallis test across extreme samples; –, not significantly associated. Q = *Quinella*-rich community type, S = *Sharpea*-rich community type, H = high-CH_4 community type.

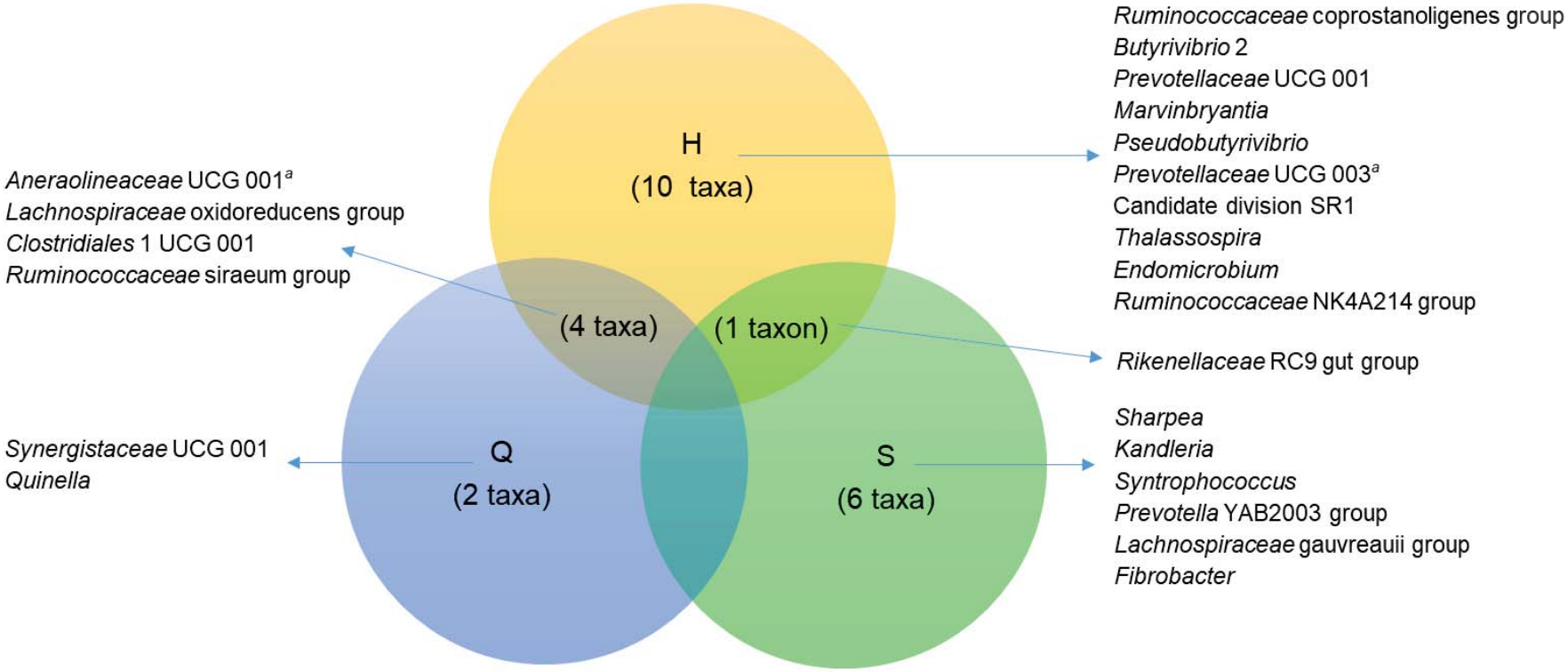


Figure 3.9 Taxa associated with different community types. The Kruskal-Wallis test was performed on the all samples (reduced taxa) dataset and the extreme samples dataset to find taxa significantly associated with Q-, H- and S-type communities. This analysis included all 86 taxa, but only those that were significantly associated with unique sectors on the Venn diagram are listed, except those in the central sector. All were significantly associated using both datasets, except those marked [a], which were significantly associated in only the all samples (reduced taxa) dataset).

3.2.4 Operational taxonomic units associated with different community types

OTU level analysis was performed to find OTUs that discriminated the two low CH_4 community types from the H-community type and from each other. Samples were normalized to a total of 780 sequence reads so that all OTU were given equal weight. Only those OTUs with more than 10 reads in the total dataset were considered for further analysis. Sparse Partial Least Squares – Discriminant Analysis (sPLS-DA) was used to identify discriminating OTUs. Two OTU sets were used for this analysis. The first dataset contained all OTUs that contained more than 10 sequencing reads and the second dataset was a subset of the first and contained only the 200 most-discriminating OTUs identified using VIP (Variable Importance in the Projection) function from mixOmics of R) from each dimension. Merging these gave 255 unique OTUs in total. Analysis of both OTU sets generated the same V-shaped pattern of separation of the samples (Figure 3.10) as was found when the samples were plotted based on genus level-classification of the sequence data (Figure 3.5). sPLS-DA-based cross validation of samples, categorised as Q, H and S-community types, was also performed. In this analysis, a 19.7 % error was found using all OTUs with >10 reads, while the error was reduced to 10.5% in the analysis conducted used the 255 unique discriminating OTUs (Table 3. 3). To examine these 255 OTUs more clearly, the coefficient (loading) for each OTU in dimensions 1 and 2 were plotted (Figure 3.11). Based on the assignment of the repset sequences (a representative sequence from each OTU), the discriminating OTUs were assigned to 44 taxa. The identities of OTUs assigned to the seven taxa (covering 76.5% of the 255 OTUs) with the greatest number of the 255 OTUs are shown in Figure 3.11. OTUs that separated the rumen samples in the negative direction in dimension 2 (Figure 3.10), where the samples previously designated as Q-type samples were found (Figure 3.11B), nearly all belonged to *Quinella*. In contrast, OTUs that discriminated samples in the negative direction in dimension 1 (Figure 3.10), where the samples previously designated as S type were found (Figure 3.11A), were classified as belonging to *Sharpea* (7 OTUs), *Kandleria* (2 OTUs), *Prevotella* 1 (12 OTUs) and *Rikenellaceae* RC9 gut group (3 OTUs). However, *Prevotella* 1 (12 OTUs) and *Rikenellaceae* RC9 gut group (5 OTUs) were also associated with the positive direction in dimension 2, and so can attributed to the H-type community. This means that potentially different species of these groups were associated with the H- and S- type communities. OTUs associated with *Christensenellaeae* R7 group were associated with the H-type community, which is associated with high CH_4 yields.

Interestingly, *Christensellaceae* are positively associated with methanogens in the human gut (Goodrich et al. 2014).

Based on this analysis, members of the genera *Quinella* and *Sharpea* were confirmed to be associated with Q- and S-type low CH_4 bacterial communities, respectively. *Quinella* is regularly found in studies of rumen microbial communities (Deng et al., 2007; Gruninger et al., 2014; Li et al., 2016), especially in the sheep rumens (Belenguer et al., 2010; Kittelmann et al., 2014; Henderson et al., 2015), but its association with CH_4 emissions is unexplained. Kittelmann et al. (2014) first reported an association of *Quinella* with low CH_4 sheep and this is confirmed here by a more detailed analysis of the same dataset. Wallace et al. (2015) reported that the members of the family *Veillonellaceae* were significantly associated with low CH_4 emitting cattle, but this was not further resolved to the genus level. Since *Quinella* is sometimes included in the family *Veillonellaceae* (based on older taxonomic schemes), it is unclear whether *Quinella* is abundant in low CH_4 emitting cattle. In the Global Rumen Census study (Henderson et al., 2015), it was found that *Quinella* was mainly present in sheep (in 89 out of 113 sheep rumen samples, 78.8%) and deer (in 59 out of 61 deer rumen samples, 93.7%), while it occurs less frequently in cattle (in 170 out of 419 cattle rumen samples, 40.6%). *Quinella* can be very abundant, with four sheep rumen samples in the Global Rumen Census having relative abundances of 18.0 % to 48.7%. It was also found in high abundance in deer (19.4 % to 25.3% of all 16S rRNA genes), in one giraffe (19.7%) and in one cattle animal (16.1%). In the samples studied here, *Quinella* made up 6.8 to 74.4% of the 16S rRNA gene sequences in samples assigned to the Q-type community. However, the physiology this bacterium, and how it contributes to low CH_4 yields, has not been reported.

In contrast, *Sharpea* has been found to be associated with low CH_4 emitting cattle (Wallace et al., 2015; Denman, 2016). However, it is seems widely distributed in ruminants based on the Global Rumen Census study (Henderson et al., 2015) , where *Sharpea* was mainly present in cattle (91 out of 419 rumen samples, 21.7%), goats (9 out of 52 rumen samples, 17.3%), sheep (14 out of 113 rumen samples, 12.4%), and deer (6 out of 61 rumen samples, 9.8%). Similarly, *Kandleria* which is closely related to *Sharpea* (Salvetti et al., 2011), followed a similar trend, and was found in cattle (117 out of 419 rumen samples, 27.9%), sheep (30 out of 113 rumen samples, 26.5%), bison (6 out of 14, 42.8%), and buffalo (5 out of 26, 19.2%), but less often in deer (5 out of 61 rumen samples, 8.2%) and goats (2 out of 52 rumen samples, 3.8%).

An extensive metagenome and metatranscriptome study was conducted by Kamke et al. (2016) on a subset of these sheep. The low CH_4 yield sheep in this study were mainly those that harboured more *Sharpea*. The authors inferred that the higher relative abundance of lactate-producing *Sharpea* spp. was associated with greater production and utilisation of lactate, and greater production of propionate and butyrate, as indicated by a higher number of genes and transcripts (D-lactate dehydrogenase, acyl-CoA dehydrogenase, propionate CoA transferase and many of the steps in the conversion of pyruvate or acetyl-CoA to butyrate) in the low CH_4 yield sheep. On the basis of community structure analysis, gene and transcript abundance, they hypothesised that the lactate produced by *Sharpea* may be converted to butyrate by *Megasphaera* spp., and that in this way less methane is produced in comparison to fermentation to acetate and butyrate performed by members of *Lachnospiraceae* and *Ruminococcaceae*, which appeared to be characteristic of the high CH_4 yield sheep.

Table 3.3 Assessment of sPLS-DA via cross validation, using a "leave-one-out" validation accuracy method. Both datasets were tested independently. The table shows the number of samples classified to each community type by applying the cross validation methodology. The samples are listed by their original classification. The number of correctly classified samples are shown in bold font.

Original sample classification	Cross-validation classification based on all OTUs with >10 reads[a] Number of samples classified to			Cross-validation classification based on 255 discriminating OTUs reads[b] Number of samples classified to		
	Q-type	H-type	S-type	Q-type	H-type	S-type
Q-type samples	**60**	11	3	**65**	8	1
H-type samples	1	**90**	18	0	**103**	6
S-type samples	0	12	**33**	0	9	**36**

[a]Overall error (%) in OTUs containing > 10 reads = 19.7%

[b]Overall error (%) in 255 discriminating OTUs containing > 10 reads per OTUs = 10.5%

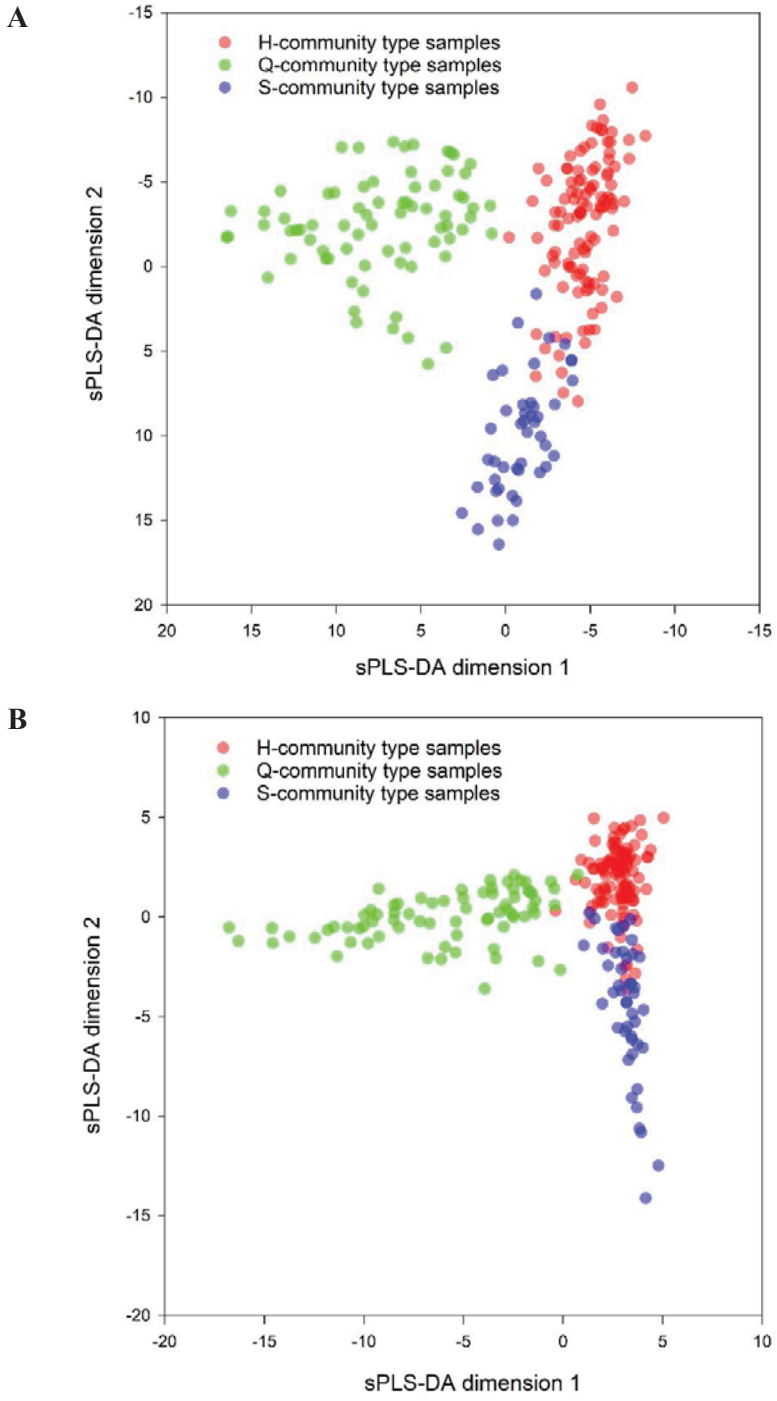


Figure 3.10 Separation of samples assigned to Q- and S- and H-communities types based on sPLS-DA scores. A) Analysis using the all OTUs with > 10 reads. B) Analysis using the 255 most discriminating OTUs (subset of OTUs with > 10 reads). The plot shows the distribution of the communities in the 228 rumen samples. The samples are marked according to the community type (Q, H and S) they were classified into (section 3.2.2).

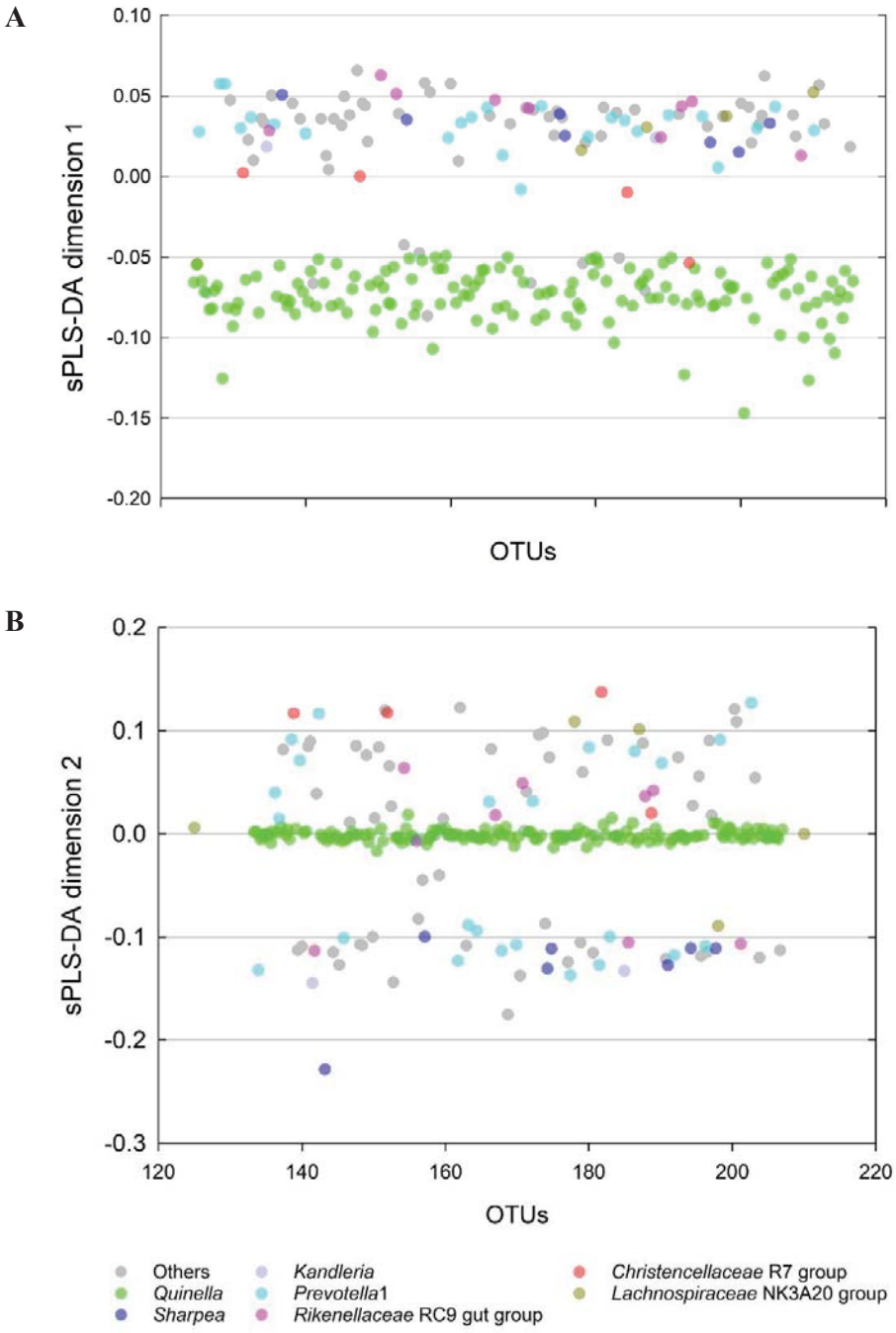


Figure 3.11 sPLS-DA loading associated with OTUs that separate samples in Figure 3.10B. A) Loadings of sPLS-DA dimension 1. B) Loadings of sPLS-DA dimension 2 of the same OTUs. The taxonomic affiliations of the OTUs are indicated by colour codes.

3.2.5 Volatile fatty acid (VFA) profiles of Q-, S- and H-type community samples

Kamke et al. (2016) have suggested how the bacteria of the S-type community might lead to formation of less CH_4 compared to the H-type community. Analysis of the VFAs in the rumen samples (data provided by C. Pinares-Patiño and A. Jonker, AgResearch) taken as part of the Kittelmann et al. (2014) study, classified by community type, suggested that samples containing Q-type communities had a significantly (p-value <0.001; one-way ANOVA) greater ratio of propionate to acetate than samples containing the other two community types (Figure 3.12). Therefore, it seems that bacteria present in the Q-type community may produce more propionate. Propionate formation is known to be an electron sink, resulting in less H_2 formation, and increased propionate formation is often found when CH_4 yields are low (Janssen, 2010). It would be interesting to investigate the fermentation products of abundant bacteria significantly associated with the Q and S community types. This will be the subject of later chapters in this thesis.

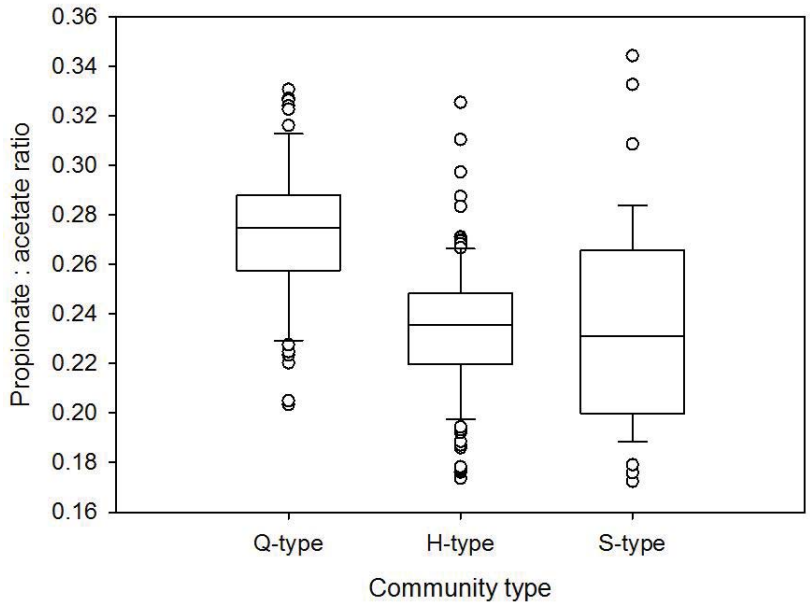


Figure 3.12 Propionate to acetate ratios in rumen samples that contained Q-, S- and H-type communities. The horizontal lines in the boxes represent the means, and the upper and lower limits of the boxes represent one quartile (2nd and 3rd) on either side of the mean. The tips of the vertical lines show the 1st and 4th quartiles, excluding extreme samples that are shown as individual circles.

Since they are abundant in the low CH_4 yield sheep rumens, associated with lower CH_4 yields, and generally less studied, *Quinella, Sharpea* and *Kandleria* (closely related to *Sharpea*) were selected for further investigation. First, a better understanding of their phylogenetic diversity was gained (next section in this chapter), and then their end products were studied (chapters 4 and 6) to understand their contributions to lower CH_4 emission by sheep that have greater abundances of these bacteria in their rumens.

3.2.6 Taxonomic refinement of *Quinella*, *Sharpea* and *Kandleria*

Strains of the genus *Quinella* have not yet been cultured, so the taxon named *Quinella ovalis* was described under rule 18a of chapter 3 of the International Code of Nomenclature of Bacteria (Lapage et al., 1992). This rule allows the naming of a bacterial taxon without a culture being isolated and deposited in a culture collection (Lapage et al., 1992). Instead, a description, a preserved specimen, or an illustration can be used as the type. The description of *Quinella ovalis* is based on features observed in a number of different studies, and a partial 16S rRNA gene sequence was obtained from enriched cells matching that description (Krumholz et al., 1993). In contrast, *Sharpea* and *Kandleria* (previously known as *Lactobacillus vitulina*) have been successfully cultured (Sharpe et al., 1973; Morita et al., 2008; Salvetti et al., 2011), but they are not well studied. Overall, the species diversity of *Quinella, Sharpea* and *Kandleria* in the rumen is not well defined. The first step towards better understanding these bacterial groups was to obtain an overview of diversity and to refine their taxonomic status. For this, good quality full length 16S rRNA gene sequence data were needed.

To obtain almost full-length good quality 16S rRNA gene sequences of *Quinella*, *Sharpea* and *Kandleria*, rumen samples that were identified by the pyrosequencing analysis described in section 3.2.2 as having high relative abundances of each genus were selected for clone library construction. Five rumen samples from different sheep were selected for each of these bacterial groups (Table 3.4). Samples from different cohorts were selected to increase the likelihood of capturing a wider diversity of these bacteria. Clone libraries were constructed following the method described in chapter 2, section 2.7.6. Of 308 cloned sequences, 26 were assigned to *Quinella ovalis*, 6 to *Sharpea azabuensis* and 3 sequences to *Kandleria vitulina* based on an initial BLAST analysis using a refined database of bacterial 16S rRNA gene sequences (Henderson et al., 2017) and after checking for possible chimeras (Huber et al., 2004). An additional 11 cloned sequences

from the 308 were included in the analysis as potentially derived from *Quinella* spp., on the basis of BLAST results which suggested that they originated from bacteria closely related to the family *Selenomonadaceae* (which contains *Quinella*). In total, 49 potential *Quinella*, 38 *Sharpea* and 22 *Kandleria* sequences were available from the new clones and public databases for inclusion in the phylogenetic refinement. All sequences were >1443 nt long.

Table 3.4 Samples used for clone library construction, and sequences obtained.

Genus the sample was selected for	Samples	Relative abundance (%) of target genus in pyrosequencing analysis	Total number of clones sequenced	Number of clones identified as closely related to		
				Quinella	*Sharpea*	*Kandleria*
Quinella	S9110.C2b	68.8	10	4	–[a]	–
	S9643.C2b	74.4	10	3	–	–
	S154.C3b	51.7	15	5	–	–
	S386.C4b	27.7	16	1[b]	–	–
	S964.C3b	59.4	18	6	–	–
Sharpea	S1253.C1b	32.0	30	0	1	–
	S1494.C1a	36.5	7	0	3	–
	S172.C3b	9.8	16	0	–	–
	S17.C4b	15.5	31	0	–	–
	S490.C3b[c]	11.5/6.1[d]	33	0	1	1
Kandleria	S1120.C1b	13.7	33	6	–	–
	S1592.C1b	24.7	34	0	1	1
	S350.C4b	9.8	22	1	–	1
	S62.C3b	6.8	33	0	–	–

[a]–, no clones were identified belonging to that particular taxon.
[b]Close relative, later assigned to new genus (see Figure 3.13).
[c]Sample selected for both *Sharpea* and *Kandleria*.
[d]Abundance of *Sharpea* followed by abundance of *Kandleria*.

Phylogenetic trees of 16S rRNA genes from members of the genera *Quinella, Sharpea* and *Kandleria* were made using all the available long length sequences (>1300 nt) in the SILVA SSURef database version 111 (Quast et al., 2013), in addition to the new clone

library sequences generated (Table 3.4). The sequences assigned to *Quinella* displayed more sequence variation than did those assigned to either *Sharpea* or *Kandler*ia (Figure 3.13). Different authors use different sequences similarity cut-offs % to define species- and genus-level taxa, varying from 93% to 98% (Stackebrandt and Goebel, 1994; Everett et al., 1999; Kenters et al., 2011). Conservative sequence similarity cut-offs for species- and genus-level taxa, respectively, of 93% and 97% were used in this study. Six of the new cloned *Quinella* sequences clustered with the reference sequence of *Quinella ovalis* (GenBank accession M62701), with sequence similarities ≥ 97.3%, and a seventh with a sequence similarity of 96.9% (Figure 3.14). Based on a criterion that sequences with similarities of approx. 97% or more may originate from the same species, these seven sequences could be designated as *Q. ovalis*. The remaining sequences likely represent seven other candidate species if only clusters with two or more sequences are considered. Three of the seven candidate species in the refined *Quinella* tree contained sequences only from the newly-formed clone libraries from section 3.2.8, representing two individual sheep rumen samples each. Sequences in *Quinella* candidate species 3 and *Quinella* candidate species 5 were from one single study (Yang et al., 2010), but formed two different species-level clusters. In addition, there were multiple sequences that did not group with these eight clusters, and may represent other species if they are not amplification artefacts. Overall, this analysis suggests that there are multiple species of the genus *Quinella* species, and more intensive investigation and generation of high-quality sequences would help confirm this. Additionally, one potentially new genus-level cluster was found that was closely related (>92.4% similarity) to *Quinella*. There appeared to be two species-level clusters within this potentially new genus (Figure 3.13).

All sequences affiliated with *Sharpea* clustered closely together as one species, *S. azabuensis* (Figure 3.13), which is the only recognised species of the genus (Morita et al., 2008). All six of the newly cloned *Sharpea* sequences fell into this species. This cluster also contained sequences from eight different studies ((Ley et al., 2008; Morita et al., 2008; Creevey et al., 2014); and five unpublished studies with data deposited in GenBank which originated from six different sources – sheep rumen, cattle rumen, horse faeces, Goeldi's marmoset faeces, compost, and swine intestine). This diversity of origins suggests that *Sharpea* is not restricted to one environment. It was surprising that the sequences from all these different samples were represented by just one genus-level cluster (Figure 3.15).

Similarly, all *Kandleria* affiliated sequences clustered with only one species, *Kandleria vitulina* (Figure 3.13). The three newly-cloned *Kandleria* sequences grouped with the pre-existing *K. vitulina* cluster with >96.2% sequences similarity. This cluster contained sequences from six different studies ((Weisburg et al., 1989; Lloyd et al., 2006; Morita et al., 2008; Noel, 2013; Creevey et al., 2014); including one unpublished study with data deposited in GenBank) with genetic material originating from three different animal species (horse, cattle and sheep). This suggests that *K. vitulina* might not be host-species specific but potentially found in the gastrointestinal tract of grazing animals.

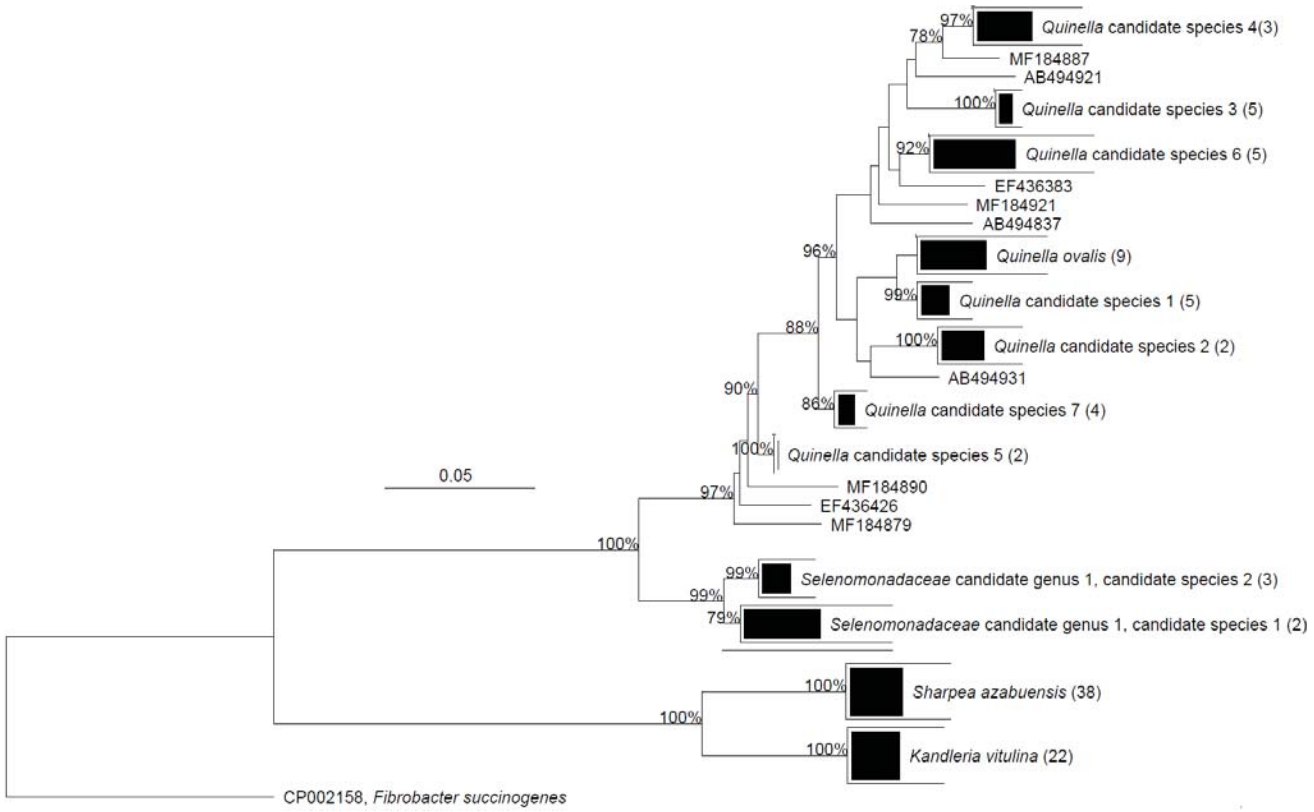


Figure 3.13 Refined phylogenetic tree of the genera *Quinella, Sharpea* and *Kandleria*. Numbers in brackets represent the total number of sequences in a particular cluster. The 16S rRNA gene sequence of *Fibrobacter succinogenes* (FibSuc43, GenBank accession CP002158) was used as an out-group sequence. The scale bar represents 0.01 changes per nucleotides residue. The numbers at the nodes are the percentage of trees that conserved that node in 1000 bootstrap resamplings.

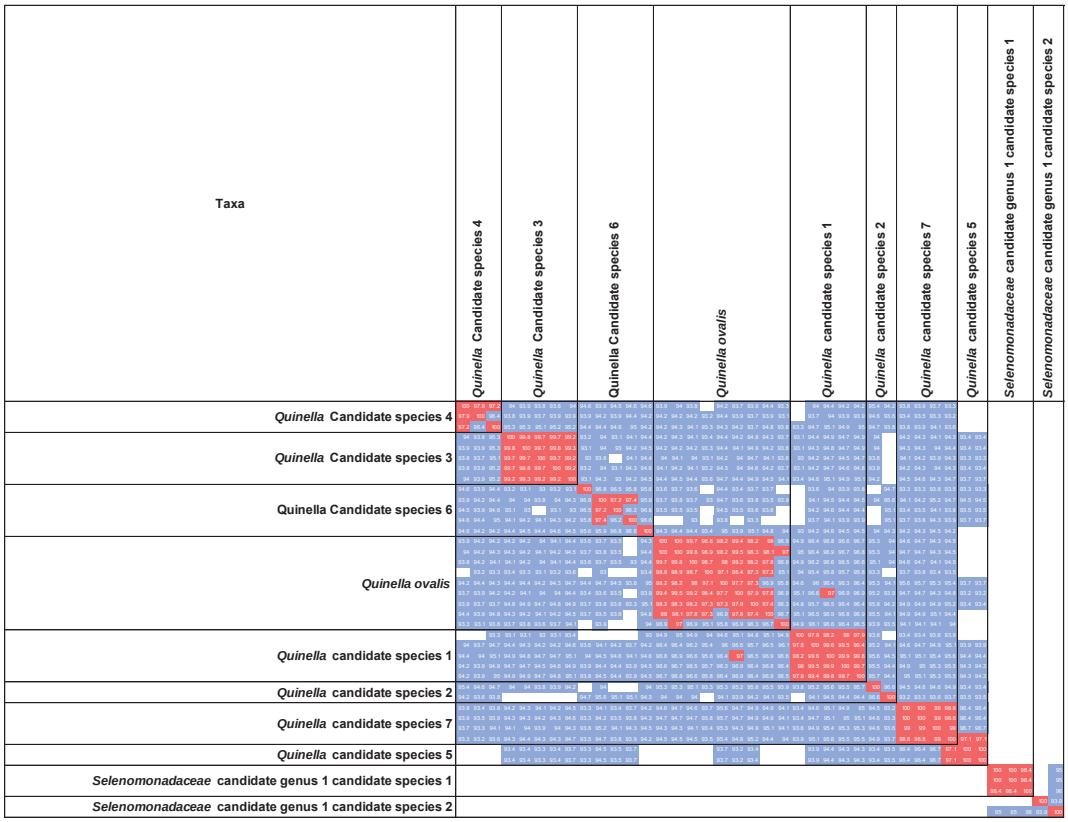


Figure 3.14 Sequence identities between sequences from *Quinella* and *Selenomonadaceae* genus 1 in the clusters in the refined phylogenetic tree of *Quinella*, *Sharpea* and *Kandleria* (Figure 3.10). The clusters are arranged in same order as they appear in the phylogenetic tree. Blue shaded cells show sequence identities of 93.0 to 96.9% (potential genus level clusters) whereas red shaded cells represented sequence identities of ≥ 97.0% (potential species level cluster, bordered with black lines).

Figure 3.15 Sequence identities between sequences from *Sharpea* and *Kandleria* in the clusters in the refined phylogenetic tree of *Quinella*, *Sharpea* and *Kandleria* (Figure 3.13). The clusters are arranged in same order as they appear in the phylogenetic tree. Blue shaded cells show sequence identities of 93.0% to 96.9 (potential genus level clusters) whereas red shaded cells represented sequence identities of ≥ 97.0% (potential species level cluster, bordered with black lines).

3.2.7 *Quinella*, *Sharpea* and *Kandleria* diversity within and between sheep

The refined phylogenetic tree of *Quinella*, *Sharpea* and *Kandleria* (Figure 3.13) was used as a reference to compare the distribution and abundance of the species in these genera in the 228 sheep rumen samples analysed in section 3.2.1. This was done by mapping the repset sequences from all sheep onto the tree using the parsimony insertion tool. Most of the possible *Quinella* species (Figure 3.16) were found in the set of analysed rumen samples, except *Quinella* candidate species 5 and *Quinella* candidate species 7. Many of the short reads formed clusters that grouped separately from the earlier-defined species. However, grouping these into the previously-defined species is not easily done because of their limited lengths (approx. 400 bp). They can, however, be used to get an insight into the diversity of *Quinella* species in these rumen samples. Some clusters, such as cluster 3, *Quinella* candidate species 2, *Quinella* candidate species 6 and *Quinella ovalis*, were more abundant than other *Quinella* clusters (Figure 3.16 and Table 3.5). Interestingly, some singleton sequences in the refined tree (Figure 3.13) formed clusters with the short sequences from the sheep rumens. Clusters 3, 7,10, 11 and 12, which were each represented by just one long-length *Quinella* sequence in the original tree (Figure 3.13) now clustered with 8, 8, 3, 11 and 8 repset sequences which represented OTUs with 2255, 705, 697, 501 and 78 pyrosequencing reads, respectively (Figure 3.16 and Table 3.5). Six clusters (clusters 2, 6, 7, 8, and 10) in the genus *Quinella*, two in *Selenomonadaceae* candidate genus 1 and one cluster adjacent to *Selenomonadaceae* candidate genus 1 did not contain any long-length *Quinella* sequences but all contained ≥3 repset sequences that represented OTUs with 71 to 3176 pyrosequencing reads. Additionally, there were also some singleton repset sequences that did not group into any cluster (Figure 3.17). These singleton sequences were from OTUs that contained up to 862 pyrosequencing reads. These findings suggest that many more potential *Quinella* species may exist for which no long-length 16S rRNA gene sequences were obtained in this study. However, this should be interpreted with caution, as errors in pyrosequencing can lead to over estimation of diversity (Huse et al., 2010; Kunin et al., 2010) , but with ≥71 reads per OTU, these singletons may represent actual organisms. The clusters with the largest number of reads assigned to them were *Quinella* candidate species 6 (9756) and *Q. ovalis* (6454), which also had the largest numbers of discriminating OTUs (Table 3.5).

Sharpea was represented by only one species-level cluster in the refined tree (Figure 3.13), but after inclusion of repset sequences there was evidence for two separate clusters (Figure 3.16). *Sharpea* repset1, a newly formed cluster, consisted of only repset sequences whereas the pre-existing *Sharpea azabuensis* contained both repset sequences and sequences from the original refined tree (Figure 3.13). This separation of *Sharpea* indicates that there might be second *Sharpea* species present in the sheep rumens that is neither in culture nor represented by long-length 16S rRNA gene sequences from culture-independent studies. *Sharpea* was also found in some samples from high CH_4 sheep, but this may have been associated with their variable CH_4 emissions. Temporal variability in these rumen populations, and their relationship to variations in CH_4 emissions remains to be investigated. Temporal variations in microbial community structure could have a large impact on CH_4 emissions. Methane production is often monitored continuously, for example in respiration chambers, but rumen microbial communities are often assed on samples taken once a day at a defined time.

Repset sequences affiliated with *Kandleria* all clustered with *Kandleria vitulina* (the only described *Kandleria* species to date) with a 100% bootstrap support, indicating that *Kandleria vitulina* might be the sole species of this genus present in the rumens of the sheep used in the present study.

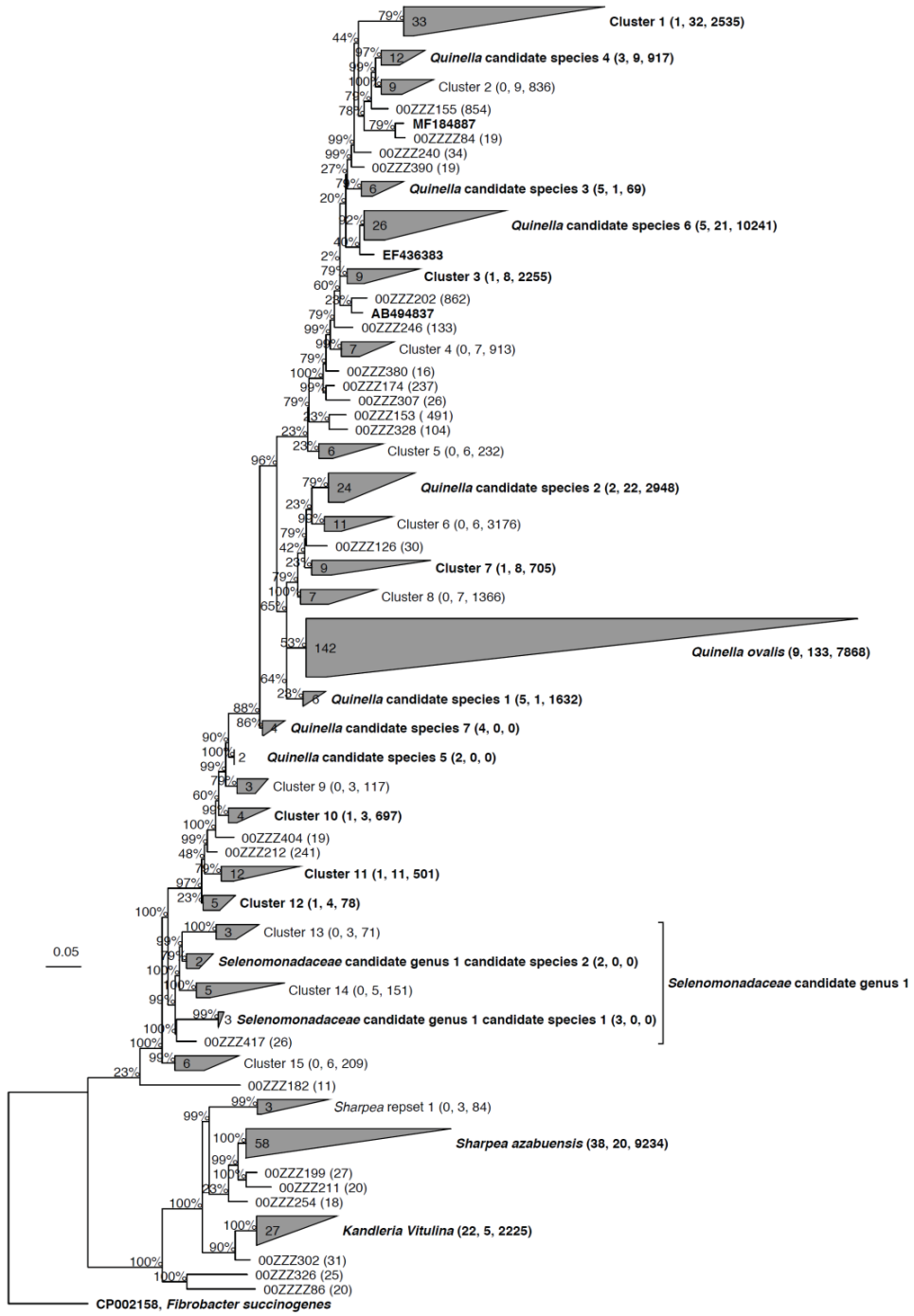


Figure 3.16 Phylogenetic distribution of low methane associated OTUs. Repset sequences representing OTUs with >10 sequences were added to the refined tree shown in Figure 3.13 using the ARB parsimony (Quick add mark) insertion function. Sequences or clusters labelled in non-bold font contain only repset sequences. Numbers in parentheses next to the cluster name show, in this order, the total number of long-length sequences from the refined tree (Figure 3.13), the total number of repset sequences (i.e., OTUs), and the total number of pyrosequencing reads in those OTUs. The 16S rRNA gene sequence of *Fibrobacter succinogenes* (FibSuc43, GenBank accession CP002158) was used as out group sequence.

Table 3.5 OTU distribution in clusters of the genera *Quinella*, *Sharpea*, and *Kandleria* in rumen samples classified as H-, Q-, and S-community types. All clusters and the singleton repset sequences representing >20 pyrosequencing reads, from Figure 3.16, are shown.

Clusters (as they appear in Figure 3.16)	Total sequences from pyrosequencing reads	Sequences in H-type community	Sequences in Q-type community	Sequences in S-type community	Number of discriminating OTU sequences[a]
Cluster 1	2535	2418	95	22	2214
Quinella candidate species 4	917	883	18	16	881
Cluster 2	836	781	47	8	399
00ZZZ155[C]	854	823	23	8	854
00ZZZ240	34	34	0	0	34
Quinella candidate species 3	69	63	4	2	69
Quinella candidate species 6	10241	10038	168	35	9756
Cluster 3	2255	2088	143	24	2207
00ZZZ202[C]	862	825	31	6	862
00ZZZ246[C]	133	126	7	0	133
Cluster 4	913	880	27	6	748
00ZZZ174[C]	237	233	1	3	237
00ZZZ307[C]	26	26	0	0	26
00ZZZ153[C]	491	491	0	0	491
00ZZZ328[C]	104	104	0	0	104
Cluster 5	232	225	7	0	177

Clusters (as they appear in Figure 3.16)	Total sequences from pyrosequencing reads	Sequences in H-type community	Sequences in Q-type community	Sequences in S-type community	Number of discriminating OTU sequences[a]
Quinella candidate species 2	2948	2796	140	12	2325
Cluster 6	3176	3101	65	10	3042
00ZZZ126[C]	30	30	0	0	30
Cluster 7	705	678	24	3	623
Cluster 8	1366	1261	100	5	1134
Quinella ovalis	7868	7288	505	75	6454
Quinella candidate species 1	1632	1477	149	6	1632
Cluster 9	117	103	9	5	83
Cluster 10	697	606	62	29	29
00ZZZ212[C]	241	190	42	9	241
Cluster 11	501	452	41	8	318
Cluster 12	78	72	5	1	39
Cluster 13	71	69	2	0	57
Cluster 14	151	137	14	0	107
00ZZZ417[C]	26	22	4	0	26
Cluster 15	209	168	30	11	126
Sharpea repset 1	84	1	5	78	-[b]
Sharpea azabuensis	9234	814	1821	6599	8918

Clusters (as they appear in Figure 3.16)	Total sequences from pyrosequencing reads	Sequences in H-type community	Sequences in Q-type community	Sequences in S-type community	Number of discriminating OTU sequences[a]
00ZZZ199[C]	27	1	12	14	27
Kandleria Vitulina	2225	253	278	1694	2155
00ZZZ302[C]	31	0	1	30	31
00ZZZ326[C]	25	3	4	18	25

[a]These OTUs were identified as discriminating OTUs (section 3.2.4).

[b]–, no sequences or discriminating OTUs.

[c]Singleton repset sequences that did not fall in any cluster but represented >20 pyrosequencing reads.

.

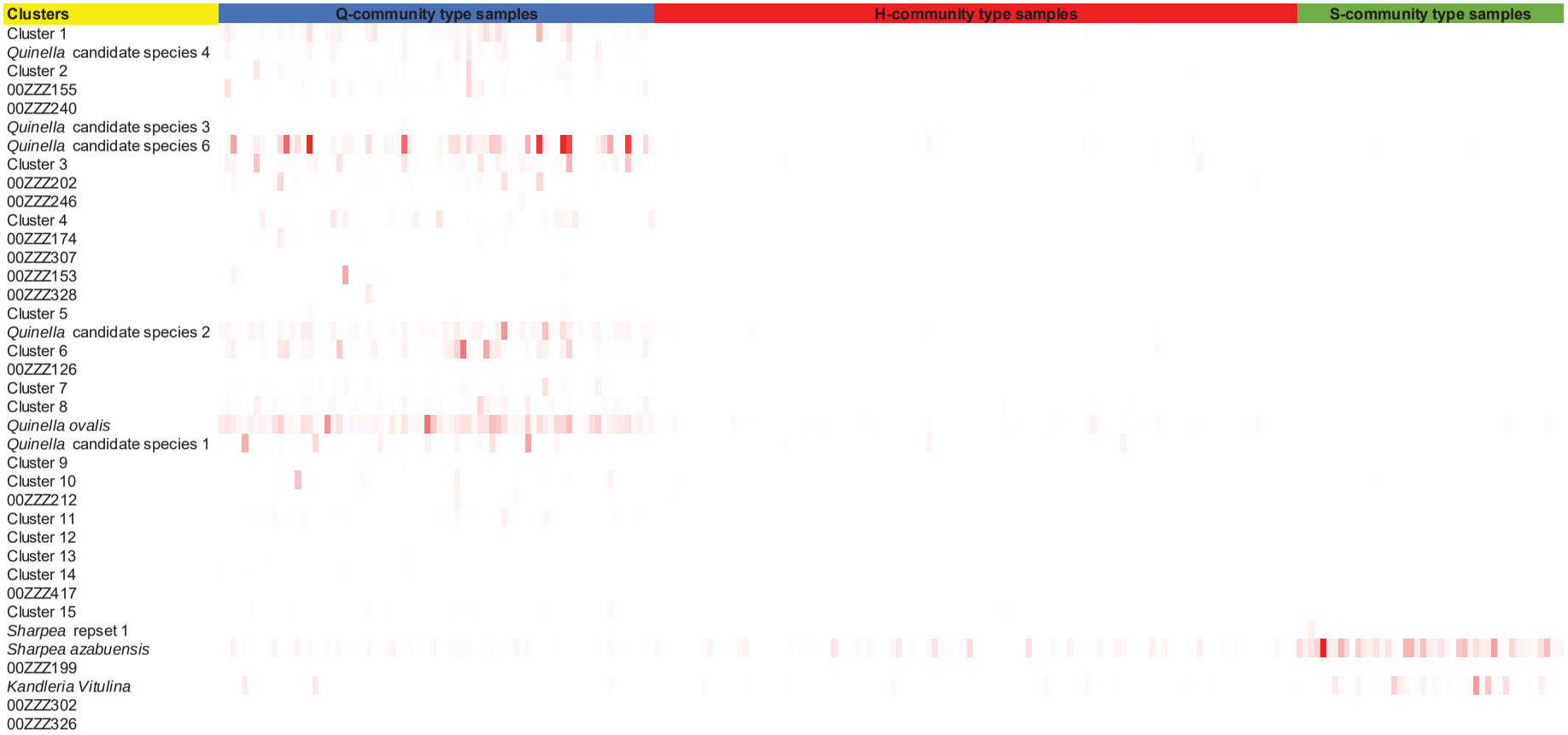


Figure 3.17 Heat-map of sequence read abundance grouped by sequence clusters in the genera *Quinella*, *Sharpea*, and *Kandleria* (Figure 3.16) in 228 rumen samples classified as H-, Q-, and S-type communities. Only clusters or OTUs that contained >20 sequences are included in the heatmap.

3.3 Conclusions

Analysis of high-quality 16S rRNA gene sequences suggested that there are at least seven additional species in the genus *Quinella*, which is currently represented by only one named species, *Q. ovalis*. In contrast, the species diversity in the genera *Sharpea* and *Kandleria* was more limited. There appears to be only one species in the genus *Sharpea*, *S. azabuensis* and only one in the genus *Kandleria*, *K. vitulina*.

Based on short reads generated by 454 pyrosequencing, it also appears that multiple species of the genus *Quinella* were present in the rumen of the sheep studied here. Some of the potentially new *Quinella* species were found to be comparatively more abundant than others in Q-type community samples. However, there were also some *Quinella* species which were absent (*Quinella* candidate species 5 and *Quinella* candidate species 7) from analysed samples, suggesting that these species may be rare or have different hosts or be favoured by different diets. Diet and host are known to shape rumen microbial community structure (Henderson et al., 2015). Additionally, analysis of the short reads indicated that there may be still other *Quinella* species currently not represented by high-quality long-length sequences. Phylogenetic reconstruction based on short-length sequences can be problematic (Quince et al., 2009; Rajendhran and Gunasekaran, 2011; Poretsky et al., 2014), so caution is needed, but at least 12 new clusters were detected that could represent additional species, based on the short reads from the 228 rumen samples studied here.

Short-read sequences from the sheep rumen samples formed two clusters within the genus *Sharpea*: *S. azabuensis* and a new cluster that was not represented by any long-length sequences. This suggested that there could have been a new species of this genus in these sheep. The short read sequences affiliated with *Kandleria* all grouped into the cluster defined by the type species, *K. vitulina*. It was interesting to find that, even though *Sharpea* was abundant in S-type community samples, it was not absent in other community types, and its presence, at lower abundance, in H- type community samples was clearly seen (Figure 3.17).

The association of *Sharpea* and *Kandleria* with low CH_4 yield sheep rumen is a recent finding (Kittelmann et al., 2014) and has only been studied using 16S rRNA gene, metagenomic and metatranscriptomic studies (Kamke et al., 2016). Investigations of the physiology of these bacteria may help in gaining a better understanding of their role in

low CH_4 emissions from sheep. Because no isolates of *Quinella* are available, and the physiology of members of this genus is unclear. The role of *Quinella* in a low-CH_4-producing rumen is unknown. To understand the physiology of *Quinella*, more information is needed either using culture-dependent (isolation) or culture-independent (metagenomic and transcriptomic) methods. VFA concentrations in the rumens of sheep associated with the different community types suggested that there may be more propionate formed in *Quinella*-rich rumens than in those containing S- or H-type communities (Figure 3.12), which could be expected to result in less CH_4 production. In contrast, the acetate:propionate ratio of samples containing S-type communities did not clearly indicate how this type might result in low CH_4 emissions. Sheep with *Sharpea*-rich rumens were studied in detail using culture independent methods by Kamke et al. (2016). They inferred that the *Sharpea* present in low CH_4 emitting sheep produced lactate, which was further converted to butyrate by *Megasphaera* spp. This hypothesis was based on differences in gene and transcript abundances in the rumen metagenomes of high and low CH_4 emitting sheep, although *Megasphaera* spp. were not detected as being significantly different in the study reported here. The study of Kamke et al. (2016) was performed later using a subset of the same sheep, and this raises questions about the stability or permanence of rumen communities. It did appear that the rumen microbial communities of the sheep in the Kamke et al. (2016) study remained colonised by the activity of *Sharpea* spp., after they had been identified as S-type by Kittelmann et al. (2014). Production of lactate and butyrate in *Sharpea*-rich sheep rumens (S-type community sheep) leads to less methane production in comparison to production of acetate and butyrate formed by members of *Lachnospiraceae* and *Ruminococcaceae* (Kamke et al., 2016). However, isolates of *Sharpea* and *Kandleria* need to be studied in detail to understand their fermentation products in pure culture and it would also be interesting to see whether their fermentation pattern is influenced by the presence of methanogens. Some of these questions about the physiology of *Sharpea*, *Kandleria*, and *Quinella* form the basis of the research described in the next chapters (chapters 4, 5, and 6).

Chapter 4 Physiology of *Sharpea* and *Kandleria*

4.1 Introduction

Sharpea and *Kandleria* were identified as key bacteria associated with the S-type bacterial community, one of the two low-CH_4 community types in sheep rumen samples (Chapter 3). Bryant et al. (1958) isolated one strain of each genus (RL1 = *Sharpea azabuensis* DSM 20406 and RL2 = *Kandleria vitulina* DSM 20405) from calf rumen samples, but, on the basis of morphology, gas production, growth temperature and carbohydrate fermentation patterns, they classified both strains as members of *Lactobacillus lactis*, which is in the family *Lactobacillaceae*. Later, on the basis of cell wall composition in addition to the phenotypic characteristics already known, Sharpe et al. (1973) designated them as a new species of *Lactobacillus* and named them *Lactobacillus vitulinus* together with two new strains (CL1 and CL2). Twenty-five years later, Morita et al. (2008) reported four bacterial strains (ST18, DI149, HM250 and HM244) from horse faeces. Phylogenetic analysis using 16S rRNA gene sequences and DNA-DNA hybridisation showed that the newly isolated strains constituted a new genus and species, which was named *Sharpea azabuensis*. The 16S rRNA gene of strain RL1 (Bryant et al., 1958) has 99.7% 16S rRNA sequence similarity with the homologous gene in ST18 (DSM 18934), which suggests that RL1 is a strain of *Sharpea azabuensis*, and not a member of *L. vitulinus* (chapter 3). Salvetti et al. (2011) then reclassified *L. vitulinus* as *Kandleria vitulina* on the basis of 16S rRNA gene sequence similarity. This newly classified *K. vitulina* also contains the other strain (RL2) that was isolated by Bryant et al. (1958). Both genera belong to the family *Erysipelotrichaceae*, and not to the family *Lactobacillaceae*.

A metagenomic and metatranscriptomic study conducted by Kamke et al. (2016) found that the relative abundance of *Sharpea* spp. was greater in low-CH_4 emitting sheep than in high-CH_4 emitting sheep. Furthermore, they observed the D-lactate dehydrogenase that was found to be associated with low-CH_4 emitting sheep was from *Sharpea* and *Kandleria*, and postulated that these bacteria produce lactate in the rumen. They also suggested that the lactate produced by these bacteria is further converted to butyrate by *Megasphaera* spp., resulting in lower CH_4 emissions from these low-CH_4 yield sheep compared to high-CH_4 yield sheep.

In the rumen, H_2 and formate produced by bacteria are used by methanogens to make CH_4. It has been previously shown that low partial pressures of H_2 can stimulate thermodynamically more favourable fermentation pathways that result in more H_2 formation and *vice-versa* (Iannotti et al., 1973; Wolin, 1976; Rees et al., 1995; Chin and Janssen, 2002). This is termed interspecies H_2 transfer. In this process, H_2 producers and users together regulate the partial pressure of H_2 in rumen, which consequently influences the amount of CH_4 that is formed. Many studies have shown this characteristic of microbes in defined mixed cultures and in mixed communities (Bryant et al., 1967; Conrad et al., 1986). Interspecies formate transfer operates in a similar way, with formate use by methanogens or other formate-using microbes postulated to favour increased formate production by bacteria that can do so (Boone et al., 1989; Wu et al., 1996; McInerney et al., 2008; Rotaru et al., 2012). The result is that, under rumen-like conditions where the partial pressure of H_2 and formate are low (van Lingen et al., 2016), production of reduced fermentation products like lactate can be decreased in favour of acetate and H_2 (or formate) formation (Janssen, 2010). This possibility was tested in the work described in this chapter, by co-culturing *Sharpea* or *Kandleria* with a H_2- and formate-using methanogen. The hypothesis was that, if the model suggested by Kamke et al. (2016) is correct, then the production of lactate by *Sharpea* and *Kandleria* should not be markedly affected by the presence of a methanogen using any H_2 or formate produced by the bacteria. Conversely, if these bacteria stopped producing lactate and formed products that can be used by methanogens when these products are removed by those methanogens, then the hypothesis of Kamke et al. (2016) would not be supported.

4.2 Results and discussion

4.2.1 Substrates that support growth of *Sharpea* and *Kandleria*

Four strains each of *Sharpea* and *Kandleria* were used for initial fermentation studies. For *Sharpea*, RL1 (DSM 20406) from Bryant et al. (1958), ST18 (DSM 18934) from Morita et al. (2008), and two isolates from the Hungate1000 project (KH1P5 and KH2P10; Creevey et al., 2014) were used. For *Kandleria*, RL2 (DSM 20405) from Bryant et al. (1958), WCE2011 and MC3001 from Noel (2013), and KHCV7 from the Hungate1000 project were used. $ST18^T$ was isolated from horse faeces; while; all other strains were isolated from ruminants. Each of these strains was found to be pure based on

the cell type observed by microscopy, which matched the formal descriptions of *S. azabuensis* ST18^T (Morita et al., 2008) and *K. vitulina* RL2^T (Salvetti et al., 2011). Sequencing the 16S rRNA genes of these cultures yielded the expected sequences without contaminating peaks in the electropherograms. Substrates used for this experiment were prepared according the method described in section 2.12.2. All eight cultures were inoculated into growth medium supplemented with different growth substrates. Of the 21 substrates tested, only D-glucose, D-fructose, D-galactose, cellobiose, and sucrose were used by all eight strains of *Sharpea* and *Kandleria* (Table 4.1). However, the amount of growth by *Kandleria* WCE2011 was generally less than the other strains. Lactose supported growth of seven of the strains, and raffinose of five. These findings are similar to those of Sharpe et al. (1973) for strains RL1 and RL2, as none of the tested isolates were able to ferment xylose, arabinose, glycogen, and glycerol. None of the strains grew on some common organic and amino acids.

Table 4.1 Substrates that supported growth of strains of *Sharpea* and *Kandleria*.

Substrate	Conc[a] (mM)	*Sharpea* strains				*Kandleria* strains			
		ST18^T	RL1	KH1P5	KH2P10	RL2^T	MC3001	WCE2011	KHCV7
D-Glucose	4	+++[b]	+++	++	++	+++	+++	+	+++
D-Fructose	4	++	++	++	++	+++	+++	++	+++
D-Galactose	4	+++	+++	+++	+++	+++	+	++	+++
D-Glucuronate	4	–	–	–	–	–	–	–	–
D-Galacturonate	4	–	–	–	–	–	–	–	–
Cellobiose	2	+++	+++	+++	+++	+++	+++	+	+++
Sucrose	2	+++	++	++	++	+++	+++	+	+++
Lactose	2	+++	++	++	+++	+++	+	–	+++
Raffinose	2	+++	+++	+++	+++	+++	–	–	(+)
D-Xylose	5	–	–	–	–	–	–	–	(+)
L-Arabinose	5	–	–	–	–	–	–	–	(+)
L-Rhamnose	5	–	–	–	–	–	–	–	(+)
Glycerol	20	–	–	–	–	–	–	–	(+)
D-Mannitol	4	–	–	–	–	–	–	–	–
L-Glutamate	20	–	–	–	–	–	–	–	–
L-Alanine	20	–	–	–	–	–	–	–	–
L-Aspartate	20	–	–	–	–	–	–	–	–
Citrate	20	–	–	(+)	–	–	(+)	–	–
Fumarate	20	–	–	–	(+)	–	–	–	–
Succinate	20	–	–	–	–	–	–	–	–
L-Lactate	20	–	–	–	–	–	(+)	–	–

[a]Conc., initial concentration in growth medium.

[b]Optical density was measured at 600 nm. Symbols: **+++**, ΔOD> 0.4; **++**, ΔOD 0.2-0.4; **+**, ΔOD 0.1-0.2; **(+)**, ΔOD 0.05-0.1, **-**, ΔOD < 0.05; all relative to controls without added substrate. The data are means of three replicate cultures.

Table 4.2 End products from fructose fermentation by *Sharpea* and *Kandleria*.

Genus	Strain	Fructose remaining[a,b] (mM)	Products formed (mM)[b]			
			Lactate	Formate	Acetate	Ethanol
Sharpea	ST18^T	0.01	17.96	2.79	1.92	0.69
	RL1	0.01	15.94	1.41	0.76	<0.01
	KH1P5	0.01	15.05	1.32	0.68	0.12
	KH2P10	0.06	15.56	1.40	1.31	0.60
Kandleria	RL2^T	0.05	15.44	1.57	0.78	1.03
	MC3001	<0.01	14.07	1.98	1.09	0.42
	WCE2011	<0.01	15.30	1.95	0.94	0.40
	KHCV7	0.05	16.14	1.90	1.07	0.17

[a]Initial fructose concentration was 10 mM.
[b]All data are means of three replicate cultures.

Table 4.3 Lactate isomers formed by *Sharpea* and *Kandleria*.

Genus	Strain	Lactate concentration (mM)[a]	
		D-lactate	L-lactate
Sharpea	ST18^T	16.84	0
	RL1	18.91	0
	KH1P5	13.37	0
	KH2P10	17.27	0
Kandleria	RL2^T	10.33	0
	MC3001	11.16	0.05
	WCE2011	8.64	0
	KHCV7	10.89	0.79

[a]All data are means of three replicate cultures.

4.2.2 End products of *Sharpea* and *Kandleria*

The end products of fermentation by *Sharpea* and *Kandleria* in pure culture were determined by growing isolates in medium containing 10 mM D-fructose. The methods are described in section 2.5.3. All isolates produced lactate as the major end product (~75% of total end product; Table 4.2). Small amounts of formate, acetate and ethanol were also produced. Lactate production was reported by Sharpe et al. (1973) and Salvetti et al. (2011) for strains that were eventually classified as *Sharpea* and *Kandleria*, but the other products were not previously reported. The *Sharpea* and *Kandleria* strains tested here produced almost exclusively D-lactate (Table 4.3).

4.2.3 Is the fermentation pattern of *Sharpea* and *Kandleria* influenced by a methanogen?

It was clear that *Sharpea* and *Kandleria* produced lactate as their major end product. However, the presence of small amounts of acetate, formate and ethanol indicate that while these bacteria carry out a predominately homolactate fermentation, there is a minor role for a mixed acid fermentation in their metabolism (Figure 4.1C). It has been reported that, in particular circumstances, homolactate fermenters (Figure 4.1A) can change their behaviour and performed a mixed acid fermentation (Figure 4.1B) to varying extents depending on the conditions (Thomas et al., 1979; Neves et al., 2005; Teusink et al., 2011). Because small amounts of acetate, formate and ethanol were being produced, it was logical to test whether the fermentation products of *Sharpea* and *Kandleria* were influenced by the presence of a methanogen.

Methanobrevibacter olleyae was chosen for co-culture experiments with *Sharpea* or *Kandleria* as it can use both H_2 and formate as electron donors (Rea et al., 2007). The experiment was conducted as shown in Figure 4.2 (see also section 2.5.2 for details). Basically, two separate experiments were set up. In the first experiment ("After"), *Sharpea* or *Kandleria* were first grown with 10 mM fructose and *M. olleyae* was added later, after any H_2 or formate had been formed. It was hypothesised that in this way there should be no change in the general fermentation pattern of the bacteria after *M. olleyae* was added compared to the products detected just before the addition of the methanogen, because the bacteria would have used all the fructose prior to the addition of the methanogens. In this scenario there would be no opportunity for the activity of *M. olleyae* to modify the fermentation patterns of the bacteria. In the second experiment ('Before'),

M. olleyae was grown first to establish a dense active population, before *Sharpea* or *Kandleria* were added together with 10 mM fructose. It was speculated that if *Sharpea* and *Kandleria* performed a mixed acid fermentation that was influenced by end-product use by the methanogen, then in the presence of *M. olleyae* the final lactate concentration would decrease, and acetate and perhaps ethanol concentrations would increase (Figure 4.1D). However, the dominance of lactate as the major end product did not change (p value > 0.9; ANOVA) when *Sharpea* and *Kandleria* were grown in the presence of *M. olleyae* compared to when there was no methanogen present, or when the methanogen was added after the bacteria had grown. The increased acetate concentration in the 'After' experiment was due to addition of 0.77 to 1.16 mM extra acetate while inoculating *M. olleyae* into the grown cultures of *Sharpea* or *Kandleria*. Similarly, 0.64 to 1.07 mM formate remained in the dense *M. olleyae* cultures when *Sharpea* and *Kandleria* strains were added to the 'Before' experiment. This residual formate was then used by *M. olleyae* together with any formate formed during fructose fermentation by the bacteria. Overall, it appeared that even in co-culture with the methanogen, *Sharpea* and *Kandleria* continued to carry out a predominantly homolactic acid fermentation (Table 4.4, Figure 4.1C).

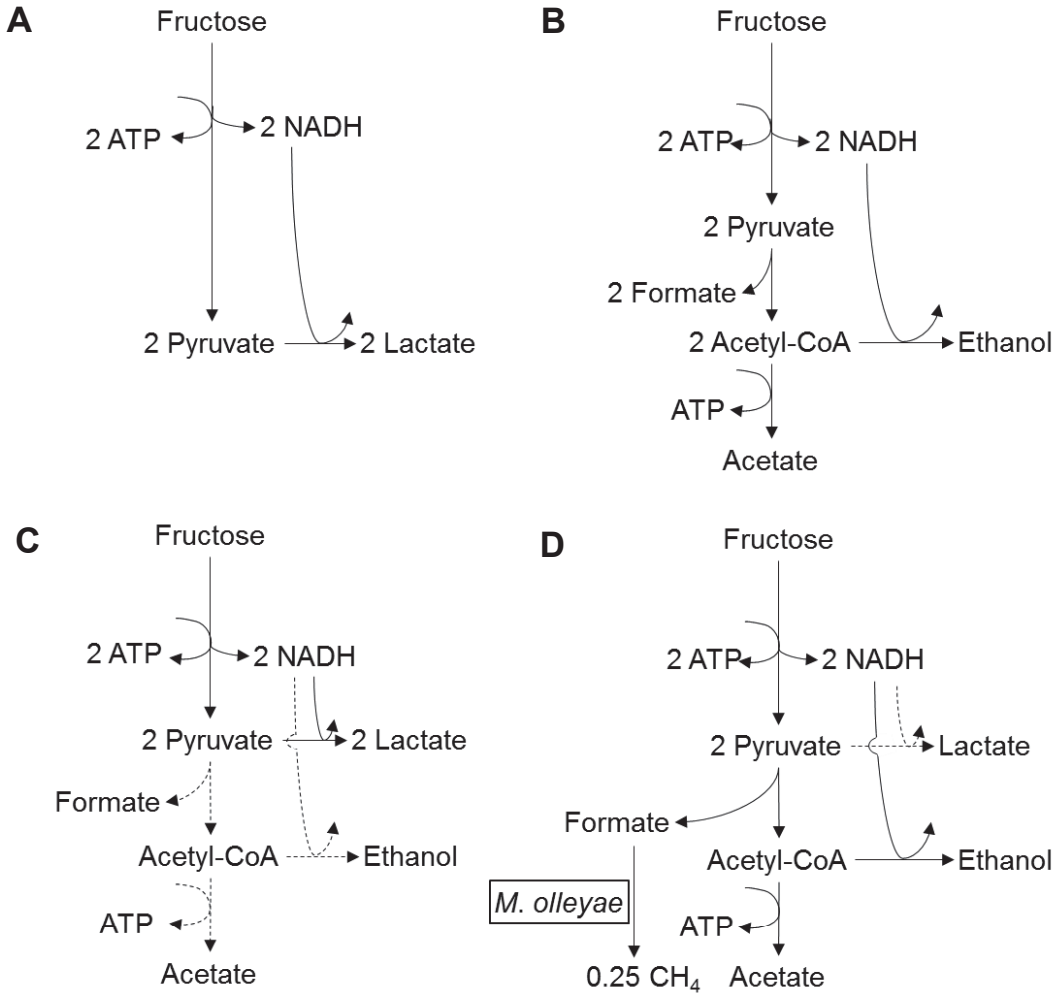


Figure 4.1 Possible fermentation schemes discussed in this chapter. A) Homolactic fermentation. B) Mixed acid fermentation. C) Predominantly lactate formation by *Sharpea* and *Kandleria* with a minor role for formate, acetate and ethanol formation. D) Predicted change in end-product formation in the case of inter-species formate transfer or change in fermentation pattern from homolactic to mixed acid fermentation in the presence of *M. olleyae*. Dashed arrows indicate less active pathways. Note that the schemes in panels C and D are not fully balanced.

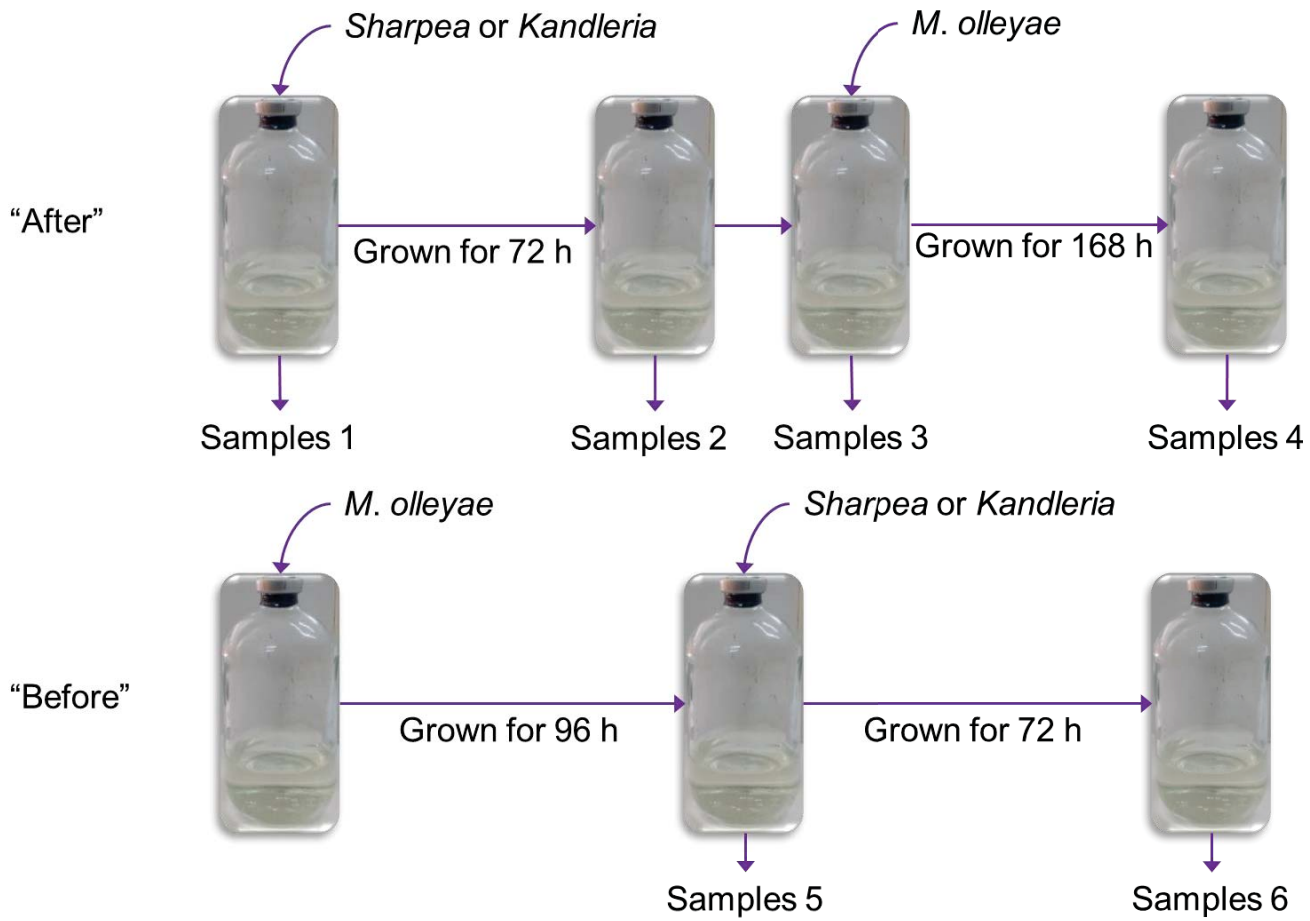


Figure 4.2 Experimental design of co-culture experiments of *Sharpea* or *Kandleria* with *Methanobrevibacter olleyae*. Two separate experiments were conducted: "After", in which *M. olleyae* was added to grown cultures of *Sharpea* or *Kandleria*, and "Before", in which *Sharpea* or *Kandleria* were inoculated into an active culture of *M. olleyae*. Samples for end product measurements were taken at different time points (denoted as Samples 1 to 6).

Table 4.4 Changes in concentrations of substrate and fermentation products in cultures of *Sharpea* or *Kandleria* with and without *Methanobrevibacter olleyae*. For experiment details see Figure 4.2.

Stages of experiment	Isolate	Products (mM)[a]				
		Fructose	Lactate	Formate	Acetate	Ethanol
Pure culture growth of *Sharpea/ Kandleria* (2-1)[b]	ST18^T	−10.96	18.00	1.98	1.13	0.29
	KH1P5	−12.23	20.72	1.48	1.30	0.29
	WCE2011	−10.05	15.63	1.96	1.01	0.13
	RL2^T	−9.82	15.61	1.63	1.01	0.23
"After". *M. olleyae* added to grown cultures of *Sharpea/ Kandleria* (4-1)	ST18^T	−11.01	17.76	−0.01	2.27[d]	0.23
	KH1P5	−12.27	19.99	0.00	2.56[d]	0.20
	WCE2011	−10.05	15.47	0.51	2.09[d]	0.07
	RL2^T	−9.82	15.82	0.00	2.24[d]	0.19
"Before". *Sharpea/ Kandleria* added to grown culture of *M. olleyae* (6-5)	ST18^T	−9.72	15.21	−0.93[c]	3.13	0.20
	KH1P5	−10.77	18.53	−0.22[c]	1.24	0.14
	WCE2011	−9.87	17.76	−0.95[c]	1.56	0.06
	RL2^T	−8.96	15.55	−0.64[c]	1.46	0.07

[a]Negative values indicate decrease in concentration. All data are the means of three replicate cultures.

[b]Changes in concentration were calculated from the differences at different sampling points (indicated by the numbers 1 to 6; see Figure 4.2 for details).

[c]0.64 to 1.07 mM formate remaining when *Sharpea* and *Kandleria* strains were added, which was then used by *M. olleyae* together with any formate formed during fructose fermentation.

[d]0.77 to 1.16 mM extra acetate added when *M. olleyae* was inoculated into cultures with *Sharpea* or *Kandleria* strains (sample 3 in Figure 4.2).

4.3 Conclusions

In vitro analysis of *Sharpea* and *Kandleria* confirmed previous findings (Sharpe et al., 1973; Salvetti et al., 2011) that these behave like classical lactic acid bacteria that produce lactate as their major end product. Small amounts of acetate, formate and ethanol were produced from fructose, suggesting that they might perform a mixed acid fermentation in certain circumstances. The production of formate also raised the possibility that, in the presence of a H_2 or formate using methanogen, these bacteria might switch to producing less lactate. However, co-culture experiments with *M. olleyae* (a methanogen that can use H_2 as well as formate) refuted this hypothesis, as lactate remained the major end product. The cultures were constructed to contain a dense active culture of methanogens. The small amounts of acetate, formate and ethanol produced in both the pure and mixed cultures may be products from a background mixed-acid fermentation that yields acetyl-CoA and reducing potential for the cell, or for some other purpose.

The findings of this study support the proposal by Kamke et al. (2016) that, in the complex rumen system, *Sharpea* and *Kandleria* produce mainly lactate. Because lactate formation, with further conversion of lactate to butyrate by the action of *Megasphaera* spp., results in less H_2 formation than the classical direct fermentation of carbohydrates to butyrate by bacteria such as members of the family *Ruminococcaceae*, less CH_4 is formed in the rumen of sheep with greater populations of *Sharpea* and *Kandleria* (Kamke et al., 2016).

Chapter 5 Investigation of *Quinella* through genome analysis from metagenomic DNA

5.1 Introduction

Quinella was one of the relatively abundant and significantly associated bacterial groups in the rumen of low methane emitting sheep (Chapter 3). Cells that are now thought to be members of the genus *Quinella* were first reported in 1913 by Woodcock and Lapage (1913), who observed them during microscopy of goat rumen contents, but mistook them for protistan parasites because of their very large cell size. Thirty years later, Quin (1943) observed *Quinella* in the rumen of sheep and thought that these large cells were a "pseudo-yeast" to which the name *Schizosaccharomyces ovis* was applied. Later, it was found that *Quinella* was not a yeast, as it could not be grown in standard yeast media and was motile and gram-negative (van der Westhuizen et al., 1950; Wicken and Howard, 1967a). On the basis of comparative analyses of nearly full-length 16S rRNA gene sequences, *Quinella* was found to be closely related to members of the *Selenomonas-Megasphaera-Sporomusa* group of bacteria (Krumholz et al., 1993), which is now designated as the class *Negativicutes* (Marchandin et al., 2010). Analysis of the diversity of 16S rRNA sequences affiliated with *Quinella*, performed as part of the current study, suggested that there is more than one species present in the sheep rumen and also showed that the relative abundance of the different candidate species varied from animal to animal (Chapter 3). To understand the physiology of *Quinella*, culture dependent and culture independent approaches could be applied (for a review of these, section 1.8). There have been some isolation attempts by Purdom (1963) and Orpin (1972), as well in this study (section 5.2.3), but no one has succeeded in isolating and maintaining a pure culture of *Quinella*. Two fermentation studies, one *in vitro* (Brough et al., 1970) and the other *in vivo* (Vicini et al., 1987), were conducted to understand *Quinella* physiology. However, the results from these studies contradict each other. Brough et al. (1970) suggested that the preferred substrate, glucose, is fermented mainly to lactate by enriched *Quinella* cells, with traces of acetate, propionate and CO_2 also detected. In contrast, Vicini et al. (1987) reported that when *Quinella* was present as the most abundant bacteria in sheep fed molasses, acetate and propionate were the major end products in the rumen, with no lactate formation.

There has been a constant development of methods for the assembly of draft and even complete genomes from metagenomic datasets (Wu et al., 2014; Kang et al., 2015; Parks et al., 2015). Metagenome samples from the marine environment (Iverson et al., 2012), human gut (Sharon et al., 2013) and mine drainage channels (Tyson et al., 2004; Hua et al., 2015), among others, have been used to construct draft genomes representing cultured and uncultured microbes. However, there are only a few studies that report partial genome reconstruction from rumen metagenomic samples (Hess et al., 2011; Solden et al., 2017; Svartstrom et al., 2017). Hess et al. (2011) reconstructed 15 genomes from metagenomes of microbes adhering to switchgrass in the cow rumen. These genomes were 60.41% to 92.98% complete and the majority of them belonged to members of the orders *Bacteroidales*, *Spirochaetales* and *Clostridiales*. Solden et al. (2017) reconstructed four genomes of the uncultured bacterial genus BS11 from metagenomes recovered from moose rumen fluid samples, which helped in understanding the hemicellulose fermentation potential of this genus. A similar study conducted by Svartstrom et al. (2017) reconstructed 99 genome from the moose rumen microbiome. Of those, 68 were represented members of the phylum *Bacteroidetes* and most of them were associated with a newly-identified clade of *Bacteroidetes* that may possess fibrolytic activity. To gain insights into the physiology of *Quinella,* a metagenomic approach was applied to cells that were physically enriched from sheep rumen samples that contained large populations of members of this genus. This work is described in this chapter.

5.2 Results and Discussion

5.2.1 Fluorescence *in situ* hybridisation (FISH) probe design for *Quinella*

The association of *Quinella* with the Q-type community was shown in chapter 3, based on short 16S rRNA sequencing reads generated by 454 pyrosequencing. Those findings used total DNA extracted from sheep rumen samples. The next step was to confirm the physical presence of *Quinella* in rumen samples by microscopy. Fluorescence *in situ* hybridisation (FISH) was used to examine rumen samples for *Quinella*, and to confirm their presence in the rumen. A *Quinella*-specific FISH probe has not been reported before, based on searches in NCBI, ProbeBase (Greuter et al., 2016) and published articles. A *Quinella*-specific FISH probe was designed based on long length sequences from the newly-refined *Quinella* phylogenetic tree (section 3.2.9). After several rounds of design

and testing (section 2.11 for methods of probe design and testing), five FISH probes were selected for further study.

Table 5.1 *Quinella*-specific FISH probes designed for this study.

Probe name	Sequence (5' to 3')	Target position (*E. coli*)[a]	G+C content (mol%)	Melting temperature (°C)
Quin1231	TTCAGCCCATTGTAGTAC	1231-1248	44.4	48.8
Quin130A	CCCGACTTTGCGGCAGAT	130-147	61.1	58.4
Quin130B	TCCCCGTCTTTGCGGCAG	132-149	66.7	61.0
Quin130C	TCCCCGACTTCGCGGCAG	132-149	72.2	63.2
Quin130D	CCCAACTTTGCGGCAGAT	130-147	55.6	56.2

[a]Based on the 16S rRNA gene sequence of *Escherichia coli* as annotated in ARB(Ludwig et al., 2004).

Probe Quin1231 targeted the 16S rRNA of *Quinella* at *Escherichia coli* positions 1231-1248, whereas probes Quin130A-Quin130D targeted approximately *E. coli* positions 130-139 (Table 5.1). The recommended G+C content for optimal probe function is >50% to ensure tight binding with the target, and usually 50% to 70% is sufficient to provide good results (Ludwig et al., 2004). The G+C content of the *Quinella* FISH probes varied from 44.4% to 72.2%, but four of the five *Quinella* probes had >50% GC content (Table 5.1). Similarly, the melting temperatures of these probe varied from 48.8°C to 63.2°C, a reflection of the variation in G+C content (as the higher G+C content is related to a higher melting temperature).

The ARB probe match tool (Ludwig et al., 2004) was applied to assess the specificity of these probes by matching them to 16S rRNA gene sequences, and all five probes were predicted to be specific to *Quinella* 16S rRNAs, assuming no probe-binding mismatches. Probe Quin1231 matched exactly to 43 of the available 44 full-length *Quinella* sequences, whereas probes Quin130A, Quin130B, Quin130C and Quin130D matched to 10, 22, 2 and 2 of the 44 full-length *Quinella* sequences respectively, and 4 of the 5 sequences in the new candidate genus closely related to *Quinella* (Table 5.2). This suggested that a combination of probes Quin130A-Quin130D might target 36 of the 44 *Quinella*

sequences. When the probe match filter was relaxed to one base pair mismatch, probe Quin1231 could potentially bind to 16S rRNA from some members of 28 other bacterial genera, mostly in the family *Selenomonadaceae*. Probe Quin130B could theoretically bind to 16S rRNAs from *Selenomonas ruminantium* and *Selenomonas bovis* with a one base-pair mismatch.

5.2.2 Use of FISH probes with rumen samples

No cultures of *Quinella* were available to test the five new FISH probes. It appears that *Quinella* cells are abundant in some low-CH_4 emitting sheep rumens (section 3.1), and that *Quinella* cells are larger relative to most other bacteria (Woodcock and Lapage, 1913; Quin, 1943; Wicken and Howard, 1967a; Orpin, 1972). This large cell size offered a way of enriching *Quinella* cells from rumen samples for genome sequencing and for studying them using microscopy. Therefore, the probes were applied to rumen samples to obtain preliminary evidence that there were abundant large cells that bound the probes. Probe Quin1231 was used on its own, whiles the four probes Quin130A to Quin130D were used in an equimolar mix (Quin130Mix). Probe specificity was assessed by varying the hybridisation temperature, and the NaCl and formamide concentrations. At 46°C and 48°C, and using 0.4 M NaCl and 40% (v/v) formamide, both Quin1231 and Quin130Mix bound only to large oval cells, indicating that these cells might be *Quinella* (Figure 5.1). Probe Quin1231 was selected for further use, as it covered all potential *Quinella* species identified in chapter 3 and so might be useful for following the enrichment of putative *Quinella* cells based on two criteria: size and FISH-probe binding. Probe Quin1231 bound some of the large oval-shaped cells in rumen samples, but other large cells were not stained (Figure 5.2). This suggests that there were other large cells in the rumen samples that were not *Quinella*, or that some *Quinella* cells were not stained with this probe or were not permeabilized during sample preparation. Other large cell types are known from the rumen, such as *Oscillospira* and Eadie's oval ("*Magnoovum eadii*" (Orpin, 1972; Clarke, 1979; Mackie et al., 2003). In these preparations, none of the smaller cells were stained using Quin1231, although they were stained using the universal bacterial probe.

Table 5.2 Matches of *Quinella* probes to *Quinella* 16S rRNA gene sequences. Sequences are in the same order as they are in the phylogenetic tree shown in Figure 3.11. The cluster names are those used in Figure 3.11.

Sequence (NCBI accession no.)	Sequence clusters[a]	Probe target region[b]	
		Probe Quin1231 (*E. coli* positions 1231-1248)	Probe Quin130A-Quin130D (*E. coli* positions 130-149)
MF184892	*Quinella* candidate species 4	CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
MF184893		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
MF184888		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
MF184887		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
AB494921		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCGAAGUCGGGGACAACG
EF436437	*Quinella* candidate species 3	CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCACAAAGUUGGGGACAACG
EF436435		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCACAAAGUUGGGGACAACG
EF436434		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCACAAAGUUGGGGACAACG
EF436436		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCACAAAGUUGGGGACAACG
EF436332		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUUGGGGACAACG
MF184915	*Quinella* candidate species 6	CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCGAAGUCGGGGACAACG
MF181920		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
MF184919		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
MF184917		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
DQ673489		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
EF436383		CUACACACGUACUACAAUGGG-CUGAACAGAGGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
MF184921		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
AB494837		CUACACACGUACUACAAUGGG-GCGAACAGAGGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG

Sequence (NCBI accession no.)	Sequence clusters[a]	Probe target region[b]	
		Probe Quin1231 (*E. coli* positions 1231-1248)	Probe Quin130A-Quin130D (*E. coli* positions 130-149)
MF184871	*Quinella ovalis*	CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184873		CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184872		CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184874		CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184899		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184876		CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
M62701		CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184889		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
AB494823		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUUUGCCGCAAAGACGGGGACAACG
MF184897	*Quinella* candidate species 1	CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184898		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184895		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184896		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184894		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184914	*Quinella* candidate species 2	CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
MF184915		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
AB494931		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
DQ673510	*Quinella* candidate species 7	CUACACACGUACUACAAUGGG-CUGAACAGAGGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
DQ673569		CUACACACGUACUACAAUGGG-CUGAACAGAGGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
DQ673570		CUACACACGUACUACAAUGGG-CUGAACAGAGGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
EF436320		CUACACACGUACUACAAUGGG-CUGAACAGAGGG	GCGCAGGUAAUCUGCCGCAAAGACGGGGACAACG
EF436424	*Quinella* candidate species 5	CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCCGCCGCAGAGACGGGGACAACG
EF436425		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCCGCCGCAGAGACGGGGACAACG

Sequence (NCBI accession no.)	Sequence clusters[a]	Probe target region[b]	
		Probe Quin1231 (*E. coli* positions 1231-1248)	Probe Quin130A-Quin130D (*E. coli* positions 130-149)
MF184890		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUCGGGGACAACG
EF436426		CUACACACGUACUACAAUGGG-CUGAACAGAAGG	GCGCAGGUAAUCUGCCGCAAAGUUGGGGACAACG
MF184879		CUACACACGUACUACAAUGGG-CUGAACAAAAGG	GCGCAGGUAAUCUGCCGCAAAGAUGGGGACAACG
DQ673559	*Selenomonadaceae* candidate genus 1 species 1	CUACACACGUACUACAAUGGGACGGAACAAAGGG	GCGUAGGUAAUCUGCCGCAAAGUUGGGGGCAACA
DQ673560		CUACACACGUACUACAAUGGGACGGAACAAAGGG	GCGUAGGUAAUCUGCCGCAAAGUUGGGGGCAACA
DQ673499		CUACACACGUACUACAAUGGGACGGAACAAAGGG	GCGUAGGUAAUCUGCCGCAAAGUUGGGGACAACA
MF184882	*Selenomonadaceae* candidate genus 1 species 2	CUACACACGUACUACAAUGGGACGGAACAAAGGG	GCGUAGGUAAUCUGCCGCAAAGAUGGGGACAACA
DQ673561		CUACACACGUACUACAAUGGGACGGAACAAAGGG	GCGUAGGCAAUCUGCCGCAAAGUUGGGGACAACA

[a]Clusters without names are singleton sequences, shown in Figure 3.11.

[b]Probe Quin1231: GUACUACAAUGGGCUGAA; Probe Quin130A: AUCUGCCGCAAAGUCGGG; Probe Quin130B: CUGCCGCAAAGACGGGGA; Probe Quin130C: CUGCCGCGAAGUCGGGGA; Probe Quin130D: AUCUGCCGCAAAGUUGGG. Positions highlighted with red indicate mismatches with one of the five *Quinella* probes.

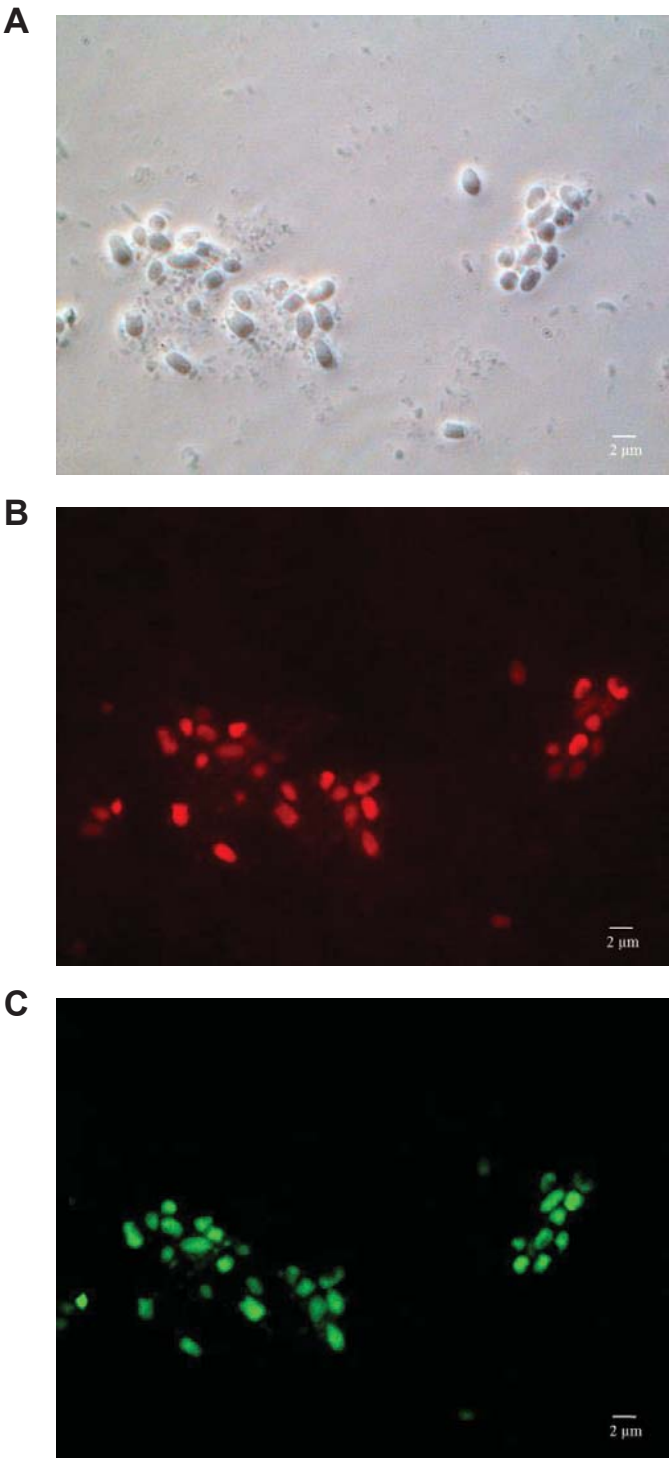


Figure 5.1 Micrographs of rumen bacterial cells hybridised simultaneously with Cy3- and Alexa 488-labelled *Quinella*-specific probes Quin130Mix and Quin1231 respectively. A) Phase contrast image of sample of rumen bacteria. B) The same field showing cells that hybridised with *Quinella*-specific probes Quin130Mix. C) The same field showing cells that hybridised with the *Quinella*-specific probe Quin1231. The scale bar indicates a distance of 2 µm.

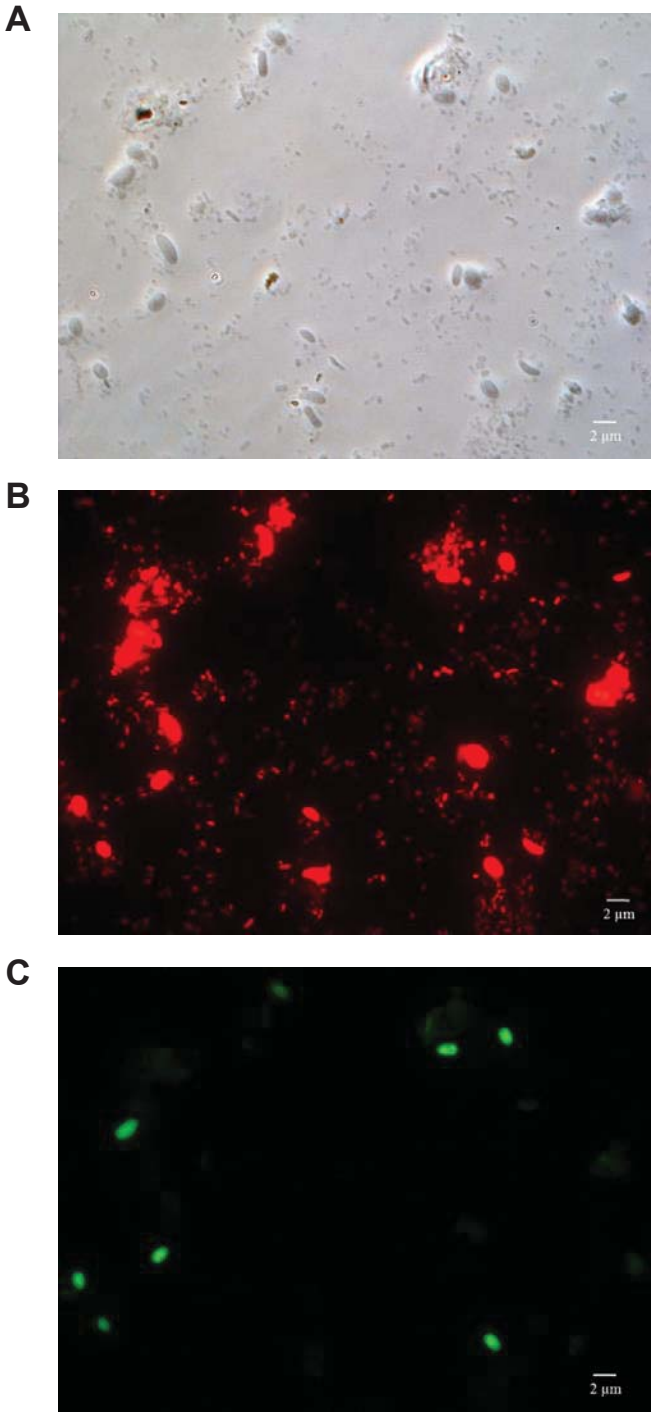


Figure 5.2 Micrographs of rumen bacterial cells hybridised simultaneously with Cy3-labelled universal bacterial probe EUB338 and Alexa 488-labelled *Quinella*-specific probe Quin1231. A) Phase contrast image of sample of rumen bacteria. B) The same field showing cells that hybridised with the universal bacterial probe EUB338. C) The same field showing cells that hybridised with the *Quinella*-specific probe Quin1231. The scale bar indicates a distance of 2 µm.

An attempt was made to use Clone-FISH to validate the *Quinella*-targeted probes. This uses a vector in an *Escherichia coli* host that expresses target 16S rRNA in the *E. coli* cells, which can then be used to test FISH probes (Schramm et al., 2002). 16S rRNA gene sequences from *Quinella* were cloned and used as positive controls, and a sequence from *Selenomonas ruminantium* was used as negative control. However, weak probe binding was observed to all constructs over a wide range of hybridisation conditions. The universal bacterial probe EUB338 functioned as expected in these trials, but of course was able to bind to the native 16S rRNA in the *E. coli* cells.

Probes Quin1231 and EUB338, which binds to 16S rRNA from most bacteria (Amann et al., 1990), were distinguished by labelling them with different fluorophores, and then used together to label cells in rumen contents. Probe Quin1231 hybridised with large oval cells (Figure 5.2C), whereas other bacterial cells can be clearly seen in the phase contrast image of the same field (Figure 5.2A) and clearly hybridised with the universal bacterial probe (Figure 5.2B). The cells that bound the Quin1231 probe were oval and about 3 to 5 µm long and 1 to 3 µm in diameter. These cells occurred singly and not in noticeable clusters, pairs or chains.

5.2.3 Attempt to culture *Quinella*

Once *Quinella* cells were identified by FISH, attempts were made to culture *Quinella* using the methods described in section 2.12.2. The same sheep that had been observed to have large populations of *Quinella* in the development of the FISH methods were sampled to provide inocula for the cultivation attempts. Unfortunately, this was unsuccessful and the cultures of large cells that were isolated (not as large as *Quinella*) were identified as *Staphylococcus warneri* based on their 16S rRNA gene sequences.

5.2.4 Enrichment of *Quinella* cell from rumen liquor samples

From the FISH experiment, it was clear that potential *Quinella* cells were comparatively larger than most other rumen bacteria, and this property was used to prepare suspensions of cells enriched with *Quinella* for further study. Wicken and Howard (1967a) developed a method for *Quinella* enrichment, by incubating rumen liquor with glucose followed by differential centrifugation. Here, a slightly different and quicker method was developed. This involved a series of filtration (using a 23-µm pore-size filter) and low-speed centrifugation steps (section 2.12.3). This cell concentration method resulted in a marked increase in the relative abundance of large oval cells postulated to be *Quinella*, as

observed using phase contrast microscopy (Figure 5.3) and labelled FISH microscopy (Figure 5.4).

Three samples were prepared for extraction of metagenomics DNA. Rumen samples were collected from 24 sheep, and 12 of these that contained large numbers of large oval *Quinella*-like cells (assessed using phase contrast microscopy) were pooled (section 2.12.1). This was designated as "sample 1", and DNA extracted from it contained a large proportion of *Quinella* 16S rRNA genes based on clone library analysis (Table 5.3). Forty seven of 96 cloned sequences were identified as originating from *Quinella*. Two rumen samples from one sheep collected 2 weeks apart were pooled and processed to produce "sample 2" that also appeared to contain a high proportion of *Quinella* cells under phase contrast microscopy and by clone library analysis (Table 5.3). A single rumen sample from one sheep was also processed in the same way to produce "sample 3", which contained enriched *Quinella* cells based on phase contrast microscopy, FISH and clone library analysis (Table 5.3).

Table 5.3 16S rRNA gene clone libraries prepared from *Quinella*-enriched samples.

Sample	Description	No. of clones sequenced	No. of *Quinella* clones identified	Inferred *Quinella* abundance (%)
1	Pooled samples from 12 sheep	111	56	50.4
2	Pooled samples (two samplings 2 weeks apart) from one sheep	95	42	44.2
3	Single rumen sample from one sheep	92	57	61.9

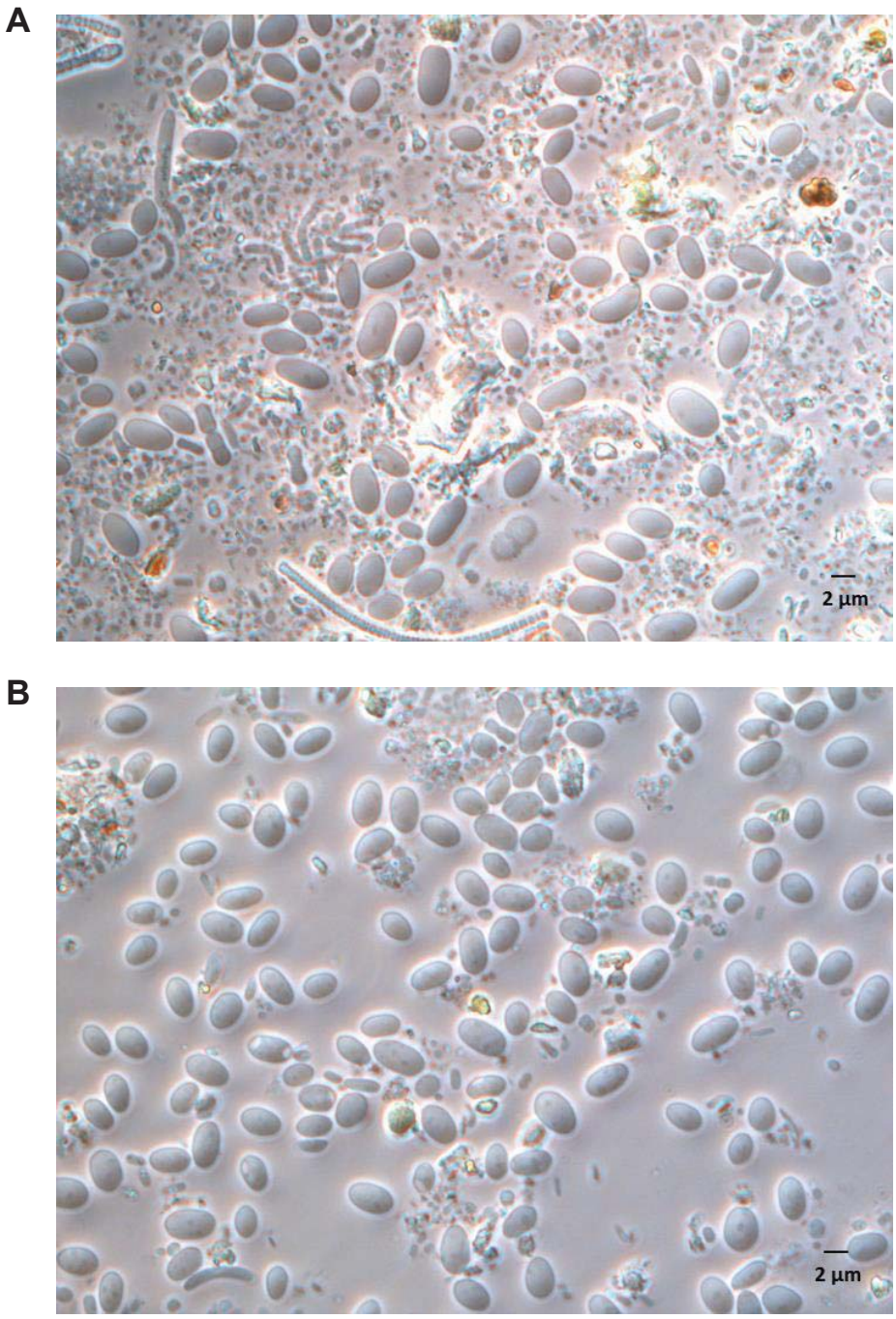


Figure 5.3 Phase contrast images of enriched *Quinella*-like cells from sample 3. A) Supernatant after the first filtration (23-µm pore size) and centrifugation (100 × *g*) steps. B) Image of preparation containing concentrated *Quinella*-like cells after the full protocol was complete. The scale bar indicates a distance of 2 µm.

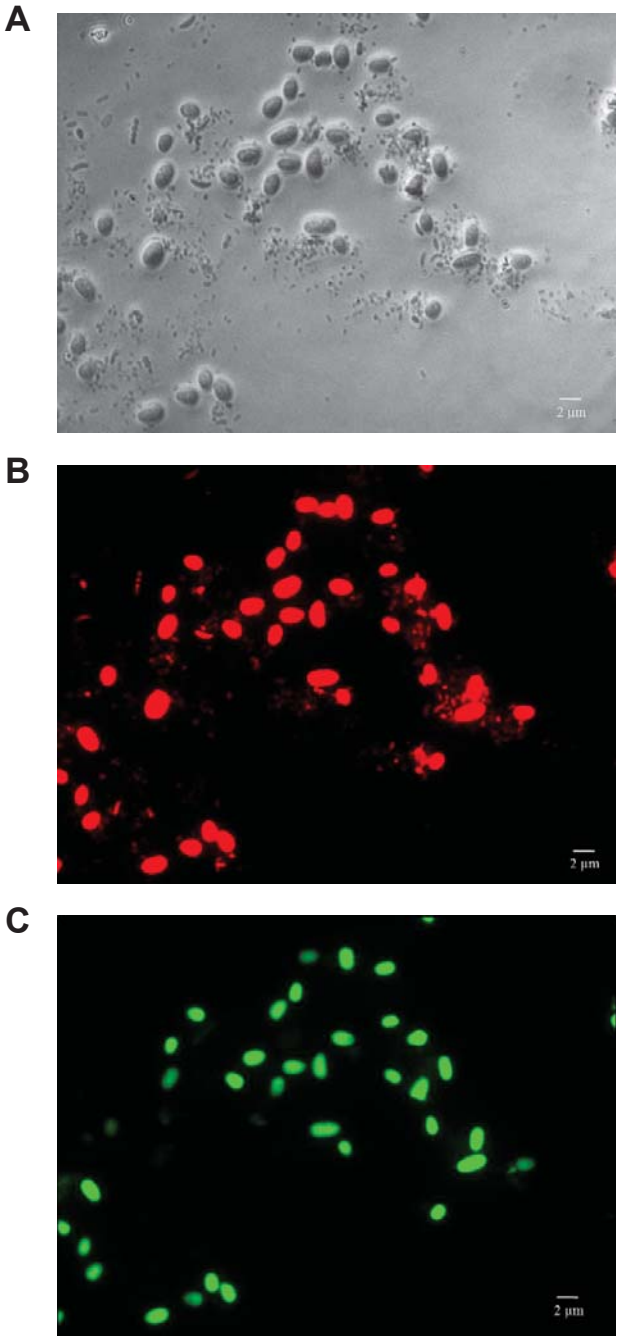


Figure 5.4 Micrographs of concentrated *Quinella* cells from sample 3 that were hybridised simultaneously with Cy3-labelled universal bacterial probe EUB338 and Alexa 488-labelled *Quinella*-specific probe Quin1231. The preparation was enriched for large cells using a sample from the rumen of a sheep with a large population of *Quinella*. A) Phase contrast image of the enriched preparation. B) The same field showing cells that hybridised with the universal bacterial probe. C) The same field showing cells that hybridised with the *Quinella*-specific probe. The scale bar indicates a distance of 2 µm.

5.2.5 Ultrastructure of *Quinella*

Phase contrast microscopy suggested that the enrichment method worked well, as the proportion of large oval cells was increased (Figure 5.3). These samples were confirmed to be dominated by *Quinella* cells based on FISH and clone library analysis. Sample 3 was used for scanning electron microscopy (SEM) and transmission electron microscopy (TEM) to study the cell size and structure more closely. The SEM and TEM images also suggested that the enrichment method worked because a large part of the cells observed were the large oval type postulated to be *Quinella*. These had a range of sizes and outer cell surface textures, which may reflect cells at different stages in the growth cycles, or perhaps different species.

SEM images revealed that the cell size was 3 to 5 µm long and 1 to 2 µm in diameter (Figure 5.5). These are smaller than the cells reported by Orpin (1972), which were 5.8 to 8.0 µm long and 2.5 to 6.0 µm in diameter. It was interesting to find that, unlike *S. ruminantium* (the genus closest to *Quinella* in phylogenetic analyses), *Quinella* in these preparations did not have thick and complex flagella (curled-up and 20 nm in diameter), and it is unclear if these *Quinella* cells possessed flagella. Of course, the flagella may have been sheared off during the cell concentration process, which including filtering and centrifugation steps. *Quinella* has previously observed to be motile but the presence of flagella was not reported (Orpin, 1972). Some *Quinella* cell were fully covered with a granular material while others were either partially covered or had a smooth cell surface (Figure 5.5). In some instances, it appeared that this granular material was peeling off from the cells, which might be an artefact of preparing the cells for SEM. It is unclear if this is the outer membrane, a polysaccharide layer, or some other structure. *Quinella* cells with smooth cell surfaces were similar in shape and size to the cells with granular cell surfaces. In TEM images, two different types of cell surfaces were also observed as well. Some *Quinella* cell surfaces were covered with an uneven electron-dense granular layer (Figure 5.6 A, B), while in others this was absent (Figure 5.6 C, D). Cells of both types had short tuft-like structures, but these were more prevalent on the cells without the electron-dense layer. The surface features observed by SEM and TEM look like fimbrae, but it is not possible at this stage to rule out that they are strands of a surface capsule or similar material.

The cells had a thin cell envelope apparently composed of peptidoglycan sandwiched between inner and outer cell membranes (a characteristic of gram negative bacteria),

confirming that *Quinella* possess gram negative cell wall structure like its cultured relatives in the family *Selenomonadaceae* (Figure 5.6). *Quinella* belongs to the phylum *Firmicutes*, which contains mainly gram positive bacteria, but it is placed in the class *Negativicutes*, which contains bacteria with gram-negative cell wall structures (Marchandin et al., 2010). Unlike most other members of this class it is oval in shape rather than being spherical (*Megasphaera* and *Veillonella*) or curved rods or crescent shaped (*Selenomonas* and *Sporomusa*) (Marchandin et al., 2010). Notably, *Quinella*-like cells showed gram positive type cell division even though they have a gram negative cell wall structure (Figure 5.6A). Condensed material seen inside the cells in the TEM images might be genomic DNA which seems to be dividing between future daughter cells (Figures 5.6A and C).

5.2.6 Diversity of *Quinella* in the concentrated samples

On the basis of clone library analyses, *Quinella* 16S rRNA genes constituted up to 62% of the total in the concentrated samples (Table 5.2). The diversity of these *Quinella* was investigated by importing the 155 cloned sequences derived from the three samples into the refined long-length *Quinella* phylogenetic tree, using the parsimony insertion tool (Section 2.10). Imported cloned sequences were distributed throughout the tree (Figure 5.7), suggesting that the analysed samples contained more than one *Quinella* species. These samples might therefore be suitable to gain a general insight into the metabolism of *Quinella*, by reconstructing genomes from DNA extracted from these samples. However, the diversity suggested that generation of genomes from single strains should not be expected.

5.2.7 Single cell sorting to generate genomic DNA sequences from *Quinella*

In an attempt to generate genome sequences from more defined staring material and so avoid producing chimeric genomes, single cell sorting was performed on sample 1. Multiple subsamples each of 50 putative *Quinella* cells were collected, guided by the fluorescent signal from labelling with the Quin1231 probe and by cell size (section 2.11.4 and 2.14). DNA was extracted from 10 of these subsamples, and multiple displacement amplification (MDA) was used to amplify genomic DNA. DNA from one subsample was successfully amplified, and this was used for DNA sequencing. This sample is referred to as sample 4.

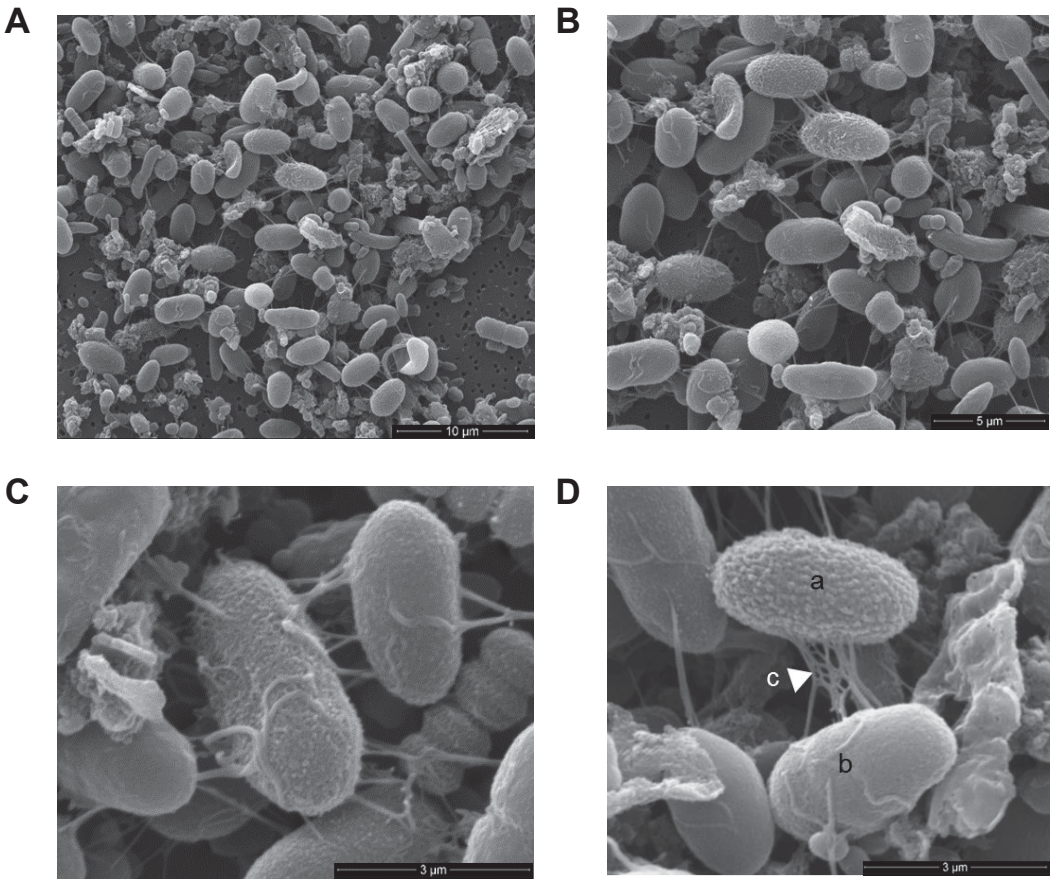


Figure 5.5 Scanning electron microscopic images of putative *Quinella* cells and other cells in sample 3. Features: a, granular cell type; b, smooth cell type; c, unidentified material peeling off from cell surface, possibly a dehydrated glycocalyx or denatured surface associated proteins.

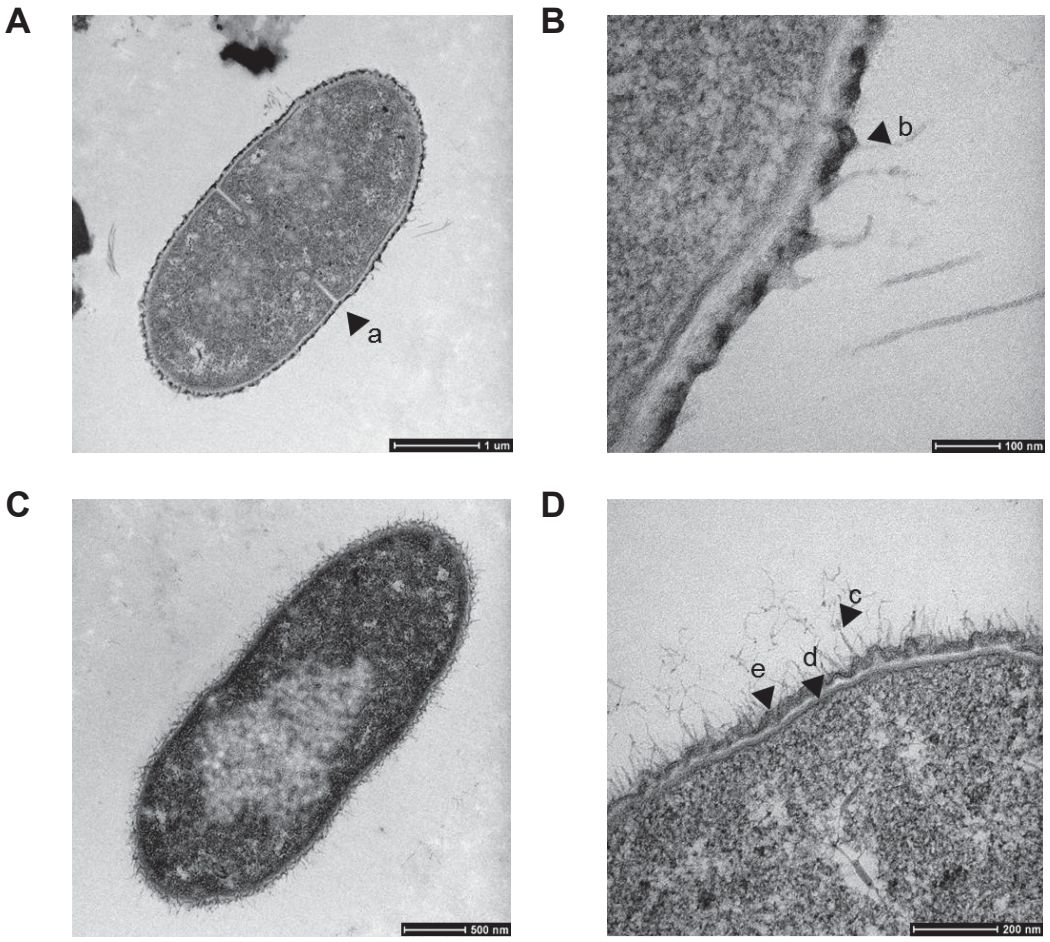
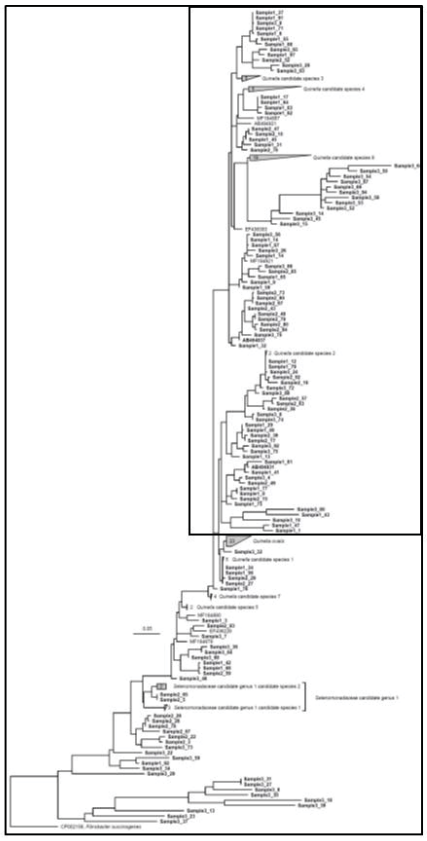


Figure 5.6 Transmission electron microscopic images of putative *Quinella* cells from sample 3. The scale bars indicate different distances for each panel, as noted. Features: a, division site resembling those found in Gram-positive bacteria (Adams and Errington, 2009); b, electron dense material on outer membrane; c, possible fimbriae or outer surface material; d, cytoplasmic membrane; e, outer membrane.




Sample1_37
Sample1_91
Sample3_9
Sample1_71
Sample1_6
Sample1_55
Sample1_66
Sample3_93
Sample1_97
Sample2_52
Sample3_28
Sample3_83
5 Quinella candidate species 3
5 Quinella candidate species 4
Sample1_17
Sample1_64
Sample1_63
Sample1_62
MF184887
AB494921
Sample2_47
Sample2_10
Sample1_45
Sample1_31
Sample2_70
19 Quinella candidate species 6
Sample3_63
Sample3_50
Sample3_54
Sample3_87
Sample3_69
Sample3_94
Sample3_58
Sample3_53
Sample3_52
Sample3_14
Sample3_45
Sample3_15
EF436383
Sample3_56
Sample1_14
Sample1_67
Sample3_26
Sample1_14
MF184921
Sample3_66
Sample2_85
Sample1_65
Sample1_9
Sample1_59
Sample2_73
Sample2_90
Sample2_67
Sample2_43
Sample2_48
Sample2_79
Sample2_80
Sample2_84
Sample3_70
AB494837
Sample1_32
2 Quinella candidate species 2
Sample1_12
Sample1_79
Sample3_24
Sample2_92
Sample2_18
Sample3_72
Sample3_88
Sample2_57
Sample2_63
Sample2_36
Sample3_6
Sample3_74
Sample1_29
Sample1_49
Sample2_39
Sample2_17
Sample3_92
Sample3_75
Sample1_13
Sample1_81
AB494931
Sample1_41
Sample3_4
Sample2_49
Sample1_77
Sample1_8
Sample2_15
Sample1_75
Sample3_80
Sample1_43
Sample3_10
Sample1_47
Sample1_1

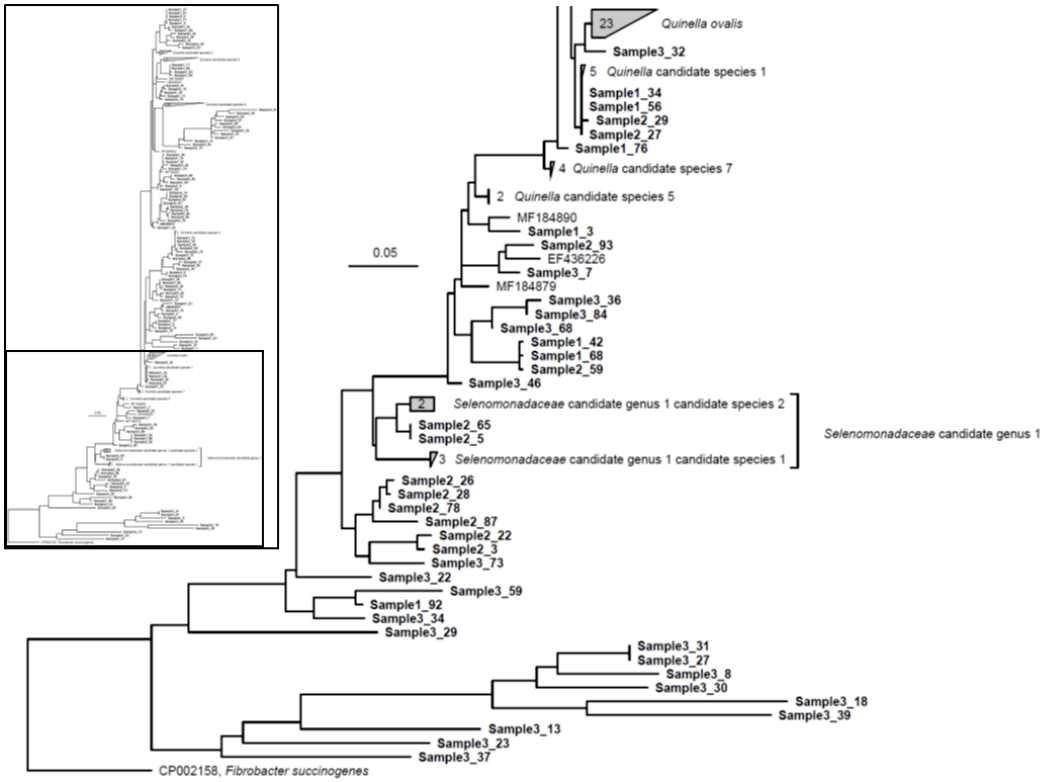


Figure 5.7 Phylogenetic tree of 16S rRNA gene sequences affiliated with *Quinella* from concentrated samples enriched for *Quinella* cells. The tree is split into two parts (see an overview of the full tree on the top left of each panel). Sequence identifiers in bold letters represent 16S rRNA gene sequences from the different samples of concentrated *Quinella* cells. Numbers inside the collapsed clusters represent the total number of sequences in a particular cluster. The 16S rRNA gene sequence of *Fibrobacter succinogenes* (FibSuc43, GenBank accession CP002158) was used as out group sequence. The scale bar represents 0.05 changes per nucleotide.

5.2.8 Shotgun sequencing results, assembly and contigs binning

The three samples with concentrated *Quinella* cells (referred to as samples 1, 2, & 3, section 5.2.4) were selected for metagenome construction. DNA was extracted from these enriched samples and this was used as the template for metagenome sequencing using the Illumina MiSeq sequencing platform. Genomic shotgun libraries of insert size of 550 bp were built with the aim of assembling full or partial *Quinella* genomes. The DNA reads were assembled and assigned to bins as described in sections 2.16.

Table 5.4 details the sequencing statistics and Table 5.5 lists the results from the assembly (using SPAdes assembler) and binning (using MetaBAT binning programme) steps. Samples 1 to 3 generated between 21 and 33 bins each. Bins with size >1 Mbp were selected for further analysis. 16S rRNA genes amplified from the DNA extracted from one pool of 50 cells and cloned in *E. coli* were identified as originating from *Quinella* (20 of 20 clones). This DNA (sample 4) was used to amplify more DNA for sequencing by MDA, and the MDA products sequenced using Illumina MiSeq. However, the reads from this sample did not assemble well, as these data appeared to be dominated by chimeric sequences from short DNA fragments that were amplified during the MDA steps. This was reflected in the assembly, where average contig lengths were <1.5 kb. This sample was excluded from further downstream analysis.

Genomic phylotyping of the bins (for method see section 2.17.1) showed that reads from members of the family *Selenomonadaceae* were the most abundant (on the basis of homology to marker genes) (Figure 5.8). This indicated that the cell concentration method was at least able to concentrate bacterial cells belonging to the family *Selenomonadaceae*. Three bins (bin8 and bin19 from sample 1 and bin4 from sample 3) were not able to be assigned to any taxonomic identity using bacterial and archaeal marker genes from Amphora2 (Wu and Scott, 2012; Kerepesi et al., 2014). A major portion of bin11 and bin17 of sample 2 were represented by the family *Synergistaceae* whereas bin9 of sample 3 was characterised by the family *Prevotellaceae*. Additionally, as expected, some other common rumen bacterial families (Henderson et al., 2015) were also detected in genomic bins, for example, *Spirochaetaceae*, *Eubacteriaceae*, *Lachnospiraceae*, and *Clostridiaceae*. Bin28 of sample 2 was classified as a mixture of the families *Methanobacteriaceae* and *Thermodesulphobiaceae*. Genomic phylotyping can also be done at a genus and species level, but because this analysis is based on published

genomes, this can lead to false identification of genomic bins. Therefore, a more robust analysis was needed to identify the phylogenetic status of *Quinella* genomic bins.

Table 5.4 Read number and sequence quality from DNA sequencing data.

Sample	Description	Reads (bp)	Quality score (%Q30)[a]
1	Pooled from 12 sheep	6,997,737	84.2
2	Pooled from one sheep from two measuring rounds	5,688,064	85.0
3	Single rumen sample from one sheep	4,541,407	84.1
4	MDA amplified DNA from 50 sorted *Quinella* cells using sample 1	7,369,816	84.2

[a]%Q30, represents the percentage of bases with a quality score of at least 30 (i.e., an inferred base call accuracy of 99.9%).

Table 5.5 Assemblies and bins generated from metagenomic DNA sequence data.

Assembly details	Sample 1	Sample 2	Sample 3
Numbers of reads	6,997,737	5,688,064	4,541,407
Numbers of contigs	290,401	277,479	416,020
Numbers of contigs >1000 bp	69,711	67,624	87,783
Largest contigs (bp)	269,413	637,526	228,730
N50[a]	1,029	1,053	904
N75[a]	664	673	614
GC content (mol%)	47.26	44.42	44.97
Number of bins generated[b]	31	33	21
Number of bins >1 Mbp	12	13	9

[a]N50 and N75 are the contig lengths (bp) such that longer or equal length contigs account for 50% or 75% of the bases in the assembly (Gurevich et al., 2013).
[b]Using MetaBAT

5.2.9 *Quinella* genome bins

Table 5.6 shows steps taken to identify the bins to be used for analyses of the metabolism of *Quinella* based on it genome. A total of 13 bins were identified as containing possible genomic DNA from *Quinella* (five from sample 1, and four each from samples 2 and 3). This was based on the presence of 16S rRNA gene sequences of *Quinella*. Following the proposed genome quality classification scheme of Parks et al. (2015), four bins were considered nearly (≥90%) complete with medium (5% to 10%) to low (≤5%) contamination (Table 5.8). These were selected for further study: bin 5 from sample 1 (from now on this will be called bin SR1Q5), bin 7 from sample 1 (SR1Q7), bin 5 from sample 2 (SR2Q5) and bin 1 from sample 3 (SR3Q1). Among other *Quinella* bins, bin 2 from sample 1, bin 18 from sample 2 and bin 5 from sample 3 were 100% complete but at the same time they were highly contaminated (≥48%). Post-binning strategies (Kang et al., 2015) were applied to reduce contamination from these bins, but these proved to be unsuccessful and so these bins were not used.

Quality assessment of the *Quinella* genome bins using CheckM (section 2.17.2) assigned three (SR1Q5, SR1Q7 and SR2Q5) of the final four *Quinella* genome bins to the phylum *Firmicutes* based on 295 lineage-specific marker genes (present in >97% of particular lineage) arranged in 158 marker sets. The fourth genome bin (SR3Q1) was only assigned to the domain Bacteria (it contained 104 lineage-specific marker genes to separate it from other domains). Markers genes present in more than one copy are indicative of contamination while the absence of these genes contributes towards incompleteness (Parks et al., 2015). Contamination in the *Quinella* genome bins ranged from estimates of 0.21 to 10.3% (Table 5.7 and Table 5.8). Of 295 *Firmicutes* lineage-specific marker genes, 18 and 33 were found in duplicate copies in bins SR1Q5 and SR1Q7 respectively. Bin SR3Q1 contained 10 genes in duplicate, six in three and four copies each, while one gene was present in five or more copies. As a result, this bin was estimated to be 10.25% contaminated. Because only one duplicate marker gene was found in bin SR2Q5, this bin was considered the least contaminated (0.21%) *Quinella* genome bin. Furthermore, of 295 lineage-specific marker genes, 30, 18 and 17 were absent from genome bins SR1Q5, SR1Q7 and SR2Q5 respectively, indicative of genome incompleteness. Notably, in genome bin SR3Q1, out of 104 bacteria domain level marker genes, six were absent, suggesting that this bin should be treated with care when predicting the overall *Quinella* physiology. The contamination level could have been further reduced by excluding

contigs with duplicated marker genes, but it was unclear which of the duplicates originated from *Quinella*, as they appear to have similar coverage, G+C contents (Figure 5.9) and gene arrangements.

The contigs with very different G+C content and coverage were investigated in more detail, but none of these contigs appeared to have multiple copy marker genes and hence had no role in bin contamination and completeness values. It was interesting to find no strain heterogeneity in bin SR2Q5, suggesting that it represented a single strain. In contrast, the other three bins (SR1Q5, SR1Q7 and SR3Q1) were highly mixed (≥48.5%) and likely contained multi-strain genomic information.

Table 5.6 Steps to generate *Quinella* genome bins from metagenomic DNA sequence data. The bins generated in step 1 (Table 5.5) were retained or removed using a series of sequential analyses.

Steps	Description/filter	Total bins left	Refer to section
1	Binning with MetaBAT	85	5.2.7
2	Bins with >1,000,000 bp size	34	5.2.7
3	Bins containing 16S rRNA gene sequences retained	23	5.2.7
4	Bins originating from members of the family *Selenomonadaceae* retained (>70 % of the bins residues identified as belonging to *Selenomonadaceae*, using Amphora markers)	20	5.2.7
5	Bins containing *Quinella* 16S rRNA genes retained (BLAST and phylogenetic tree construction)	13	5.2.8
6	Bin with >90% completeness and <11% contamination retained	4	5.2.8
7	Bins with amplified full length 16S rRNA gene sequences identified as *Quinella* retained (BLAST and phylogenetic tree construction)	4	5.2.9

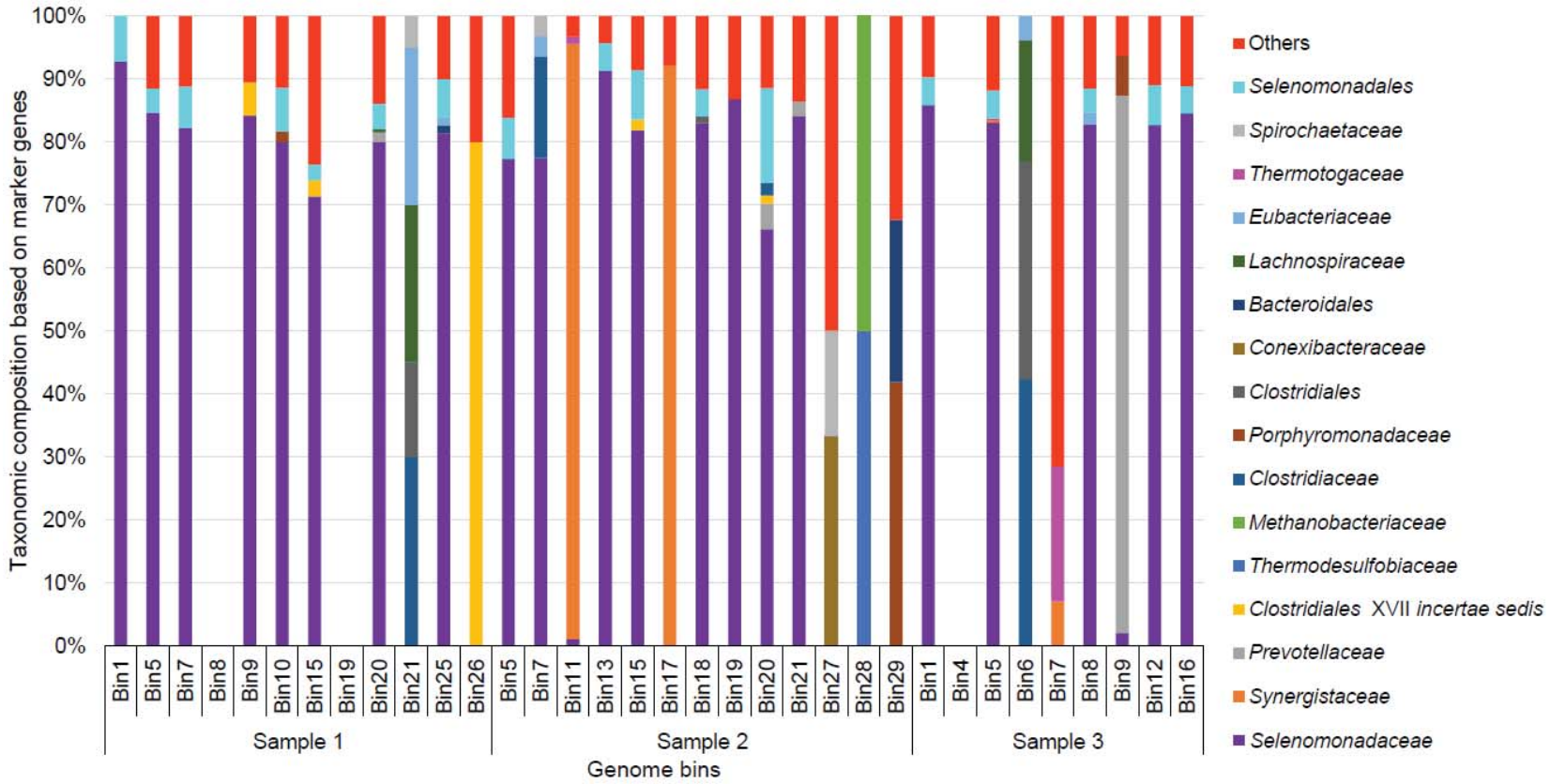


Figure 5.8 Phylotyping of genomic bins. Taxonomic composition of genomic bins was assessed using the AmphoraNet web-based server which uses 31 bacterial and 104 archaeal protein coding marker genes for genomic phylotyping (Kerepesi et al., 2014). Bin8 and Bin19 from sample 1 and Bin4 from sample 3 were not able to be assigned any taxonomic identities.

The 16S rRNA gene sequences in each of these four bins were identified as *Quinella* sequences by BLAST analyses against a curated database of rumen bacterial 16S rRNA genes (Henderson et al., 2017). This database contained sequences from the improved *Quinella* phylogeny (chapter 3). Genome bin size varied from 1.8 to 2.6 Mbp, suggesting that these four genome bins might contain different *Quinella* species. However, the differences in genome bins size may also reflect differences in the level of contamination. For example, genome bin SR3Q1 (2.6 Mbp) was the largest bin, but was estimated to contain 10.2 % contamination while bin SR2Q5 which contained almost no contamination was the smallest (1.8 Mbp). Genome sizes of other members of the family *Selenomonadaceae* (to which *Quinella* belongs) vary from 1.26 Mbp (*Dialister micraerophilus*) to 5.28 Mbp (*Pelosinus fermentans*) (Yutin and Galperin, 2013). Significant variations in genome sizes have been observed previously within a genus. For example, in the genus *Selenomonas*, which is the closest genus to *Quinella* has 10 species currently identified (Yarza et al., 2008) with , genome sizes varying from 2.3 Mbp (*Selenomonas* sp. F0592; GenBank accession NZ_CP012071) to 3.0 Mbp (*Selenomonas ruminantium* subsp. *lactilytica*; (Kaneko et al., 2015). The variation in the *Quinella* genome bin sizes is therefore consistent with what is observed for genomes from the genus *Selenomonas.* Consequently, the number of contigs also differed among the *Quinella* genome bins (Table 5.8). The G+C contents of these bins were also varied. The lowest G+C content was found in SR3Q1 (49.2 mol%) whereas the highest was observed in SR2Q5 (56.0 mol%). Again, this pattern was similar to that found in the genus *Selenomonas*, where G+C content varied from 50.7 mol% (*Selenomonas ruminantium* subsp. *lactilytica*; (Kaneko et al., 2015) to 57.1 mol% (*Selenomonas* sp. F0592; GenBank accession NZ_CP012071). Therefore, the variation in the G+C content of genomic DNA may not be unusual. However, genome completeness, contamination and strain heterogeneity may also play a major role in the observed G+C content. Missing regions or contaminated regions as well as mixed genomes at strain level might also affect the overall G+C content of genome bins.

In summary, four potential *Quinella* genomic bins were generated that showed different levels of completeness, contamination, G+C content and genome bin size. This finding suggested that these genome bins may represent different *Quinella* species. This is subjected to further analysis using 16S rRNA gene sequences to understand their phylogenetic diversity and functional genome based analysis to understand genome level diversity.

Table 5.7 Lineage-specific quality control assessment of *Quinella* genome bins using CheckM.

Characteristic	*Quinella* genome bins												
	1Q2	SR1Q7[a]	SR1Q5[a]	1Q1	1Q9	2Q18	2Q15	SR2Q5[a]	2Q13	SR3Q1[a]	3Q5	SR3Q16	SR3Q12
No. of genomes used[b]	5449	100	100	100	100	5656	5449	100	100	5449	5656	5449	100
No. of markers used	104	295	295	295	295	56	104	295	295	104	56	104	295
Lineage-specific marker sets	58	158	158	158	158	24	58	158	158	58	24	58	158
No. of markers absent	0	18	30	87	191	0	3	17	146	6	0	18	80
Single copy markers	3	244	247	198	90	0	43	277	117	75	0	43	112
Duplicate markers	7	33	18	9	14	8	32	1	31	10	0	24	51
Triplicate markers	17	0	0	1	0	15	20	0	1	6	0	15	39
Four copies of markers	25	0	0	0	0	14	6	0	0	6	0	3	12
Markers present in 5+ copies	52	0	0	0	0	19	0	0	0	1	56	1	1
Genome completeness	100	94.2	91.0	76.0	28.8	100	96.6	92.7	43.9	91.4	100	78.6	76.1
Genome contamination	48.5	8.7	5.7	4.8	3.6	329.2	68.0	0.2	10.7	10.3	672.8	65.1	65.6
Strain heterogeneity	57.0	48.5	77.8	41.7	57.1	69.4	71.1	0.0	79.4	78.4	76.6	48.5	29.6
16S rRNA gene length (bp)	793[c]	426	122	966	444	915[c]	134[c]	399	136	373	976[c]	124	124

[a]Genome bins coloured in green were the shortlisted *Quinella* genome bins used for further analysis.

[b]Number of reference genomes used to identify lineages (root, 5656; Bacteria, 5449; *Firmicutes*, 100).

[c]Multiple 16S rRNA sequences were found, but only the length of sequences with an e-value of 0 when compared to known *Quinella* 16S rRNA genes are listed.

Table 5.8 *Quinella* genome bin specifications.

Genome bin specification	*Quinella* genome bins			
	SR1Q5	SR1Q7	SR2Q5	SR3Q1
Genome bin size (bp)	2,125,473	2,584,672	1,821,931	2,614,227
G+C content (mol%)	49.0	52.9	56.0	49.1
Number of contigs	132	169	42	68
Length of largest contigs (bp)	62,332	68,294	179,392	219,728
Genome completeness (%)	90.9	94.2	92.6	91.4
Genome contamination (%)	5.7	8.6	0.2	10.3
Strain heterogeneity (%)	48.5	77.8	0	78.4

5.2.9 Confirmation of taxonomic identification of bins

The partial 16S rRNA sequences that were extracted from the four *Quinella* genome bins (SR1Q5, SR1Q7, SR2Q5 and SR3Q1) were compared against a curated bacterial 16S rRNA gene database (Henderson et al., 2017). All four 16S rRNA sequences showed best matches with 16S rRNA genes from *Quinella* spp., with best e-values shown by SR1Q7 and SR2Q5 (0) while SR1Q5 and SR3Q1 had e-values of 6.00E-61 and 4.00E-46, respectively.

To obtain full-length 16S rRNA gene sequences, the genome bin sequence information was used to design primers (Figure 5.10A), which were then used to generate clone libraries containing almost full-length16S rRNA gene sequences from the DNA extracted from samples 1 to 3 (section 2.7.2). The DNA fragments amplified from DNA extracted from samples 1 to 3 were of the expected sizes (1851 to 2390 bp long) (Figure 5.8B). Two libraries were prepared from sample 1, targeting the 16S rRNA gene sequences found in bins SR1Q5 and SR1Q7 by using two different primers sets (Figure 5.10A). Different primer sets were used to target 16S rRNA genes in bin SR2Q5 and SR3Q1 from samples 2 and 3, respectively (Figure 5.10A). This therefore produced one library targeting the 16S rRNA gene and flanking regions for each of the four genome bins. Cloned amplicons were sequenced using multiple primers (section 2.17.3).

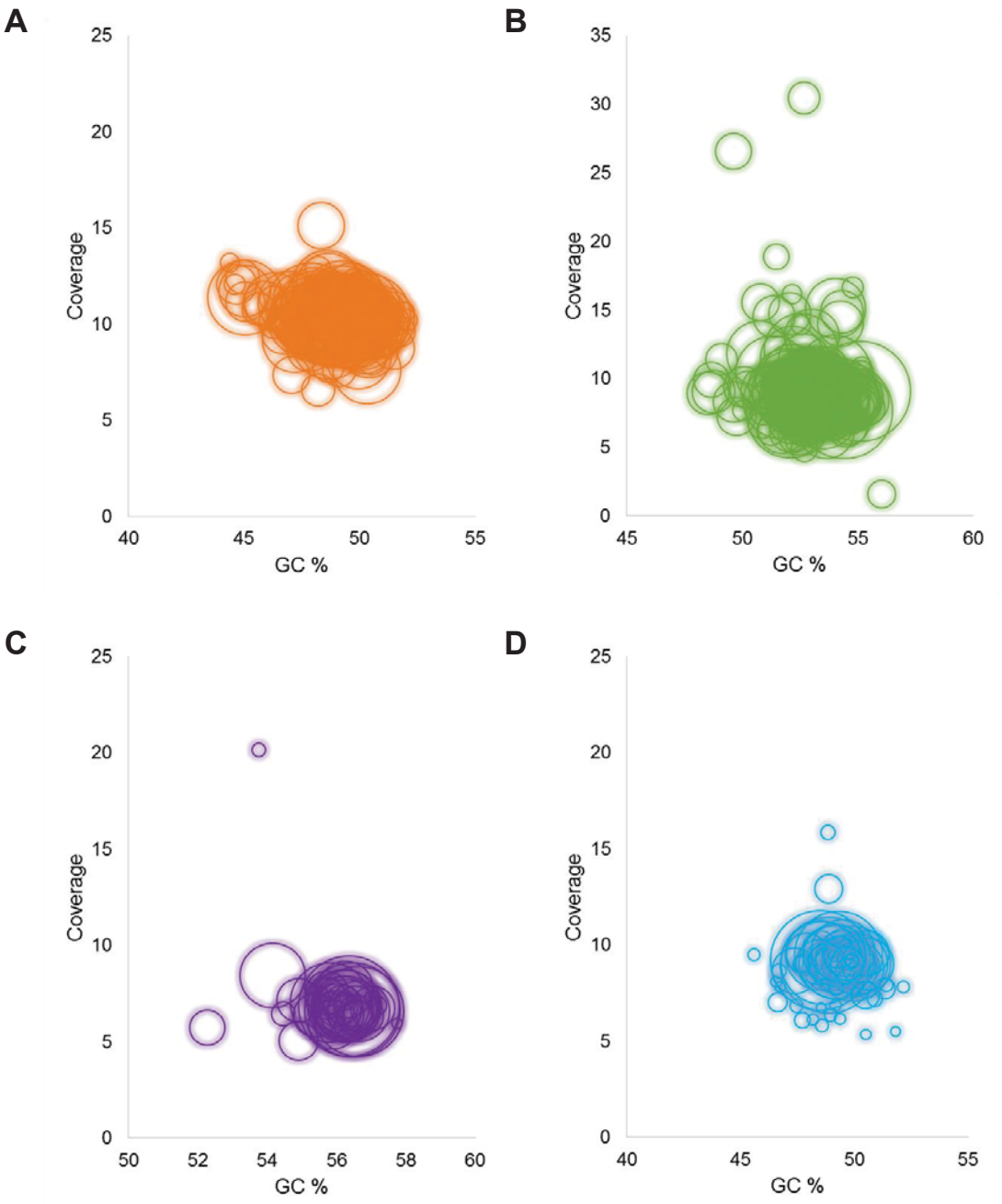


Figure 5.9 Comparison of *Quinella* genome bins (A) SR1Q5 (B) SR1Q7 (C) SR2Q5 (D) SR3Q1. G+C content (mol%; x-axis), coverage (y-axis), and relative contig lengths (size of the circles) are shown. The coverage is the k-mer coverage (read k-mers per contig k-mer) for the largest k value calculated by SPAdes, which was 127 in this case.

The best five clones from each of the four clone libraries were selected based on the electropherogram quality, and on matches to the flanking regions and partial 16S rRNA genes in the bin that was being targeted, and these were used for the phylogenetic analysis.

The cloned sequences amplified using the same forward primer were highly similar but not identical over the full length of the cloned fragments (Table 5.9, upper triangle of matrix). The greatest similarity was found in between clones targeting bin SR1Q5 (≥98.9%) followed by clones of bin 2Q5 (≥98.2%) and then clones from bin SR3Q1 (≥98.4%) whereas the lowest level of similarity occurred for clones targeting bin SR1Q7 (≥91.0%). The sequence similarities between only the 16S rRNA genes in clones amplified using the same forward primer was >99% (Table 5.9, lower triangle of matrix). The similarities between cloned sequences amplified using different forward primers was lower (≤82.6% over the full cloned sequences, ≤95.7% for the 16S rRNA gene regions of the clones). This indicates that the 16S rRNA genes amplified using the different primers were from potentially different *Quinella* species. The variations within any one clone library could have been due to amplification of DNA from closely-related strains within the same DNA sample.

Next, the clones were compared with the contigs in the four genome bins that contained partial 16S rRNA genes. When the clone regions overlapping with the reference contigs sequences from the four bins were aligned, all four clone libraries contained at least one clone that showed ≥99.6% similarity over the partial 16S rRNA gene and the adjacent gene region targeted by the primer set. These best matches were for all five clones targeting bin SR1Q5 (100%), clone SR1Q7_49 (99.1%) targeting bin SR1Q7, clone SR2Q5_87 (99.8%) targeting bin SR2Q5 and clone SR3Q1_120 (99.9%) targeting bin SR3Q1. When this analysis was restricted to just the 16S rRNA regions in the 20 clones, all clones had ≥99.0% similarity (SR1Q5, 100%; SR1Q7, ≥99.0%; SR2Q5, ≥99.2%; and SR3Q1, ≥99.2%) and at least one clone in each library had 99.8% to 100% similarity to the partial 16S rRNA genes in the bins that were targeted by that primer set (Table 5.9). This suggests that some of the nearly full length 16S rRNA genes in the clone libraries were from the same strains or very close relatives of the strains that contributed the contigs containing the 16S rRNA genes in each of the four bins.

The analysis was repeated limiting it to only the parts of the cloned sequences that matched the bin contig sequences outside the 16S rRNA gene. Clones assigned to bin SR1Q5 all had 100% similarity with the references contig. In the other clones, the

similarities were ≥85.2% with the respective contigs from the bins they were assigned to. This showed that the differences between the clones and the contigs from the bins they were targeting were mainly in these regions adjacent to the 16S rRNA genes. Translating these regions showed that this also reflected amino acid level differences, suggesting that the clones might represent different strains of same *Quinella* species. Overall, it seems that the cloned 16S rRNA genes are representative of the bins they were assigned to. This variations in the flanking regions of the cloned sequences showed that each of samples 1 to 3 contained different but closely-related strains that shared very similar 16S rRNA genes. However, this doesn't allow conclusions to be made about the sequence data in the bins. If the starting samples contained very closely-related strains, it is likely that the bins represent aggregates of genomes from closely related strains.

The 16S rRNA gene sequences were extracted from the 20 clones and used to construct a phylogenetic tree. The cloned sequences representing different *Quinella* genome bins grouped in different parts of the tree (Figure 5.11), as expected if the *Quinella* genome bins were from separate *Quinella* species. All five 16S rRNA clone sequences assigned to SR1Q5 grouped together and branched with *Quinella ovalis* with >97% similarity with other sequences in that cluster (excluding sequences Unl25493; >96%). Genome bin SR1Q5 may therefore contain genomic sequences that represent strains of the species *Quinella ovalis* or a very close relative of it. Based on the placement of the other 16S rRNA genes, it appears that the other three genomic bins represent three different species of the genus *Quinella* (Figures 5.11, 5.12 and 5.13). Furthermore, the 16S rRNA gene sequences assigned to the four genome bins fell in different part of the *Quinella* tree and covered most the known diversity of *Quinella*, which suggested that the genomes bins should be useful to gain an insight into the physiology of members of this genus.

Repset sequences from the 454 sequencing study (see chapter 3) representing the OTUs that contained ≥100 sequencing reads were added to this tree using the ARB parsimony (quick add mark) insertion function (Figure 5.13). This was done to see whether any of the 16S rRNA gene sequences assigned to the four genome bins represent abundant OTUs from the original sheep study. Some of the highly abundant OTUs clustered with the 16S rRNA gene sequences representing genomic bins SR1Q5 and SR1Q7, whereas the 16S rRNA gene sequences representing genome bins SR2Q5 and SR3Q1 did not cluster with any repset sequences. This is not surprising, because these genome bins came from

samples from a different group of sheep than the one used to generate the repset sequences (chapter 3).

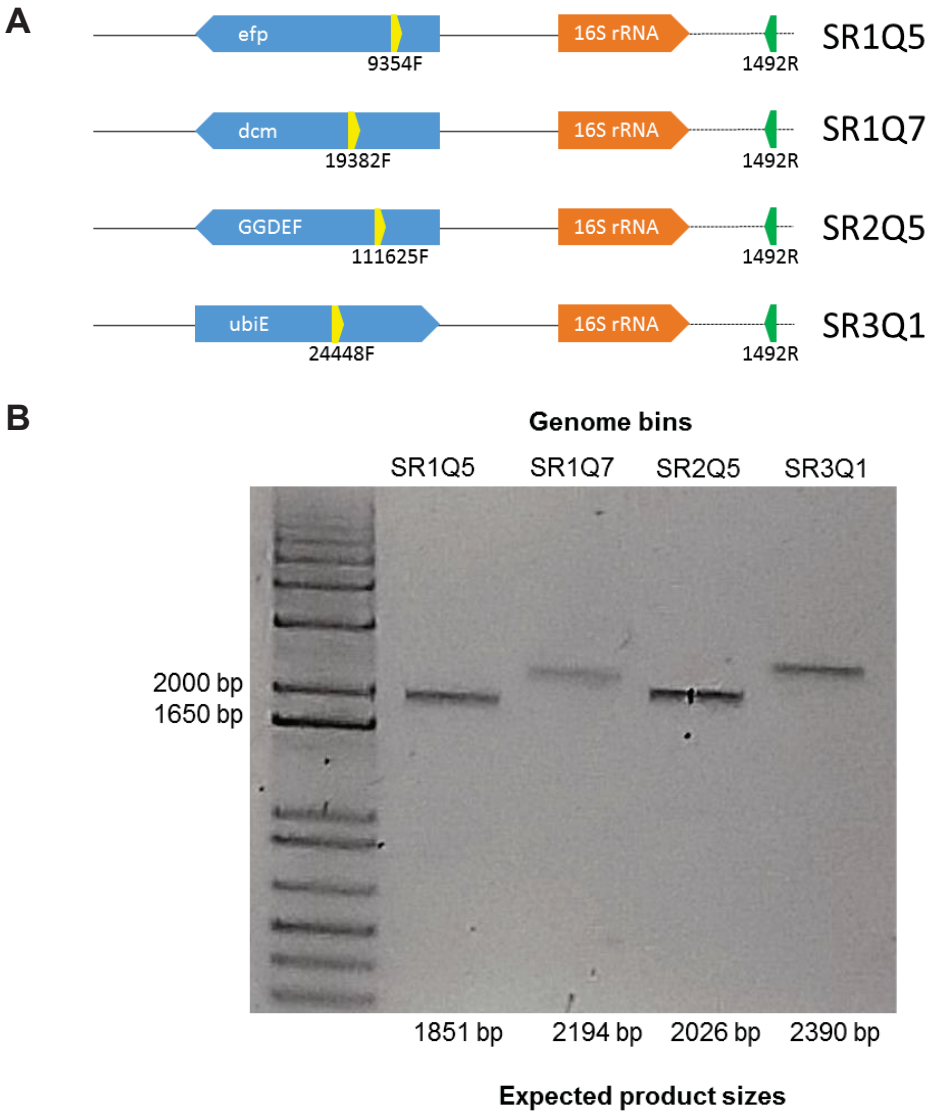


Figure 5.10 Primer targets and use to amplify 16S rRNA genes from DNA of *Quinella*-enriched samples. A) Arrangement of contigs containing 16S rRNA genes and the adjacent genes, showing the regions targeted by the combination of newly-designed forward primers (yellow) and a universal bacterial primer 1492R (green). Primers 9354F and 19382F were used to amplify products from DNA extracted from sample 1, each in combination with the universal bacterial primer 1492R. Primers 111625F and 24448F were used with DNA from samples 2 and 3 respectively, again each in combination with 1492R. B) Agarose gel of selected PCR amplified products showing that the products generated using the different primers were of the sizes expected (Table 2.3). The left-hand lane contains a DNA marker ladder.

Table 5.9 Similarity matrix of cloned sequences. The upper triangle of the matrix gives the similarity (percentage of identical nucleotides) between full lengths of the cloned sequences (colour code white to red, low to high). The lower triangle of the matrix gives similarities between 16S rRNA gene regions of the cloned sequences (colour code white to purple, low to high). The three columns on the right give the similarities of the parts of the cloned sequences with their respective contigs in the genome bins.

		Similarity between cloned sequences (%)																				Similarity with overlapping region of the genome bins (%)		
Genome bins		SR1Q5					SR1Q7					SR2Q5					SR3Q1					With full overlapping region	With 16S rRNA gene only	Other than 16S rRNA gene region
	Clone no.	3	4	9	10	11	38	46	47	49	54	66	67	74	81	87	107	111	116	119	120			
SR1Q5	3		99.6	99.4	99.5	99.1	81.2	81.0	80.7	81.4	82.4	82.6	82.6	82.0	82.4	81.5	82.0	81.7	82.3	82.1	81.9	100.0	100.0	100.0
	4	99.7		99.6	99.5	99.0	81.3	81.1	80.8	81.3	82.3	82.4	82.5	82.0	82.3	81.4	82.1	81.8	82.3	82.1	82.0	100.0	100.0	100.0
	9	99.6	99.6		99.5	98.9	81.2	81.0	80.7	81.3	82.2	82.5	82.6	82.1	82.4	81.5	82.1	81.8	82.3	82.1	82.1	100.0	100.0	100.0
	10	99.8	99.7	99.5		99.0	81.2	81.0	80.6	81.2	82.2	82.5	82.6	82.1	82.4	81.5	82.1	81.8	82.2	82.0	82.0	100.0	100.0	100.0
	11	99.7	99.5	99.4	99.6		81.0	80.8	80.5	81.1	82.0	82.2	82.3	82.0	82.1	81.7	81.8	81.5	82.1	81.9	81.8	100.0	100.0	100.0
SR1Q7	38	94.4	94.2	94.1	94.3	94.1		94.6	93.7	99.6	96.8	79.8	80.1	80.0	81.1	79.5	75.3	75.1	74.2	75.3	75.3	99.0	99.5	98.6
	46	94.3	94.1	93.9	94.1	94.0	99.0		97.0	94.6	91.9	79.6	79.8	79.7	80.8	79.2	76.0	75.8	74.9	76.0	76.1	92.7	99.3	88.9
	47	94.4	94.3	94.1	94.3	94.1	99.4	99.0		93.6	91.0	79.2	79.4	79.4	80.4	78.9	75.9	75.6	74.8	75.9	76.0	90.5	99.8	85.2
	49	94.5	94.3	94.1	94.3	94.2	99.9	99.0	99.4		96.9	80.0	80.1	80.2	81.2	79.5	75.2	74.9	74.1	75.2	75.2	99.1	99.8	98.9
	54	94.4	94.2	94.1	94.3	94.1	99.9	99.0	99.4	99.9		81.0	81.0	80.9	81.2	80.1	74.1	73.7	74.0	74.1	74.0	99.0	99.8	98.6
SR2Q5	66	94.6	94.4	94.3	94.5	94.4	94.6	94.7	94.6	94.6	94.6		99.6	98.7	99.5	98.2	79.1	78.5	79.1	79.1	79.0	99.1	99.4	98.9
	67	94.5	94.3	94.2	94.4	94.3	94.6	94.6	94.6	94.6	94.6	99.8		98.9	99.6	98.2	79.1	78.7	79.2	79.1	79.1	99.2	99.2	99.1
	74	94.5	94.4	94.3	94.3	94.3	94.5	94.5	94.5	94.5	94.5	99.8	99.6		99.0	98.5	79.0	78.5	79.4	79.1	79.0	99.1	99.5	98.7
	81	94.5	94.3	94.2	94.4	94.3	94.5	94.5	94.5	94.5	94.5	99.8	99.7	99.7		98.6	80.1	79.7	79.7	80.0	80.1	98.9	99.2	98.6
	87	94.5	94.3	94.2	94.4	94.4	94.5	94.5	94.5	94.5	94.5	99.8	99.7	99.9	99.7		78.3	77.8	78.8	78.3	78.3	99.8	99.5	100.0
SR3Q1	107	94.2	94.0	93.9	94.1	93.9	95.5	95.4	95.4	95.6	95.5	94.3	94.3	94.1	94.2	94.2		98.9	99.0	99.5	99.5	99.7	99.5	99.8
	111	93.8	93.6	93.4	93.6	93.5	95.0	94.9	94.9	95.0	95.0	93.8	93.8	93.6	93.7	93.7	99.2		98.4	98.9	98.9	99.8	99.7	99.8
	116	93.6	93.5	93.4	93.5	93.4	94.9	94.9	94.9	95.0	94.9	93.7	93.7	93.7	93.6	93.7	98.9	98.4		99.1	99.1	99.8	99.7	99.8
	119	94.3	94.1	94.0	94.2	94.1	95.6	95.6	95.6	95.7	95.6	94.4	94.4	94.3	94.3	94.3	99.6	99.0	99.0		99.6	99.8	100.0	99.7
	120	94.3	94.1	94.0	94.2	94.1	95.6	95.6	95.6	95.7	95.6	94.4	94.4	94.3	94.3	94.3	99.6	99.0	99.0	99.7		99.9	99.7	100.0

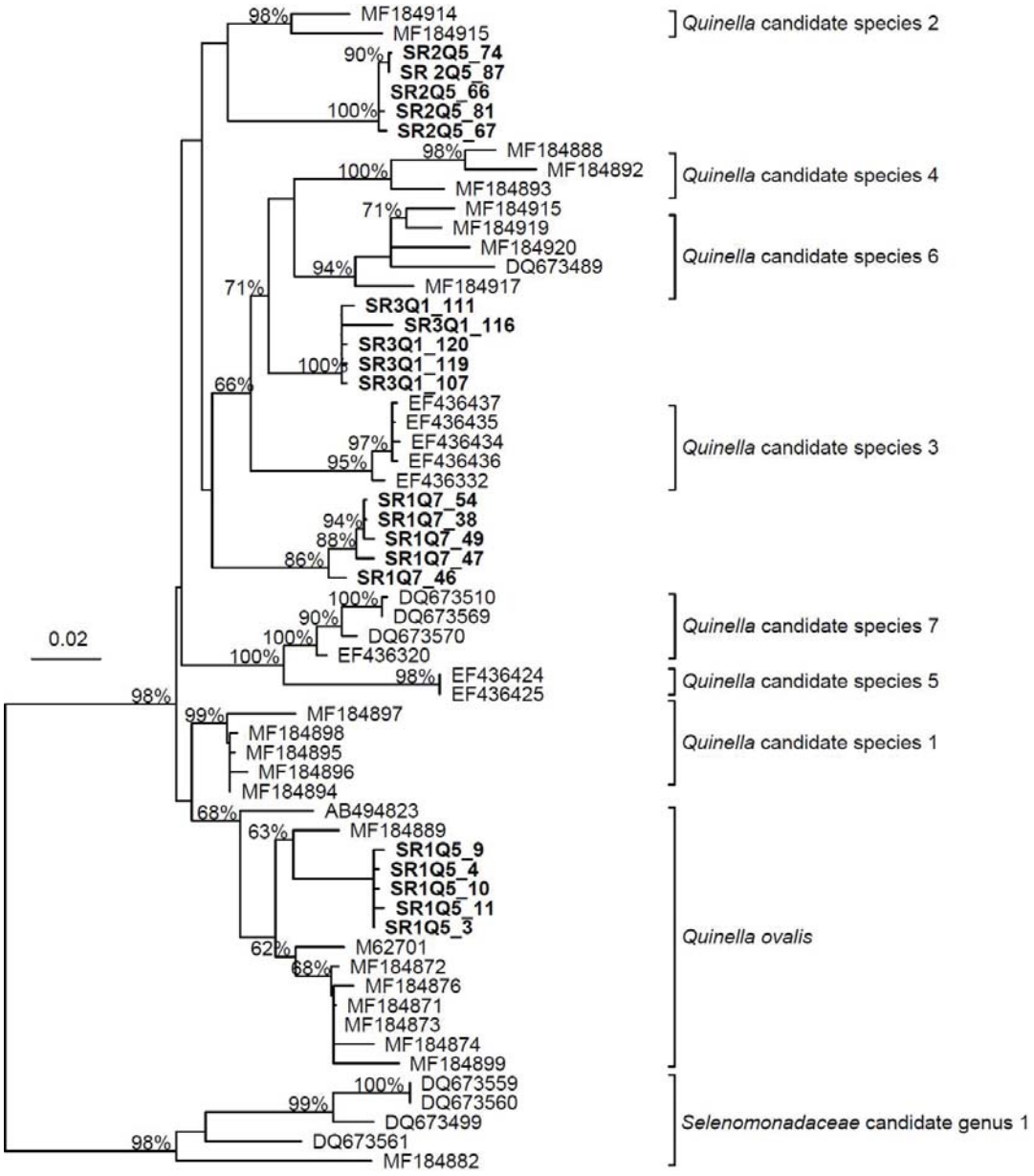


Figure 5.11 Phylogenetic tree of *Quinella* 16S rRNA gene sequences, including those amplified from the DNA samples used to generate the *Quinella* genome bins (in bold font). The scale bar represents 0.02 changes per nucleotide. Bootstrap values of <60% and some of those within defined clusters are not shown.

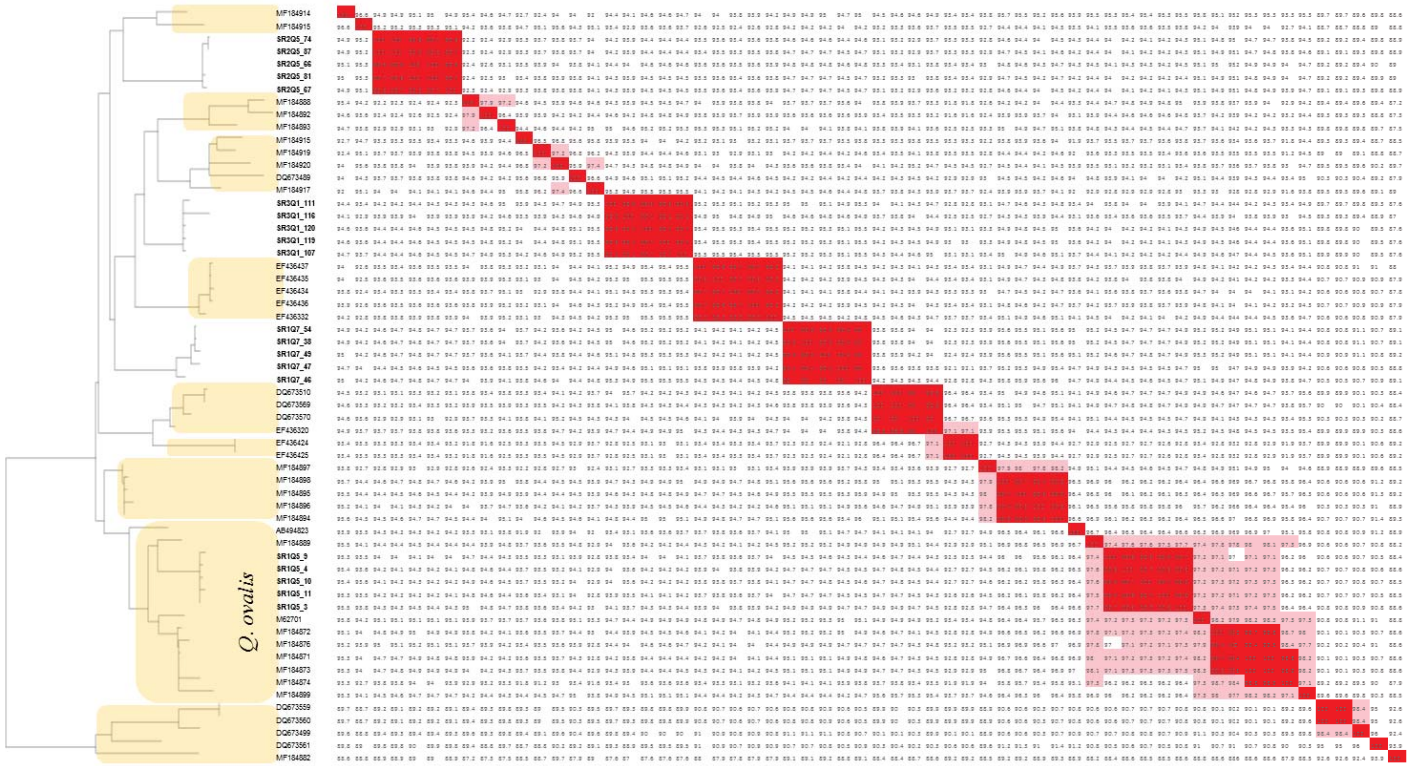


Figure 5.12 Sequence similarity matrix of clone library sequences amplified from the DNA samples used to generate the genome bins, and other sequences assigned to *Quinella*. Clone library sequence identifiers are represented in bold letters while similarity values are coloured using two different cut-offs dark red (≥98.7 %) and light red (97% to 98.7%), following the recommendations for delineating species made by Stackebrandt and Goebel (1994) and Kim et al. (2014) respectively. Sequences defining potential species of *Quinella* are delineated by yellow shapes in the phylogenetic tree on the left, based on the tree shown in Figure 3.11.

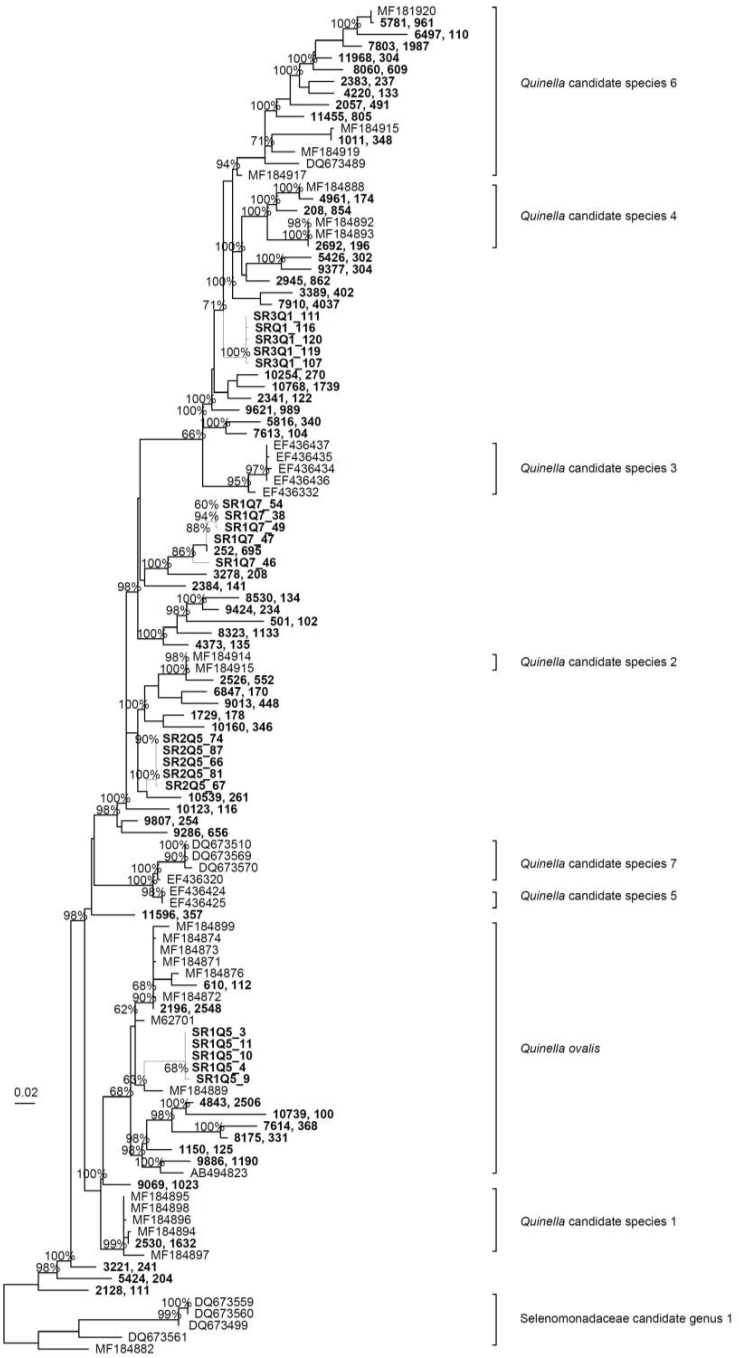


Figure 5.13 Phylogenetic distribution of repset sequences representing abundant OTUs of *Quinella* and 16S rRNA gene sequences amplified from the DNA samples used to generate the *Quinella* genome bins. Both type of sequences are in bold font. Repset sequences representing OTUs with ≥100 sequences were added to the tree (Figure 5.11) using the ARB parsimony (quick add mark) insertion function. Numbers after the comma give the total number of pyrosequencing reads in those OTUs. The candidate species are those defined in Figure 3.13. The scale bar represents 0.02 changes per nucleotides. Bootstrap values of <60% and some of those within defined clusters are not shown.

5.2.10 *Quinella* genome bin annotation

The contigs from the four *Quinella* bins were annotated using the methods described in section 2.18. A summary of the genome bin annotation is shown in Table 5.10. A summary of COG functional annotation based on orthologous protein is provided in Table A5.1 and Figure 5.14. Of the 8761 total genes identified in *Quinella* genome bins, 28.2 % were predicted to be of unknown function and annotated as hypothetical. It was interesting to find that, even though the overall genome size varied among *Quinella* genome bins, they all shared similar percentages of genes for particular COG (Clusters of Orthologous Groups) categories (Figure 5.14). Furthermore, in orthologous gene family analysis, it was also found that out of 1058 gene families from all four *Quinella* genome bins, 330 gene families were present in all four genome bins (Figure 5.15). However, genome bins SR3Q1 and SR1Q7 were found to be different from SR1Q5 and SR2Q5 in terms of the numbers of unique gene families. This finding was also supported by a Functional Genome Distribution (FGD) analysis (Figure 5.16), where genome bins SR1Q5 and SR2Q5 were more similar to each other, and SR1Q7and SR3Q1 were more similar to each other. FGD basically calculates the similarity between pairs of genome bins using amino-acid sequences predicted from the ORFeome (Altermann, 2012). This is a BLAST-based ORF-position-independent approach. It is based on amino acid sequence similarities of ORFs compared between genomes and is considered to be a function-based analysis. The groupings in the FGD tree were slightly different from the 16S rRNA gene-based analysis, in which SR1Q7and SR3Q1 were close to each other while SR1Q5 and SR2Q5 branched separately.

Table 5.10 *Quinella* genome bin statistics.

Attribute	*Quinella* genome bins			
	SR1Q5	SR1Q7	SR2Q5	SR3Q1
Genome size (bp)	2,125,473	2,584,672	1,821,931	2,614,227
DNA coding (%)	92.8	90.2	92.8	91.7
G + C (mol%)	49.0	52.9	56.0	49.1
Number of contigs	132	169	42	68
Total number of ORFs	2,067	2,445	1728	2521
Number of rRNAs	6	6	10	6
Number of tRNAs	348	438	316	465
Genes assigned to COGs	1769	2101	1543	2152
Genes with Pfam domains	1770	2030	1520	2130
Genes with signal peptides	184	203	171	222
Genes with transmembrane helices	426	475	363	508
CAZymes	68	83	62	93
CRISPR	0	1	0	2

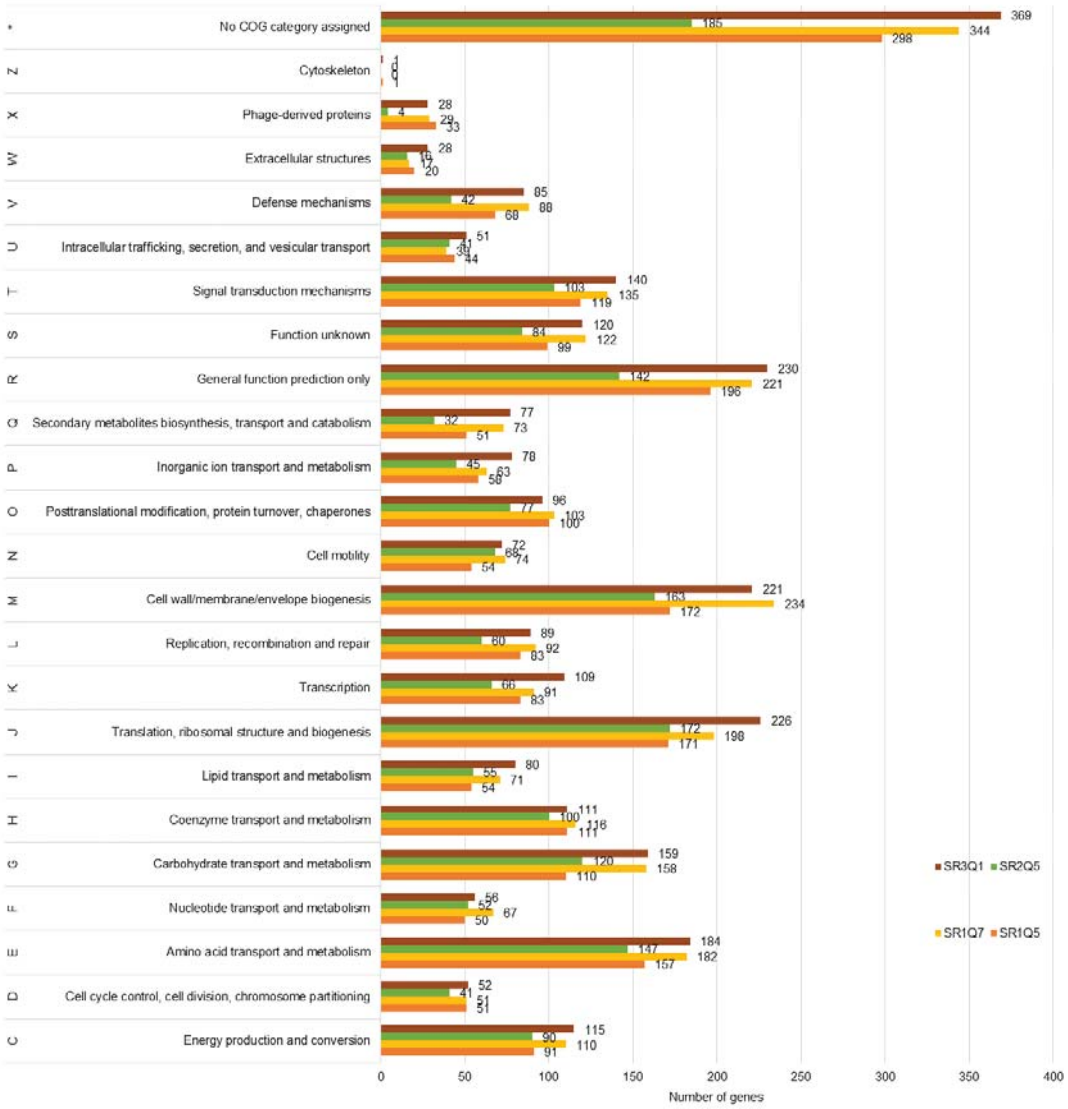


Figure 5.14 Functional classification of the predicted genes in the four *Quinella* genome bins based on the clusters of orthologous proteins (COGs) database (Tatusov et al., 2001b). The COG category codes (letters) are given on the extreme left.

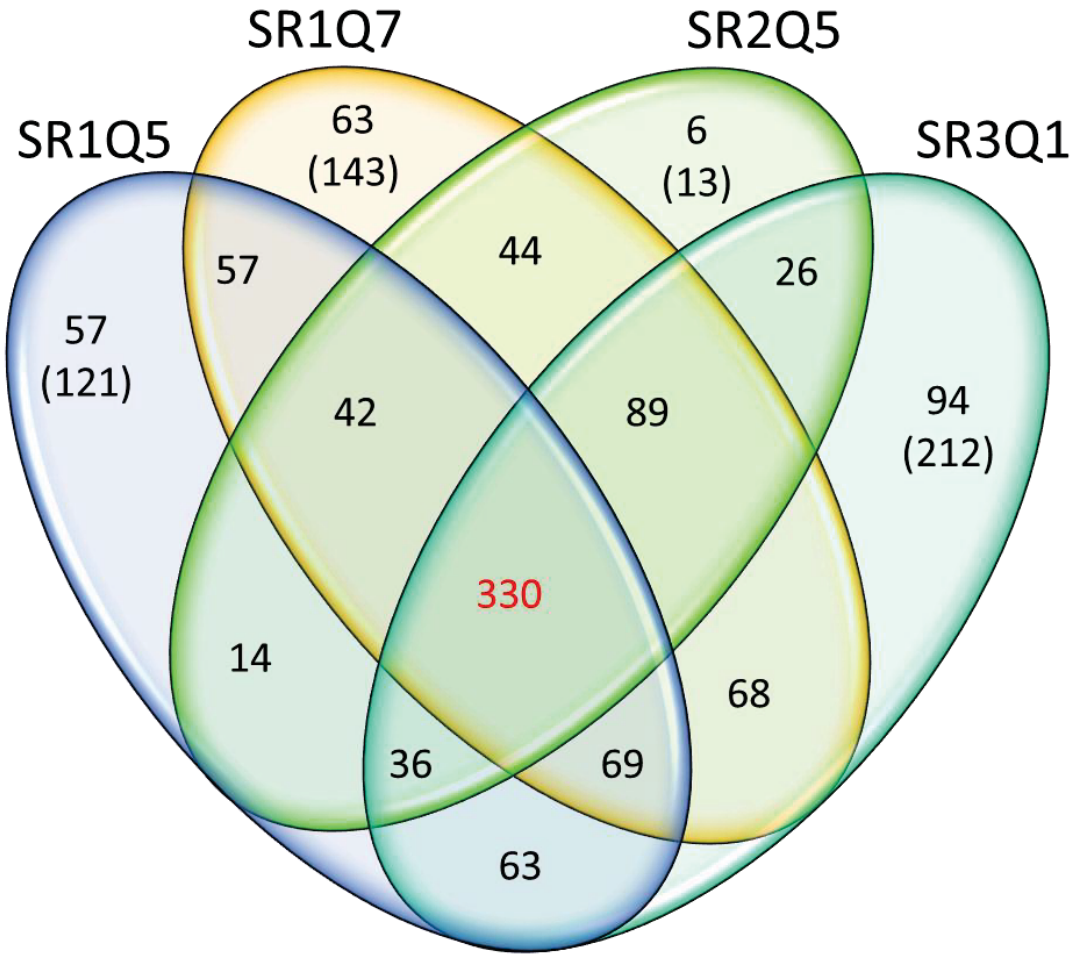


Figure 5.15 Venn diagram of orthologous protein families among all four *Quinella* genome bins. Unoverlapped regions represent unique gene families, with the numbers in the parentheses representing total number of genes present in those families.

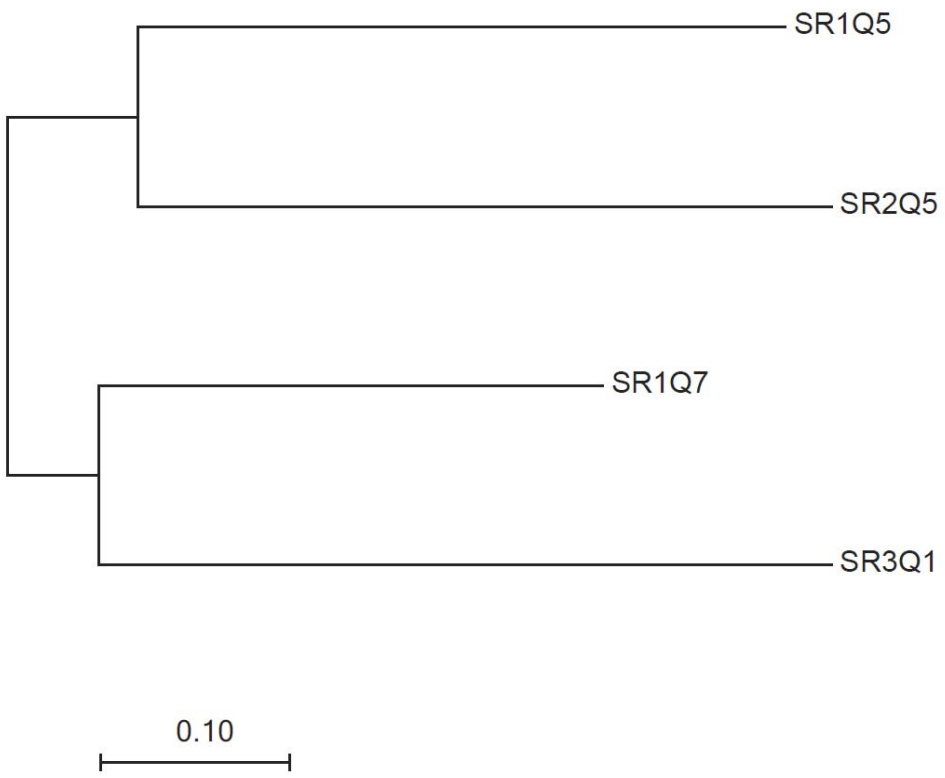


Figure 5.16 Functional genome distribution (FGD) tree of *Quinella* genome bins. The individual bins in FASTA format were concatenated using a universal spacer-stop-spacer sequence and automatically annotated using GAMOLA2 (Altermann et al., 2017). Predicted ORFeomes of all bins were subjected to an FGD analysis (Altermann, 2012) and the resulting distance matrix was imported into MEGA7 (Kumar et al., 2016).The functional distribution was visualized using the UPGMA method (Sneath and Sokal, 1962). The tree is drawn to scale, with branch lengths in the same units as those of the functional distances used to infer the distribution tree. The scale bar is a dissimilarity measure between genome bins based on the similarity and absence of ORFs in the genome dataset used to generate the tree (Altermann, 2012).

5.3 Conclusions

In agreement with the literature, FISH confirmed that *Quinella* cells were large and oval shaped. Electron microscopy showed that *Quinella* cells in the sheep studied here were about 3 to 5 µm long and 1 to 3 µm in diameter and that different cells can have different outer cell surface textures. This might be because *Quinella* cells present in samples were in different stages of growth cycles, or perhaps different species of *Quinella* were present. Some differences may be preparation artefacts. Analysis of 16S rRNA gene sequences also supported the existence of multiple candidate species of *Quinella*, even in a single rumen sample (see also chapter 3). TEM image showed that *Quinella* possesses a Gram negative cell wall structure even though on the basis of 16S rRNA gene sequences similarity it is placed in the phylum *Firmicutes*, which contains mainly gram positive bacteria.

After an unsuccessful attempt to isolate *Quinella*, its large oval cell size was used as a guide to develop a cell concentration method which resulted in a significant increase in the number of putative *Quinella* cells in processed rumen samples. DNA was extracted from three of these samples, and the DNA sequenced to generate metagenomic sequence data for genome analysis. Four genomic bins were assembled that contained, based on multiple analyses, largely *Quinella* DNA sequences, and the genomic information in them was estimated to be >90% complete and <11% contaminated. These *Quinella* genomic bins varied in apparent genome size and G+C content, but this is not unusual as it has been observed before in the closely related genus *Selenomonas*. Analysis of full length 16S rRNA gene sequences retrieved from the three samples, and attributed to the four bins based on sequence similarity of flanking regions, suggested that these genomic bins represented different *Quinella* species. All four *Quinella* genomic bins shared 330 protein families, and the number of unique protein families varied from 6 to 94, further supporting the potential diversity of *Quinella* genomes captured in this study.

The proteins predicted to be coded by the DNA sequences in the four genome bins were used to deduce the physiology of ruminal *Quinella* spp. This work is described in the next chapter.

Chapter 6 Insights into the physiology of *Quinella*, and its possible role in the rumen of low-methane-emitting sheep

6.1 Introduction

As discussed in previous chapters, *Quinella* spp. have not been studied as viable long term cultures that can be used to understand the physiology of members of this genus. However, two fermentation studies, one *in vitro* (Brough et al., 1970) and another *in vivo* (Vicini et al., 1987), have been conducted to understand the physiology of *Quinella*. The results from these studies appear to contradict each other. Brough et al. (1970) suggested that the preferred substrate, glucose, is fermented mainly to lactate, with traces of acetate, propionate and CO_2 also detected. In contrast, Vicini et al. (1987) suggested that, when *Quinella* spp. were abundantly present in sheep fed molasses, acetate and propionate were the major end products with no lactate formation. The latter study was based on measured products in the sheep rumen, which can be affected by the presence of other microorganisms. Other than these studies, very little work has been done on *Quinella* spp. and there were almost no further reports until Kittelmann et al. (2014) identified the association of rumen-dwelling *Quinella* spp. with low CH_4 emissions in sheep.

In the previous chapter, four genome bins were generated from DNA extracted from samples enriched for cells that match the morphology of *Quinella* and bound DNA probes that targeted the 16S rRNA of *Quinella* spp. These genome bins contained partial 16S rRNA genes that grouped with a radiation of bacteria identified as *Quinella* spp. Longer 16S rRNA gene sequences were amplified from the DNA samples that were the source of the genome bins, and these closely matched the partial genes in the genome bins, and grouped with 16S rRNA gene sequences designated as *Quinella*. Based on these findings, and on other analyses of the sequences in the four genome bins (see Chapter 5), the assumption was made that the genome bins contain significant parts of the genomes of different species of *Quinella*. An analysis of the deduced proteins coded by these partial genomes was undertaken to give an insight into the potential physiology of *Quinella* spp. and predict the role of member of this genus in rumen fermentation.

6.2 Results and discussion

6.2.1 Degradation of polysaccharides

Major plant polysaccharides are basically grouped into four major fractions: cellulose, starch, hemicellulose and pectin (Chesson and Forsberg, 1997). The first step in rumen fermentation is the hydrolysis of these polymeric components of feed to form simpler sugars. The genes in the four *Quinella* genome bins genes were compared with different databases (NCBI non-redundant protein sequences database, Pfam database, TIGRFAMs database) using BLAST in GAMOLA2 (Altermann et al., 2017) as the search tool to identify their ability to attack plant polysaccharides. A separate carbohydrate-active enzymes (CAZyme) search was conducted using the CAZyDB database (Lombard et al., 2014), which contains a curated reference set of carbohydrate degrading enzymes. A summary of the CAZyme analysis of the four *Quinella* genome bins is presented in Table 6.1 and details can be found in Appendix Table A6.1.

Enzymes in GH families 1, 3, 13, 23, 77 and 84 were present in all genome bins. Details of these enzymes are listed in Table 6.2. Signal peptide and transmembrane topology analysis using Phobius (Käll et al., 2004) and SPOCTOPUS (Viklund and Elofsson, 2008) suggested that all of these enzymes were non-cytoplasmic. Among these enzymes, β-glucosidase, α-phosphotrehalase, transglycosylase and *O*-GlcNAcase ((protein)-3-*O*-(*N*-acetyl-D-glucosaminyl)-L-serine/threonine *N*-acetyl-glucosaminyl hydrolase) contained potential signal peptides, which suggested they are secreted and act extracellularly.

Glycosyl hydrolases (GH), a CAZYme class of enzymes that hydrolyse the glyosidic bonds between carbohydrates or between carbohydrate and non-carbohydrate components of polysaccharides, contains most of the enzymes that are involved in cellulose and hemicellulose degradation. Of those enzymes, endo-1,4-β-D-glucan hydrolase (EC 3.2.1.4), exo-1,4-β-D-glucan cellobiohydrolase (EC 3.2.1.91) and β-D-glucosidase (EC 3.2.1.21) are the key enzymes needed for cellulose hydrolysis (Forsberg et al., 1997). None of the GH enzymes found in the *Quinella* genome bins appear to be endo-1,4-β-D-glucan hydrolases or exo-1,4-β-D-glucan cellobiohydrolases, but β-glucosidase was present in all genome bins. This enzyme cannot degrade cellulose alone, as it is involved in the third and last step of cellulose hydrolysis (conversion of cellobiose and cellodextrins to glucose). Thus, it appears that none of the *Quinella*

genomes encode the primary enzymes for cellulose-degradation. However, the β-glucosidase may act on cellobiose and cellodextrins, released by other rumen microbes, and may provide an advantage in the highly competitive rumen environment. Similarly, 1,4-β-D-xylan xylanohydrolase (EC 3.2.1.8) and 1,4-β-D-xylan xylohydrolase (EC3.2.1.37) are key enzymes involved in xylan degradation (Krause et al., 2003). These enzymes fall in GH families 10 and 11. The amino acid sequences for these enzymes were also searched for in the four *Quinella* genome bins but none of these CAZyme families were found.

Acetyl-xylan esterase (EC 3.1.1.72) and feruloyl esterases (EC 3.1.1.73), which belong to carbohydrate esterases (CE) families 1 to 7 and hydrolyse ester-linked side groups of xylan (Henrissat and Bairoch, 1993) and, were absent from the *Quinella* genome bins. However, 8 to 19 CE enzymes were detected in *Quinella* genomic bins and they belonged to CE family 10 (esterase, putative carboxylesterase and Tat pathway signal sequence) which are not involved in primary attacks on complex plant polysaccharides.

Gene sequences coding for pectin-degrading enzymes (pectin lyase, polygalacturonase and pectin methylesterases) which fall in CAZyme families PL8, PL16, GH28 and CE8 were also searched for in the *Quinella* genome bins but were not found.

Commonly found starch-degrading enzymes are α-amylase, β-amylase, dextrin 6-α-glucanohydrolase, pullulanase and iso-amylase (Cerrilla and Martínez, 2003). None of the *Quinella* genome bins contained genes sequences coding for any of these starch-degrading enzymes.

Quinella cells have been reported to contain storage polysaccharides (Brough et al., 1970). Three of the enzymes found in all the genome bins may have a role in glycogen metabolism (1,4-α-glucan branching enzyme, α-phosphotrehalase and 4-α-glucanotransferase). The predicted extracellular location of these enzymes and the presence of a signal peptide in the α-phosphotrehalase makes the inference of their roles in intracellular glycogen metabolism speculative.

Almost one-third of the of the glycosyl transferases (GT) found in the *Quinella* genome bins were identified as poly-β-1,6 *N*-acetyl-D-glucosamine synthases, potentially involved in polymerisation of UDP-*N*-acetylglucosamine in bacterial cell wall synthesis. It was interesting to find lytic transglycosylase, a lysozyme type enzyme, which in some cases is reported to contribute to pathogenesis (Cloud-Hansen et al., 2006). However, it

is predicted that, together with transglycosylase, lytic transglycosylase may take part in the conversion of peptidoglycan to 1,6-anhydro sugars (Scheurwater et al., 2008) and so these may play a role in cell wall metabolism. *O*-GlcNAcase was found in all *Quinella* genome bins but its function in bacteria is still unknown (Alonso et al., 2014). It too acts on *N*-acetylglucosamine, one of the two amino sugars in peptidoglycan. The *O*-GlcNAcase and transglycosylase had predicted extracellular locations (see above), which is expected for enzymes that act on the cell wall.

Overall, these analyses suggested that *Quinella* does not contain genes that encode enzymes responsible for the primary degradation of plant polysaccharides. Overall, it seems that *Quinella* may rely on other microbes to hydrolyse the polysaccharide components of feed, and it appears able to compete effectively for the breakdown products of that primary attack for its own growth, as well as presumably using soluble sugars in the plant material. In some of the sheep analysed in this study (section 3.2.7), *Quinella* 16S rRNA gene sequences made up as much 74.4% of all bacterial 16S rRNA gene sequences surveyed. Of course, this may be a result of some PCR bias or due to a large number of 16S rRNA genes in each genome, but it may also suggest that *Quinella* is exceptionally good at scavenging smaller carbohydrates released by the primary polysaccharide degraders.

Table 6.1 CAZyme counts in the *Quinella* genome bins (SR1Q5, SR1Q7, SR2Q5, and SR3Q1)[a]. E-values < 1e-18 and coverage > 0.35 (Yin et al., 2012) were considered as good matches, and genes meeting both of these criteria were considered to be potential CAZymes.

Carbohydrate-active enzymes (CAZymes) classes	*Quinella* genome bins				
	SR1Q5	SR1Q7	SR2Q5	SR3Q1	Total unique CAZymes
Auxiliary activities (AAs)	0	0	1	0	1
Carbohydrate esterases (CEs)	13 (1)	8	19 (1)	16 (3)[b]	5
Glycoside hydrolases (GHs)	13	15 (1)	12 (1)	18 (2)	4
Carbohydrate-binding modules (CBMs)	0	0	0	0	0
Glycosyl transferases (GTs)	33 (2)	58 (4)	39 (1)	53 (4)	11
Polysaccharide lyases (PLs)	0	0	0	0	0

[a]See Appendix A6.1 for information of individual genes identified in each of the CAZymes classes.
[b]Values in parentheses represent unique CAZYmes present only in one bin out of four.

Table 6.2 GH enzyme family and related enzymes found in all *Quinella* genome bins.

GH family	Enzymes	Presence of signal peptides[a]	Possible function
GH1	β-Galactosidase	No	Galactose breakdown
GH3	β-Glucosidase	Yes	Cellobiose and cellodextrins breakdown
GH13	1,4-α-Glucan branching enzyme	No	Glycogen degradation
	α-Phosphotrehalase	Yes	Glycogen degradation
GH23	Lytic transglycosylase	No	Converts peptidoglycan to 1,6-anhydro sugars
	Transglycosylase	Yes	Convert peptidoglycan to 1,6-anhydro sugars
GH77	4-α-Glucanotransferase	No	Glycogen degradation
GH84	*O*-GlcNAcase	Yes	Unknown

[a]All are predicted to be non-cytoplasmic enzymes.

6.2.2 Degradation of sugars to pyruvate

Each of the four *Quinella* genome bins contained all the genes needed to code for the enzymes for glucose fermentation to pyruvate (Figure 6.1), with the exception of pyruvate kinase in bin SR2Q5 (Appendix Table A6.2). Transporters were searched for using BLAST with the TransportDB 2.0 database (Elbourne et al., 2017). Phosphotransferase systems (PTS) for glucose, sorbitol, fructose, maltose, mannose, galactitol, glucitol and ascorbate were detected (Appendix Table A6.3). The phosphoenolpyruvate-dependent PTS system consists of the shared enzyme I (phosphoenolpyruvate-protein phosphotransferase, PstI) and phosphocarrier protein (Hpr), and the substrate-specific enzyme II (Tchieu et al., 2001). These enzymes were present in all *Quinella* genome bins. Enzymes I and Hpr are listed in Appendix Table A6.2. The substrate-specific enzymes II, were present either as multi subunits proteins in operons or in fused forms (Appendix Table A6.3). By using PTS, glucose, for example, would be imported and phosphorylated, and then converted to two pyruvate molecules via the standard glycolytic pathway, characterised by an ATP-dependent phosphofructokinase and a fructose bisphosphate aldolase. The gene coding for triosephosphate isomerase, a key enzyme in glycolysis, was present as a fused gene with phosphoglycerate kinase in all four *Quinella* genome bins. This type of fusion has been reported before in *Thermotoga maritima* (Schurig et al., 1995). *Quinella* genome bin SR2Q5 lacked the gene for pyruvate kinase that catalyses the conversion of phosphoenolpyruvate (PEP) to pyruvate (Waygood and Sanwal, 1974), but pyruvate kinase was present in the other three *Quinella* genome bins (Appendix Table A6.2). This *Quinella* strain may use alternative enzymes to make pyruvate. For example, in the fermentation of glucose to two pyruvate, it could use the glucose PTS to convert one molecule of PEP to one molecule of pyruvate and the other pyruvate may be formed by carboxylation of PEP to oxaloacetate followed by decarboxylation of oxaloacetate to pyruvate. However, it seems most likely that this gene was just missed from bin SR2Q5 at the assembly step. In the formation of two molecules of pyruvate from one molecule of glucose, one molecule of ATP will be used (fructose-6-phosphate to fructose-1,6-bisphosphate) and three ATP will be released (two at the step of 1,3-bisphospho glycerate to 3-phosophoglycerate and one at the step of phosphoenolpyruvate to pyruvate), resulting in the net gain of two ATP.

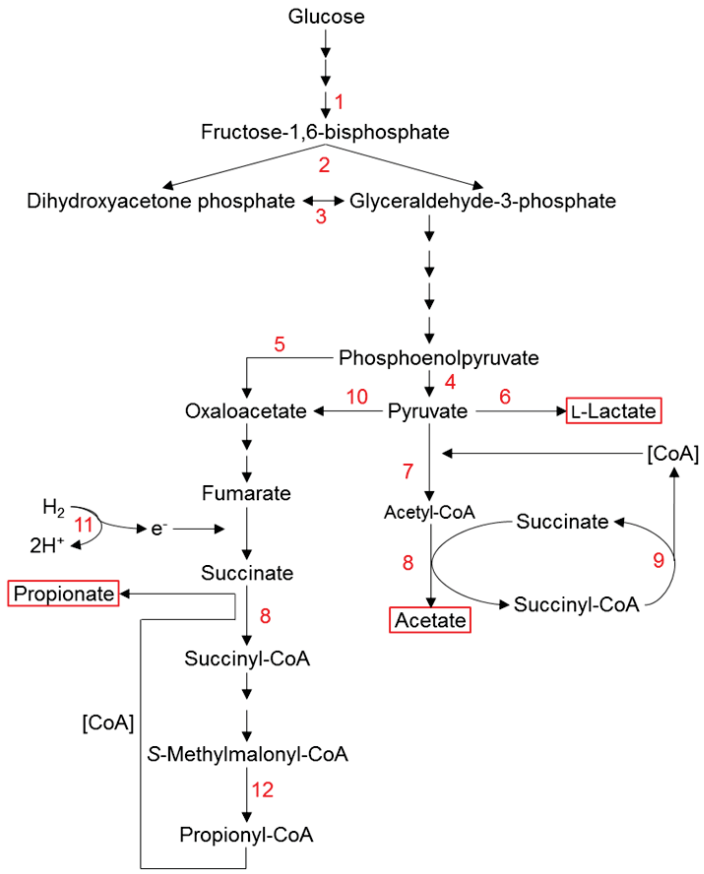
6.2.3 Other sugar fermentation pathways

The *Quinella* genome bins were also searched for genes that code for enzymes involved in the pentose phosphate and Entner-Doudoroff pathways. These may act as parallel or alternative pathways to glycolysis (Cronín et al., 1989; Chen et al., 2016). The *Quinella* genome bins contained three (ribose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase) of four genes (transaldolase is only present in SR3Q1) that code for the enzymes in the non-oxidative phase of the pentose phosphate pathway. As the genes for the key enzymes of the reversible oxidative phase (glucose-6-phosphate dehydrogenase, 6-phosphogluconolactonase and 6-phosphogluconate dehydrogenase) were missing, it doesn't seem possible for it to result in sugar transformation to pyruvate. However, the pathway can operate to make intermediate products required for a number of biosynthetic pathways (D-ribose 5-phosphate, D-ribulose 5-phosphate, D-xylulose 5-phosphate and D-erythrose 4-phosphate).

Genes coding for key enzymes (6-phosphogluconate dehydratase and 2-keto-3-deoxygluconate 6-phosphate aldolase) of the Entner-Doudoroff pathway (Murray and Conway, 2005) were missing in all *Quinella* genome bins, suggesting that this pathway was not present in *Quinella*.

6.2.4 End products from pyruvate

All four *Quinella* genome bins contained a gene coding for L-lactate dehydrogenase (Figure 6.1) that could be involved in lactate utilisation or in lactate formation from pyruvate. Analysis of whether *Quinella* is a lactate user or producer is discussed below, in section 6.2.4.1. Additionally, KEGG analysis of the *Quinella* genome bins suggested that it might produce propionate through the succinate-to-propionate conversion pathway (Figure 6.1). Propionate formation by *Quinella* is discussed later in this chapter with a deeper investigation of key enzymes involved (sections 6.2.5). The *Quinella* genome bins were also analysed for the formation of other volatile fatty acids (formate, acetate and butyrate), and other end products, specially ethanol and hydrogen.



Enzyme numbers	Enzyme names	Section numbers
1	Phosphofructokinase	6.2.2
2	Fructose bisphosphate aldolase	6.2.2
3	Triose phosphate isomerase	6.2.2
4	Pyruvate kinase	6.2.2
5	Phosphoenolpyruvate carboxykinase	6.2.1.1
6	L-Lactate dehydrogenase	6.2.4.1
7	Pyruvate:ferrodoxin/flavodoxin oxidoreductase	6.2.4.2
8	Succinate-CoA transferase	6.2.4.2
9	Succinyl-CoA synthetase	6.2.4.2
10	Oxaloacetate decarboxylase	6.2.5.1
11	Hydrogenase	6.2.5.3
12	Methylmalonyl-CoA decarboxylase	6.2.5.1

Figure 6.1 Deduced fermentation pathways of *Quinella*. Numbers in the flow chart represent enzymes which are listed in the table together with the section numbers where they are discussed.

6.2.4.1 Lactate dehydrogenase

BLAST analysis of the *Quinella* genome bins suggested that all of them contained genes that code for L-lactate dehydrogenase (L-LDH), which indicates that *Quinella* may either produce lactate or use it for growth or possibly both. There appears to be no way to differentiate between lactate producing and lactate using L-LDHs based on amino acid sequence motifs. However, recently, a metagenomic and metatranscriptomic study (Kamke et al., 2016) on *Sharpea*-enriched low-methane-yield sheep rumen microbiomes suggested that there may be 11 distinct *ldh* clusters from bacteria present in the rumen. The authors of that study predicted that the L-LDH from *Sharpea* spp. was associated with lactate formation, while L-LDH from *Megasphaera* spp. was associated with lactate utilisation. Furthermore, substrate utilisation experiments and measurement of fermentation end products conducted with *Sharpea* and *Kandleria* spp. showed that these organisms are lactate producers (see Chapter 4). Sequences of L-LDH from *Megasphaera* spp. together with some other known lactate users (*Propionibacterium freudenreichii*, *Clostridium propionicum*, *Veillonella parvula* and *Desulfovibrio vulgaris*) and those from *Sharpea* and *Kandleria* spp. and other known lactate producers (*Lactobacillus acidophilus*, *Streptococcus equinus*, *Butyrivibrio fibrisolvens*, *Olsenella umbonata*, and *Ruminococcus gnavus*) were used for a phylogenetic analysis of L-LDH from the *Quinella* genomes bins. L-LDH from *Selenomonas ruminantium*, which is both a producer and user of lactate (Kanegasaki and Takahashi, 1967), was also included. All sequences were >300 amino acids long, and were considered to be nearly full length. A phylogenetic tree was constructed using the malate dehydrogenase sequence from *M. elsdenii* as an out group (Figure 6.2). The analysis showed that L-LDHs from lactate users and lactate producers did not group separately. LDH from the lactate producers *Sharpea*, *Kandleria* and *Olsenella* clustered together with 99% bootstrap support, whereas others formed separate groups (*B. fibrisolvens* and *R. gnavus* in one, and LDH1 and LDH2 of *L. acidophilus* with *S. equinus* in another). LDH from the lactate-using *V. parvula, C. propionicum* and *M. elsdenii* formed one cluster with 95.8% bootstrap support, but were separated from LDH from *D. vulgaris* and *P. freudenreichii*, which are also lactate users that formed separate branches in the tree. All *Quinella* LDH sequences were >93 % similar to each other and formed a stable group with LDH from *S. ruminantium*, with bootstrap support of 91.2%, perhaps suggesting that LDH in *Quinella* might be involved in both lactate use and formation. However, experimental proof will be needed to confirm this. Orpin (1972),

observed the requirement for lactate as a sustrate while growing *Quinella* in mixed suspensions, but it was not clear whether the lactate was used by *Quinella* or other contaminating bacteria. Therefore, the role of lactate in the metabolism of *Quinella* remains unresolved.

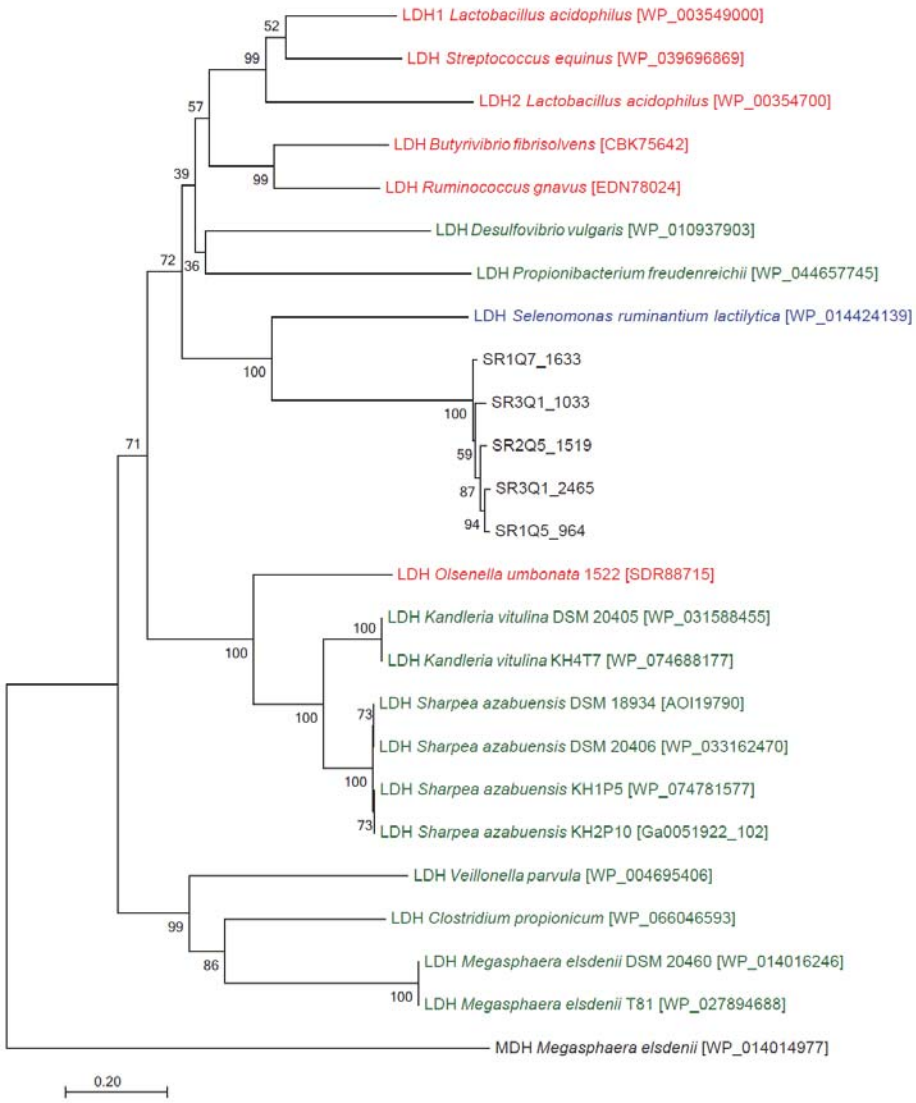


Figure 6.2 Phylogenetic analysis of L-lactate dehydrogenases from the *Quinella* genome bins. Malate dehydrogenase of *M. elsdenii* was used as the out group. The tree was constructed using the Jukes-Cantor genetic distance model with the Neighbor-Joining method and 500 bootstrap resamplings. LDH, lactate dehydrogenase; MDH, malate dehydrogenase. Lactate producing species are coloured red while lactate utilising species are coloured green. *Selenomonas ruminantium*, coloured blue, can both use and produce lactate. The scale bar represents 0.2 changes per amino acid residue. The numbers at the nodes are the percentage of trees that conserved that node in 500 bootstrap resamplings. GenBank accession numbers are given after each sequence name.

6.2.4.2 Acetate formation

The deduced activities of the proteins coded by genes in the genome bins suggested that *Quinella* may produce propionate as a major end product. However, to produce propionate from pyruvate requires four electrons. Two of those electrons can be gained from the conversion of 0.5 glucose to 1.0 pyruvate but to be able to get the other two electrons, *Quinella* would have to oxidize pyruvate to acetate or use external electrons, for example, from hydrogen produced by other microbes using a hydrogenase (section 6.2.5.3). The electrons are required in the malate dehydrogenase step (in the form of NADH) and the fumarate reductase step (in the form of reduced quinones). The possibility that *Quinella* can produce acetate will be discussed here. The most common way for anaerobes to produce acetate from pyruvate is to first convert pyruvate to acetyl-CoA using pyruvate:ferredoxin (or flavodoxin) oxidoreductase (PFOR) or a pyruvate:formate lyase.

All four *Quinella* genome bins contained genes that code for PFOR (Appendix Table A6.2), that transfers electron to either ferredoxin or flavodoxin. The analysis was conducted to first confirm the presence of PFOR in all four *Quinella* genome bins and then to identify whether they are ferredoxin associated or flavodoxin associated. For the analysis, PFOR amino acid sequences of flavodoxin-type (nifJ) was represented by *Escherichia coli*, *Anabaena variabilis* and *Klebsiella pneumoniae* while *Desulfovibrio africanus* was used as a ferredoxin-type PFOR (Figure 6.3). Amino acid sequences of *Selenomonas ruminantium* and *Megasphaera elsdenii* PFOR were also included in the analysis as they were identified as closest relatives in GAMOLA2 annotation of PFOR in the *Quinella* genome bins. On the basis of sequence similarity and phylogenetic analysis (Figures 6.3A and 6.3B), *Quinella* PFOR were most similar to NifJ type PFOR with highest sequence similarity with *K. pneumoniae* (≥60.4%), and they also had ≤55.5% amino acid sequence similarity with *D. africanus*. Metal-binding cysteine residues (Pieulle et al., 1995) were also found conserved in all analysed sequences suggesting that both type of PFOR are very similar to each other. However, presence of genes for NAD(P)H:flavin oxidoreductase (Appendix Table A6.2), which would transfer electrons from the reduced flavodoxin to NADH, and the absence of any genes for enzymes that can transfer electrons from ferredoxin through to quinone, in all four *Quinella* genomic bins, suggests that the *Quinella* PFOR is a flavodoxin-linked oxidoreductase. Nevertheless, it was surprising not to find any enzymes that transfer electrons from

NADH to quinone or ubiquinone (Figure 6.4). Usually NADH:quinone oxidoreductase catalyses this reaction, but genes for this were absent in all four *Quinella* genome bins.

A

```
          E. Coli  649 GLGDALPVSALP---PDGIWPMGTTRWEKRNIAEEIPIWKEELCTQCNHCVAACPHSAIR
     A. variabilis  656 REGDELPVSALPN---DGTYPTGTAKWEKRNIAQEIPVWDTDVCIQCGKCVMVCPHSVIR
     K. pneumoniae  649 QCGDQLPVSAFVG-MEDGTFPSGTAAWEKRGIALEVPVWQPEGCTQCNQCAFICPHAAIR
      D. africanus  649 QQGDKLPVSAFE---ADGRFPLGTSQFEKRGVAINVPQWVPENCIQCNQCAFVCPHSAIL
       M. elsdenii  648 LEGDDLPVSAYDG-IEDGRWPTGTCAEEKRGVAMFVPSWDAEKCIGCNQCSFVCPHAAIR
   S. ruminantium  648 QEGDKLPVSKFNGEMVDGTFPVGGAAYEKRGTAINVPVWNTDKCIGCNQCSYVCPHAAIR
        SR1Q5_1976  352 QEGDALPVSKFVE-LADGTFPVGGTAYEKRGTAIKVPQWDKSKCIGCNQCSFVCPHAAIR
         SR1Q7_152  650 QEGDELPVSKFVG-LADGTFPVGGTAYEKRGTAIKVPAWDKSKCIGCNQCSFVCPHAAIR
         SR2Q5_868  650 QEGDILPVSKFVK-LADGTFPVGGTAYEKRGTAIKVPAWDKSKCIGCNQCSFVCPHAAIR
        SR3Q1_2206  650 QEGDNLAVSKFVE-LADGTFPVGGTAYEKRGTAIKVPAWDKSKCIGCNQCSFVCPHAAIR

          E. coli  706 AKVVPPEAMENAPASLHSLDVKSRDMRGQKYVLQVAPEDCTGCNLCVEVCPAKDRQNPEI
     A. variabilis  713 SKVYEPEQLENAPSTFKSANAKDHDWHGLKFTIQVAAEDCTGCGICVDVCPAKNKAQPRK
     K. pneumoniae  708 PALLNGEEHDAAPVGLLSKPAQGAK--EYHYHLAISPLDCSGCGNCVDICPA------RG
      D. africanus  706 PVLAKEEELVGAPANFTALEAKGKELKGYKFRIQINTLDCMGCGNCADICPP------KE
       M. elsdenii  707 PFLMTPDEAAKAPKGYQDKEIKVAP--GYKYNIIVSVMDCLGCGSCTHVCP--------K
   S. ruminantium  708 PVLTTEEELKNAPAGFPSKQTKAIK--DYTFTVAVSTMDCLGCGNCAQVCP--------V
        SR1Q5_1976  411 PVLTTPEELKNAPEGMEAIPSKISR-GAYNLTIAVSTLDCLGCGNCAQVCP--------K
         SR1Q7_152  709 PILTTADERKNAPEGFESIPSKISR-GAYDLTIAVSTYDCLGCGNCAQVCP--------K
         SR2Q5_868  709 PVLTTNDELANAPEGFESIPSKISR-GAYNLTIAVSTYDCLGCGNCAQVCP--------K
        SR3Q1_2206  709 PILTTNDELKNAPEGMESIPSKISR-GAYNLTIAVSTMDCLGCGNCAQVCP--------K
```

B

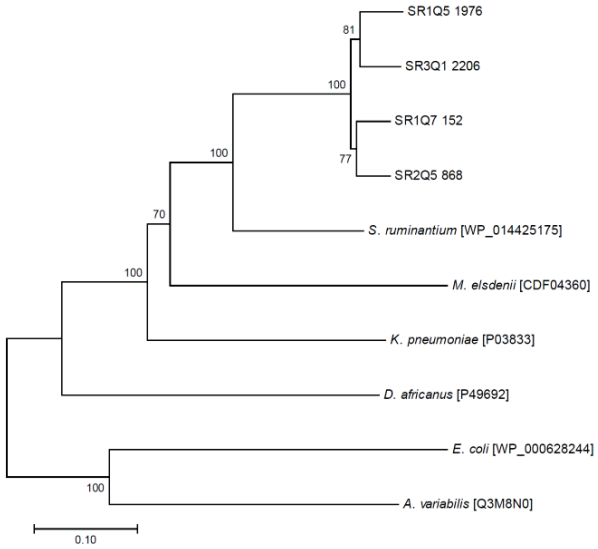


Figure 6.3 Identification of PFOR in the *Quinella* genome bins. A) Amino acid alignment of *Quinella* PFOR with *E. coli*, *A. variabilis*, *K. pneumoniae*, *D. africanus*, *S. ruminantium*, *M. elsdenii.* All metal binding cysteine (**C**) residues were found conserved in all analysed sequences. B) Phylogenetic analysis of PFOR from the *Quinella* genome bins. The tree was constructed using the Jukes-Cantor genetic distance model with the Neighbor-Joining method and 500 bootstrap resamplings. The scale bar represents 0.1 changes per amino acid residue. The numbers at the nodes are the percentage of trees that conserved that node in 500 bootstrap resamplings. GenBank accession numbers are given after each sequence name.

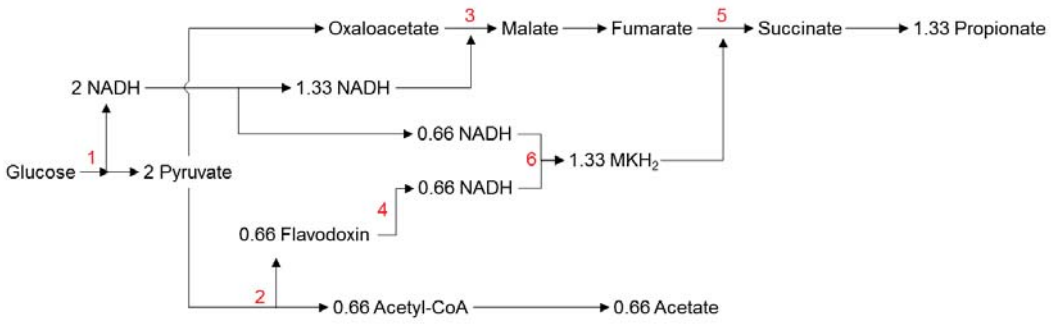


Figure 6.4 Pathway showing possible electron flow from the glycolytic pathway and conversion of pyruvate to acetyl-CoA. Numbers in the flow chart represent enzymes. (1) enzymes involve in glycolytic pathway, (2) PFOR, (3) malate dehydrogenase, (4) NAD(P)H:flavin oxidoreductase, (5) fumarate reductase (discussed latter in this chapter), (6) NADH:quinone oxidoreductase. An enzyme catalysing step 6 was not found in any of the *Quinella* genome bins.

From acetyl-CoA, there are four different routes for acetate formation (Figure 6.5), three of which generate ATP (Gottschalk, 1986; Mack and Buckel, 1997; Ferry, 2015). All four *Quinella* genome bins were searched for the genes involved in these pathways. None of the genome bins contained genes that code for phosphate acetyl transferase or acetate kinase, suggesting that the acetate-forming pathways characterised by these two enzymes (Figure 6.5A) is not present in *Quinella*. Similarly, acetyl-CoA synthetase (Figure 6.5B) was also absent in all four *Quinella* genome bins, which left two other options (Figure 6.5C and 6.5D) for acetate formation by *Quinella*. All four *Quinella* genome bins contained DNA sequences that coded for putative acetyl-CoA hydrolases and succinyl-CoA synthetases, suggesting that *Quinella* could use pathways C or D in Figure 6.5. The highest percentage identities of these two enzymes were with proteins coded by the *Selenomonas ruminantium* NBRC 103574 genome (>78% identity), annotated as a acetyl-CoA hydrolase and a succinyl-CoA synthetase. Acetyl-CoA hydrolase can convert acetyl-CoA to acetate by hydrolysis without ATP formation (Figure 6.5C). Succinyl-CoA synthetase cannot be used for acetate formation unless succinate CoA-transferase is also present (Figure 6.5D). Further analysis of the putative acetyl-CoA hydrolase from *Quinella*, using different databases used in the GAMOLA2 annotation pipeline, suggested that the protein annotated as an acetyl-CoA hydrolase in *S. ruminantium*, and by extension the homologue in *Quinella*, may actually be a succinate CoA-transferase. If that is the

case, then *Quinella* would code for enzymes to convert acetyl-CoA to acetate and generate ATP through a succinate/succinyl-CoA cycle (Figure 6.5D).

To better understand this possible acetyl-CoA hydrolase/succinate CoA-transferase, the deduced protein sequences were analysed in more detail. Mack and Buckel (1997) conducted a study on the glutaconate CoA-transferase from *Acidaminococcus fermentans*. Glutaconate CoA-transferase is another CoA-transferase that catalyses the transfer of CoA from acetyl-CoA to glutaconate. Mack and Buckel (1997) showed that the change of glutamate (E54) in the active site to aspartate (D54) converted the glutaconate CoA-transferase to a glutaconate-CoA hydrolase. Additionally, this glutamate residue in the active site is essential to transfer the CoA moiety to the acceptor acid (Solomon and Jencks, 1969). Using this clue, putative acetyl-CoA hydrolase sequences from the four *Quinella* genome bins were aligned with amino acid sequences of glutaconate CoA-transferase of *A. fermentans*, the putative acetyl-CoA hydrolase of *S. ruminantium*, the reviewed acetyl-CoA hydrolase of *Saccharomyces cerevisiae* and the reviewed succinate CoA-transferase of *Clostridium kluyveri*. From the alignments it was deduced that the *Quinella* genome bins contained a succinate CoA-transferase instead of an acetyl-CoA hydrolase. This was because the diagnostic glutamate (E54 in the *A. fermentans* glutaconate CoA-transferase, E432 in *Quinella*) was found to be conserved in all four *Quinella* genome bins, as found in known CoA-transferases (Mack and Buckel, 1997) (Figure 6.6). In contrast, an aspartate (D447 in *Saccharomyces cerevisiae*) was found in the homologous position in known acetyl-CoA hydrolases (Figure 6.6). Furthermore, two conserved motifs identified by Tielens et al. (2010) in CoA-transferases family 1 (the only family out of three that is involved in acetate formation (Heider, 2001) was also found in all analysed CoA-transferases, suggesting that not only do all the *Quinella* genome bins contain CoA-transferase, but that the putative acetyl-CoA hydrolase of *S. ruminantium* may also be a succinate CoA-transferase. Tielens et al. (2010) also divided CoA-transferases family 1 into three sub-families based on amino acid sequence homology. The CoA-transferases in the *Quinella* genome bins, and of *S. ruminantium*, *S. cerevisiae* and *Clostridium kluyveri*, belong to subfamily 1C on the basis of overall sequence homology and the sequence similarity in the conserved motif GxGGxD (Figure 6.6).

Once the likely presence of succinate CoA-transferase in the *Quinella* genome bins was confirmed, it was logical to predict that *Quinella* may convert acetyl-CoA to acetate using

succinate CoA-transferase to generate succinyl-CoA from succinate, then use succinyl-CoA synthetase to regenerate succinate with the formation of ATP from ADP (Figure 6.5D). Succinyl-CoA synthetase is composed of two subunits (α and β), and works bidirectionally (Fraser et al., 1999). DNA sequences predicted to encode both of these subunits were found in all four *Quinella* genome bins except that the β subunit was absent from genome bin SR3Q1. In conclusion, on the basis of genes identified in the genome bins, it appears that *Quinella* may catalyse a succinate-dependent conversion of acetyl-CoA to acetate using succinate CoA-transferase and succinyl-CoA synthetase. This pathway has not been reported in bacteria, but is present in parasitic helminths and protists (Rivière et al., 2004; Van Grinsven et al., 2008) and in the rumen fungus *Neocallimastix* sp. L2 (Marvin-Sikkema et al., 1993). Based on the analysis presented here, it is possible that *S. ruminantium* may also use this pathway. However, a biochemical study conducted on *S. ruminantium* to identify the key enzymes involved in fermentation of hexoses to different end products suggested that acetyl-CoA synthetase is involved in the conversion of acetyl-CoA to acetate (Michel and Macy, 1990). This suggested that *S. ruminantium* uses, pathways A or B as illustrated in Figure 6.5. It would be interesting to investigate whether succinate CoA-transferase and succinyl-CoA synthetase are actually involved in acetate formation (pathway D in Figure 6.5) in *S. ruminantium*.

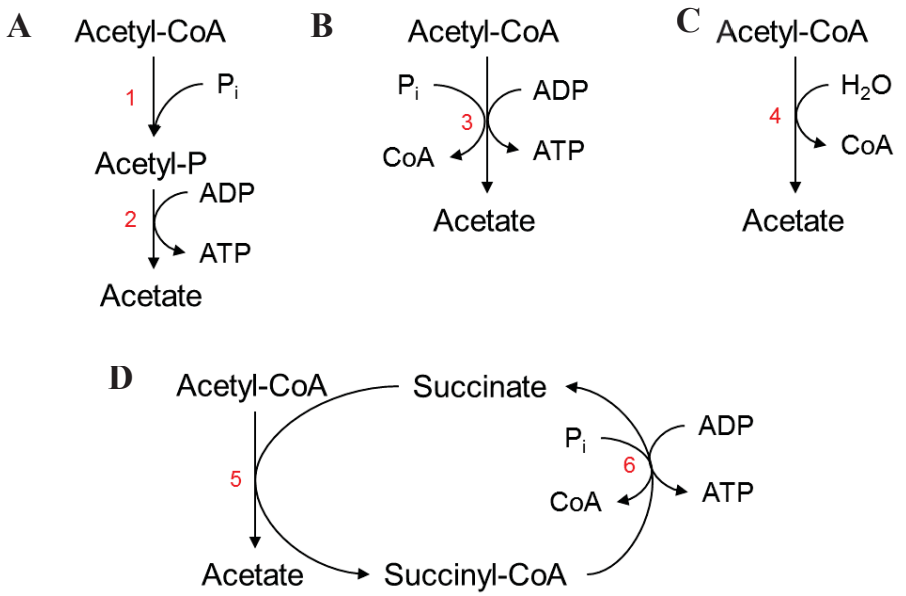


Figure 6.5 Possible acetate formation pathways from acetyl-CoA. A) Two step conversion of acetyl-CoA to acetate involving (1) phosphate acetyl transferase and (2) acetate kinase. B) Acetyl-CoA conversion to acetate using (3) acetyl-CoA synthetase. C) Acetyl-CoA hydrolysis to acetate using (4) acetyl-CoA hydrolase. D) Acetyl-CoA conversion to acetate using (5) succinate CoA-transferase and (6) succinyl-CoA synthetase.

```
                                    GxGGxxD
Saccharomyces cerevisiae ACH1   384 NVNGSRMLNGLGGSADFLRNAKLSIMHAPSARPTKVDPTGISTIVPMASHVDQTEHDLDI
    Clostridium kluyveri Cat1   412 HVMGSKMMNGIGGSGDFARNAYLTIFTTESI----AKKGDISSIVPMVSHVDHTEHDVMV
      Selenomonas ruminantium   372 HITGTKMMNGIGGSGDFARNAYLTIFYTPSI----AKGGKISAVVPMCSHIDHTEHDVDI
                  SR3Q1_778     373 HVTGTKMMNGIGGSGDFARNAYLTIFYTPST----AKGGKISSIVPFCSHIDHTEHDVDI
                 SR2Q5_1058     373 HVTGTKMMNGIGGSGDFARNAYLTIFYTPST----AKGGKISSVVPMCSHIDHTEHDVDV
                  SR1Q7_972     373 HVTGTKMMNGIGGSGDFARNAYLTIFYTPST----AKGGKISSVVPFCSHIDHTEHDVDV
                  SR1Q5_482     373 HVTGTKMMNGIGGSGDFARNAYLTIFYTPST----AKGGKISSVVPFCSHIDHTEHDVDI

Saccharomyces cerevisiae ACH1   444 LVTDQGLADLRGLSPKERAREIINKCAHPDYQALLTDYLDRAEHYAKKHNCLHEPHMLKN
    Clostridium kluyveri Cat1   468 IVTEQGVADLRGLSPREKAVAIIENCVHPDYKDMLMEYFEEACKSSGGN-T---PHNLEK
      Selenomonas ruminantium   428 IITEQGIADLRGKAPRERALEIINNCAHPDYRPILLDYFERATEATHHANT---PHILEE
                  SR3Q1_778     429 IISEYGIADLRGLEPRTRALEVINKCAHPDYRPMLLDYYERALAATKKAAT---PHLLHE
                 SR2Q5_1058     429 IISEYGIADLRGLEPRRRALEVINKCAHPDYRPLLLDYYERALAATKKAAT---PHLLNE
                  SR1Q7_972     429 IISEYGIADLRGLEPRTRALEVINKCAHPDYRPMLLDYYERALAATKKAAT---PHLLHE
                  SR1Q5_482     429 IISEYGIADLRGLEPRTRALEVINKCAHPDYRPMLLDYYERALAATKKAAT---PHLLHE
                                    ExG
```

Figure 6.6 Identification of succinate CoA-transferase in the *Quinella* genome bins. Two conserved motifs (ExG and GxGGxxD) of family 1 CoA-transferases (Tielens et al., 2010) were also found in the likely succinate CoA-transferases of *Quinella* and *Selenomonas ruminantium*. Mutation of glutamate (**E**471 in the *Clostridium kluyveri* protein) to aspartate (**D**447 in the *Saccharomyces cerevisiae* protein) results in the conversion of CoA-transferase to CoA-hydrolase (Mack and Buckel, 1997), and this served as an additional key feature to differentiate between these two enzymes.

6.2.4.3 Possibility of formate, butyrate and ethanol formation by *Quinella*

All four *Quinella* genome bins were searched for genes that code for enzymes involved in formate, butyrate and ethanol formation (Figure 6.7). Pyruvate dehydrogenase catalyses the conversion of pyruvate and CoA to acetyl-CoA and formate, but in the absence of oxygen this reaction is catalysed by pyruvate formate lyase (Knappe et al., 1974; Crable et al., 2011). These enzymes were searched for in all four *Quinella* genome bins, but none appeared to have either of these enzymes. A gene (SR3Q1_647) coding for pyruvate formate lyase-activating enzyme (PFL-AE) was present in *Quinella* genome bin SR3Q1, but its absence in other bins suggested that it might be a part of contaminated sequence. Even if that is not the case, on its own it does not consitute a functional pyruvate formate lyase system. This finding suggests that the *Quinella* might not be able to produce formate. Additionally, it was interesting to find the genes for all three subunits of formate dehydrogenase (an enzyme that catalyses the oxidation of formate to CO_2) in *Quinella* genome bin SR3Q1. Furthermore, a gene possibly coding for a formate dehydrogenase accessory protein that is required to assemble formate dehydrogenase was found in genome bins SR1Q7 and SR2Q5 but none of the subunits of formate dehydrogenase were found in these genome bins. So, the species represented by *Quinella* genome bin SR3Q1 might be able to produce formate in some way, or the contig containing the formate dehydrogense and related genes was a contaminated contig in this bin.

Acetaldehyde dehydrogenase and NAD-dependent alcohol dehydrogenase are needed to produce ethanol (Brown and Patterson, 1973). All four *Quinella* genome bins were searched for these enzymes, first using the automated GAMOLA2 search tool and then by BLAST-based searching of the reviewed ethanol dehydrogenases amino acid sequence of *Streptococcus mutans* against *Quinella* genome bins. Even though GAMOLA2 assigned gene 413 of genome bin SR1Q7 as an ethanol dehydrogenase, the BLAST search using the ethanol dehydrogenase of *Streptococcus mutans* suggested an absence of this gene in all *Quinella* genome bins. Gene 413 from *Quinella* genome bin SR1Q7 was most similar (27.8 to 37.7 % identity) to a lactaldehyde reductase from *Selenomonas bovis*, based on BLAST, and not known ethanol dehydrogenases. This suggests that *Quinella* may not able to produce ethanol.

A series of enzymes is needed for butyrate formation. Acetyl-CoA is first converted to butyryl-CoA in a four-step process catalysed by acetyl-CoA acetyltransferase, hydroxybutyryl-CoA dehydrogenase, 3-hydroxybutyryl-CoA dehydratase and

butyryl-CoA dehydrogenase. Butyryl-CoA is then converted to butyryl-phosphate by phosphotransbutyrylase and then finally to butyrate by butyrate kinase (Asanuma et al., 2005; Paillard et al., 2007). None of these enzymes were found in any of the *Quinella* genome bins, indicating that *Quinella* does not produce butyrate.

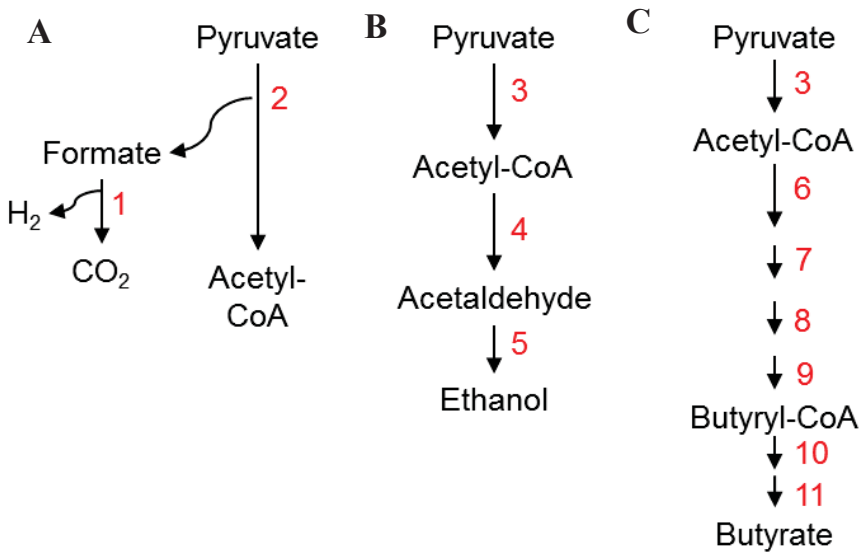


Figure 6.7 Enzymes involved in formate, butyrate and ethanol pathway. A) Formate formation and metabolism: (1) formate dehydrogenase, (2) pyruvate formate lyase. B) Ethanol formation: (3) pyruvate:ferredoxin (or pyruvate:flavodoxin) oxidoreductase, (4) acetaldehyde dehydrogenase, (5) alcohol dehydrogenase. C) Butyrate formation: (3) pyruvate:ferredoxin (or pyruvate:flavodoxin) oxidoreductase, (6) acetyl-CoA acetyltransferase, (7) hydroxybutyryl-CoA dehydrogenase, (8) 3-hydroxybutyryl-CoA dehydrogenase, (9) butyryl-CoA dehydrogenase, (10) phosphotransbutyrylase and (11) butyrate kinase.

6.2.5 Detailed analysis of key enzymes involved in propionate formation

6.2.5.1 Methylmalonyl-CoA decarboxylase and oxaloacetate decarboxylase

Methylmalonyl-CoA decarboxylase (MMCD) and oxaloacetate decarboxylase (OACD) are membrane-bound class II decarboxylases that catalyse Na^+ transport from the cytoplasm to the periplasm or cell exterior (Samols et al., 1988). The Na^+ gradient that is generated can be used to form ATP (Dimroth et al., 1998). In this section, both of these enzymes will be discussed together. MMCD is composed of four (Hilpert and Dimroth, 1983; Bott et al., 1997) or five subunits (Huder and Dimroth, 1993): alpha (α or MmdA), beta (β or MmdB), gamma (γ or MmdC), delta (δ or MmdD) and epsilon (ε or MmdE). In contrast, OACD only contains three subunits: alpha (α or OadA), beta (β or OadB) and gamma (γ or OadC) (Hilpert and Dimroth, 1983). Initial BLAST analysis of inferred amino acid sequences from all four *Quinella* genome bins suggested that *Quinella* has all the genes that code for the α, β, γ and δ subunits of MMCD but only the α subunit of OACD. Members of the genus *Veillonella* have been extensively studied to characterise MMCD and its subunits (Hilpert and Dimroth, 1983, 1991; Denger and Schink, 1992; Huder and Dimroth, 1993; Huder and Dimroth, 1995), whereas *Klebsiella pneumoniae*, *Salmonella typhimurium* and *Vibrio cholerae* have been used for OACD characterisation (Schwarz et al., 1988; Laussermair et al., 1989; Woehlke et al., 1992; Jockel et al., 1999; Dahinden et al., 2005; Studer et al., 2007) . Only one crystal structure of MMCD is available. This is from *Escherichia coli* (Benning et al., 2000), but this reference is of limited use for making generalisations because of the low sequence similarity with the majority of reviewed MMCD. There is no complete crystal structure available for the OACD complex, but an X-ray crystal structure of the carboxyltransferase domain (part of the α subunit) from *Vibrio cholerae* is available and has been studied in detail (Studer et al., 2007).

6.2.5.1.1 Detailed analysis of *Quinella* MMCD and OACD subunits

All four *Quinella* genome bins appeared to contain genes coding for MMCD of the four subunit type rather than the five subunit type. These subunits were identified by BLAST analysis of the *Quinella* genomes against several databases (nr database, pfam and Tigerfam) using GAMOLA2 as a search tool, and then confirmed by BLAST searching of the *Veillonella parvula* MMCD subunits (Gronow et al., 2010) against the complete inferred amino acid sequences of all four *Quinella* genome bins. When the same method

was applied using *Klebsiella pneumoniae* OACD subunits sequences as references, only α subunits of OACD were found. BLAST searches for the β subunit were unsuccessful. The highest identity sequences found were the MMCD β subunit genes. Similarly, no OACD γ subunit sequences were found in any of the *Quinella* genome bins.

Each subunit of MMCD and OACD from the *Quinella* genome bins was analysed separately to identify conserved amino acid residues and to understand key features. For MMCD, homologues from *Veillonella parvula*, *Propionigenium modestum*, *Anaerolinea thermophila*, *Thermococcus gammatolerans*, and *Pyrococcus abyssi* were used as reference amino acid sequences. For OACD, homologues from *Klebsiella pneumoniae*, *Vibrio cholerae, Salmonella typhimurium,* and *Klebsiella variicola* were used. All subunits sequences were aligned and phylogenetic trees were constructed.

The inferred amino acid sequence lengths of the α subunit of MMCD and OACD are different and the proteins possess different functions. MmdA sequences of *Quinella* were 58.2 to 79.0% similar to reference MmdA sequences and less than 11% identical to OadA references. This strongly supports their identification as α subunits of MMCD and not of OACD. In contrast, amino acid sequences of OadA from the *Quinella* genome bins were 40.2 to 42.4% identical to the reference OadA sequences and less than 12% identical to MmdA reference sequences, again confirming their likely classification. Furthermore, in a phylogenetic analysis, the *Quinella* decarboxylase α subunits (MmdA and OadA types) sequences grouped with their respective references sequences, with greater than 99% bootstrap support (Figure 6.8). This indicated that all four *Quinella* genome bins contained genes coding for both OadA and MmdA proteins.

The α subunit comprises of three domains (Figure 6.9): a *N*-terminus domain of 450 amino acid residues and possessing the carboxyltransferase catalytic site, a *C*-terminus domain of 70 amino acid residues which contains a conserved biotin-binding lysine (35 residues upstream of the *C*-terminus), and an association domain made up of a total of 40 amino acids of both the *N*- and *C*-termini and that is necessary for binding the α subunit to the *C*-terminus of the γ subunit (Woehlke et al., 1992; Dahinden et al., 2005; Studer et al., 2007). It was interesting to find that OadA from the *Quinella* genome bins only contained the *N*-terminus domain and that the biotin-binding and association domains were missing when compared with reference sequences (Figure 6.9). So, even though OadA was present in *Quinella* genomes, it may not function as a normal OadA, since it only appeared to be able to catalyse the carboxyltransferase reaction.

In contrast, the *Quinella* genomes contained genes coding for full-length (around 509 amino acid residues) sequences of the α subunit of MMCD. However, unlike normal OACD, α subunits of MMCD are only involved in catalysing the carboxyltransferase activity (Hoffmann et al., 1989) and do not contain association and biotin-binding domains. In MMCD, the biotin-binding domain necessary for function is part of the γ subunit (MmdC). Because of the unavailability of structures, it was not possible to identify key amino acid residues of the α subunit of MMCD. Furthermore, sequence similarity between carboxyltransferase domain of OadA and MmdA is poor (Huder and Dimroth, 1993), which restricts using OadA as references to predict key amino acid residues in MmdA.

The β subunits of MMCD and OACD are highly hydrophobic, integral membrane proteins (Woehlke et al., 1992), and contain binding sites for Na^+ translocation across the cell membrane (Dimroth and Thomer, 1983; Hoffmann et al., 1989). Beta subunit amino acid sequences of MMCD and OACD are >50% similar to each other and are argued to serve the same function (Huder and Dimroth, 1993). Genes encoding for β subunits of membrane-bound decarboxylases were present in all four *Quinella* genome bins, but *Quinella* bin SR1Q5 contained two copies. When the β subunit genes of bin SR1Q5 were inspected in detail, it was found that each of these were present at the end of a different contig. The overlapping regions of these sequences were 100% identical to each other, so these genes were assembled together and the consensus sequence was used as one MmdB sequence (SR1Q5_307/1610). When the putative β subunit amino acid sequences (Appendix Table A6.2) from the *Quinella* genome bins were aligned with reference β subunit sequences of MMCD (MmdB) and OACD (OadB), it was found that all sequences were >43% similar to each other. However, the *Quinella* β subunit amino acid sequences were more similar to MmdB (>76% identity with *V. parvula* MmdB) than to OadB (<51% similar to *K. pneumonia* and *S. typhimurium* OadB). Furthermore, a region of approximately 57 amino acid residues is found in all OadB sequences downstream from the *N* terminus (Huder and Dimroth, 1995), but this region was absent from the *Quinella* β subunit amino acid sequences (Figure 6.10A). So, based on the high percentage similarities and the absence of these extra amino acids, it was concluded that the β subunits present in *Quinella* genomes are MmdB.

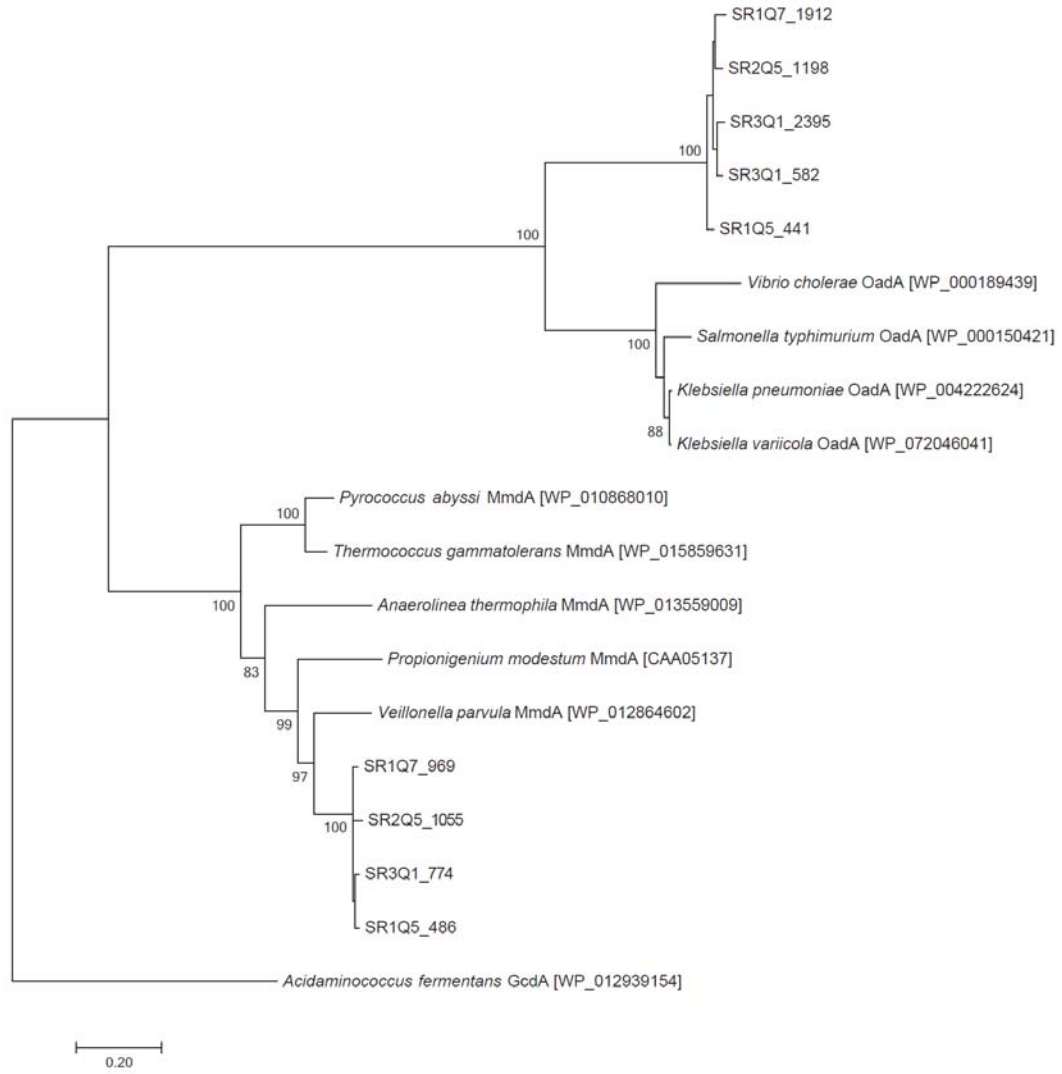


Figure 6.8 Phylogenetic tree of α subunits of MMCD and OACD, based on amino acid sequences. The *Acidaminococcus fermentans* glutaconyl-CoA decarboxylase α subunit sequence (GcdA) was used as an out group. The tree was generated using the Neighbor-joining method with the Jukes-Cantor genetic distance model. The scale bar represents 0.2 changes per amino acid residue. The numbers at the nodes are the percentage of trees that conserved that node in 500 bootstrap resamplings. GenBank accession numbers are given after each sequence name.

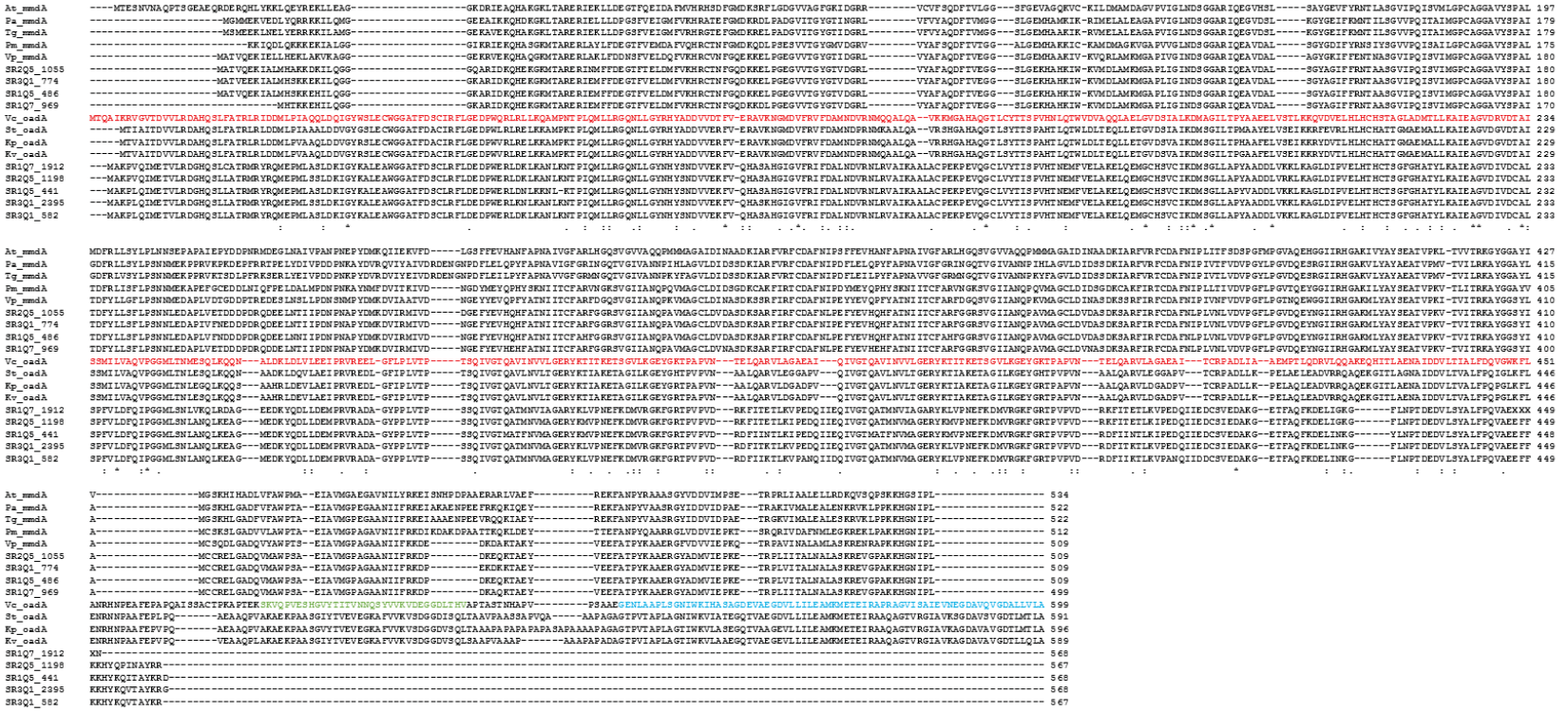


Figure 6.9 Alignment of alpha subunits (MmdA, OadA) of MMCD and OACD of *Quinella* with reference sequences. The MmdA references are from *Anaerolinea thermophila* (At), *Pyrococcus abyssi* (Pa), *Thermococcus gammatolerans* (Tg), *Propionigenium modestum* (Pm), and *Veillonella parvula* (Vp), and the OadA reference sequences are from *Klebsiella pneumoniae* (Kp), *Vibrio cholerae* (Vc) *Salmonella typhimurium* (St) and *Klebsiella variicola* (Kv). Three domains; *N*-terminus domain (red), association domain (green) and a *C*-terminus domain (light blue) containing a biotin-binding lysine residue (purple) were marked in the *Vibrio cholerae* sequence.

The β subunit membrane topology model of *K. pneumoniae* predicted by Jockel et al. (1999) was used as a reference to understand MmdB from the *Quinella* genomes. Topology analysis suggested that the β subunit contained 9 transmembrane regions (Figure 6.10B). The extended 57 amino acid residues of OadB are in the first cytoplasmic loop and may provide stability to the enzyme complex. In the *Quinella* β subunits this loop is smaller, like in other MmdB (Huder and Dimroth, 1993). Site-specific mutagenesis studies (Di Berardino and Dimroth, 1996; Jockel et al., 1999; Jockel et al., 2000b; Jockel et al., 2000a) on OadB of *Klebsiella pneumoniae* have uncovered some key amino acid residues and predicted their roles. These residues are also found in the closely-related MmdB (Jockel et al., 1999). Mutations in D203 (helix III) and S382 (helix VIII) lead to complete loss in decarboxylase and Na^+ transport activity of OACD, supporting their identification as Na^+ binding sites (Di Berardino and Dimroth, 1996; Jockel et al., 2000b). However S382 can be replaced by T382 or D382 without affecting the decarboxylase and Na^+ transport ability. Similarly, G377 (helix VIII) may be involved in a Na^+ conducting channel or serve as a contact site for other helices, and so cannot be replaced by any other residue. A change of N373 (helix VIII) to L373 or D373 slightly affects the decarboxylase activity. Furthermore, Y227 (helix IV) is an important residue for high catalytic activity of enzyme complex but changes to C227 or F227 can increase the catalytic activity. Y229 (helix IV) takes part in a proton conduction network which leads to decarboxylation of carboxybiotin. All of these important residues (D203, S382, G377, N373, Y227 and Y229) were found to be conserved in the putative MmdB of all four *Quinella* genome bins when aligned with OadB of *K. pneumoniae* (Figure 6.10A).

Genes coding for γ subunits of OACD were not found in any of the *Quinella* genome bins. In contrast, all four *Quinella* bins contained genes coding for full length γ subunits (MmdC) of MMCD, which serve almost the same function as the biotin-binding *C*-terminus domain of the α subunit of OACD. The γ subunits also play a major role in MMCD assembly, by binding to α and δ subunit of MMCD (Kumar et al., 1982). These γ subunit sequences were aligned with MmdC sequences of *V. parvula* and *P. modestum* and OadA *C*-terminus sequences of *K. pneumoniae* and *S. typhimurium.* The *Quinella* MmdC sequences were 54.7% to 59.7% similar to MmdC of *V. parvula* and 47.4% to 49.8% similar to that of *P. modestum*. They were only 31.7 % to 38.7 similar to the OadA *C*-terminus sequences. Like other proteins containing a biotin prosthetic group, MmdC of *Quinella* also contained a putative biotin-binding lysine residue (35 residues upstream

of the *C*-terminus) within a highly-conserved amino acid motif, LEAMKM (Bott et al., 1997) (Figure 6.11A). Furthermore, the alanine-proline linker (around residues 31 to 60) was also found in the *N*-terminus of all of *Quinella* MmdC. This linker provides flexible movement for the biotin group to move between the catalytic centres of different biotin-containing enzymes (Huder and Dimroth, 1993).

Genes coding for the δ subunit of MMCD were found in all four *Quinella* genome bins. The δ subunit of *Quinella* was found to be more similar to that of *V. parvula* (32.8% to 37.4% amino acid similarity) than to that of *P. modestum* (21.2% to 24.3%), but were only 14.8 to 18.7% identical to the γ subunit of OADC of *K. pneumoniae* and *S. typhimurium*. Even though very low sequence similarity was found between the δ subunits of MMCD and the γ subunits of OACD, they share similarities in their hydrophobicity patterns and the presence of identical alanine-proline linkers (Huder and Dimroth, 1995) (Figure 6.11B), and appear to have similar functions (Buckel, 2001). Importantly, the *N*-termini of both are predicted to be anchored in the membrane, indicating that the hydrophobic *N*-terminus could bind to the hydrophobic β subunit while the hydrophilic *C*-terminus could bind to the α subunit (Buckel, 2001). MmcD sequences lack the triple histidine that serves as a Zn^+ binding site at the *C*-terminus, suggesting that divalent metal ions are not present in biotin-containing carboxyltransferase enzymes that react with thioester (CoA-bound) substrates, in contrast to those that react with keto acid substrates (Towbin et al., 1979). The *Quinella* MmcD sequences also lack these triple histidine residues. In *V. parvula*, there was an extra subunit reported (epsilon), but in other bacteria with MMCD, like *P. modestum*, this subunit is absent, as it appears to be from all four *Quinella* genome bins. The function of this subunit is still unknown but its strong sequence similarity with the *C*-terminus of δ subunit indicated that its origin may be from a duplication of the δ subunit (Huder and Dimroth, 1993).

In summary, all four *Quinella* genome bins contain genes for the essential subunits needed to assemble Na^+-translocating methylmalonyl-CoA decarboxylases (MMCD). These have the α subunit (MmcA) containing the carboxyltransferase catalytic site, the integral membrane β subunit (MmcB) with important conserved amino acid residues involved in Na^+ translocation, the γ subunit (MmcC) containing a biotin-binding lysine residue, and the δ subunit (MmcD) with the alanine-proline linker. On the basis of available literature, a probable schematic of MMCD (Figure 6.12) may be drawn, where the hydrophobic integral membrane β subunit is connected to the hydrophobic *N*-terminus

δ subunit, and the highly hydrophilic *C*-terminus of the δ subunit is connected to the γ subunit.

Quinella does not appear to have all the genes needed for oxaloacetate decarboxylase (OADC), only containing the gene for the α (carboxyltransferase, OadA) subunit. While it can be speculated that the OadB and OadC subunits were missed in the sequencing, this seems unlikely as the four *Quinella* bins all have genes for OadA and all do not have genes for OadB or OadC for, unless the latter two were rejected at the assembly steps. It is also possible that OACD activity is conferred in a different way in *Quinella*. One possibility is that functional OACD in *Quinella* uses the OACD α (carboxyltransferase) subunit in a hybrid with other subunits from the MMCD complex: the β subunit (which is similar to the β subunit of OACD except for 52 extra amino acid residues in a cytoplasmic loop), the γ subunit and the δ subunit (which has a similar hydrophobicity to that of γ subunit of OACD but no sequence homology). This speculation is made appealing by the unusual structure of the α subunit of the OACD of *Quinella*. In contrast to other OadA, it lacks the biotin-containing domain. In MMCD, that function is found in the γ subunit. A normal OadA in conjunction with MmdC would contain two biotin-binding domains, so that loss of this domain from the *Quinella* OadA would render the interaction with the MMCD components possible (Figure 6.12). Here it is also speculated that the OACD would catalyse the conversion of pyruvate to oxaloacetate (Figure 6.1) instead of its well-known direction (oxaloacetate to pyruvate). However, even in the absence of a functional OACD (because of the lack of OACD specific β and γ subunits), *Quinella* could still form oxaloacetate from phosphoenolpyruvate by using phosphoenolpyruvate carboxykinase, and genes for this were present in all four genome bins (Figure 6.1, Appendix Table A6.2). This was confirmed by BLAST against the Swiss-Prot database (Boutet et al., 2016). The amino acid sequences of phosphoenolpyruvate carboxykinases in the *Quinella* genome bins were ≥70.0% similar to amino acid sequences of reviewed phosphoenolpyruvate carboxykinase (ATP-forming) of *Agathobacter rectalis*, *Eubacterium eligens*, *Parabacteroides distasonis and Bacteroides fragilis*. Using phosphoenolpyruvate carboxykinase to form oxaloacetate will result in formation of one ATP molecule. In contrast, the two-step conversion process will result in generation of one ATP molecule by pyruvate kinase during conversion of phosphoenolpyruvate to pyruvate, followed by the generation of oxaloacetate from pyruvate via the reversed OADC, which would be driven by one Na^+ from outside the

cell to the inside. This is equivalent to using approximately 3/10 ATP to carry out the carboxylation. Oxaloacetate decarboxylation (the reverse reaction) in *K. pneumoniae* extrudes one Na^+ (Dimroth et al., 2001), which then can be used to generate ATP via an ATP synthase that generates approximately 3 ATP per 10 Na^+ (Meier et al., 2011). The two-step process therefore yield approximately 7/10 ATP for the same conversion. It is unclear which option is used by *Quinella*.

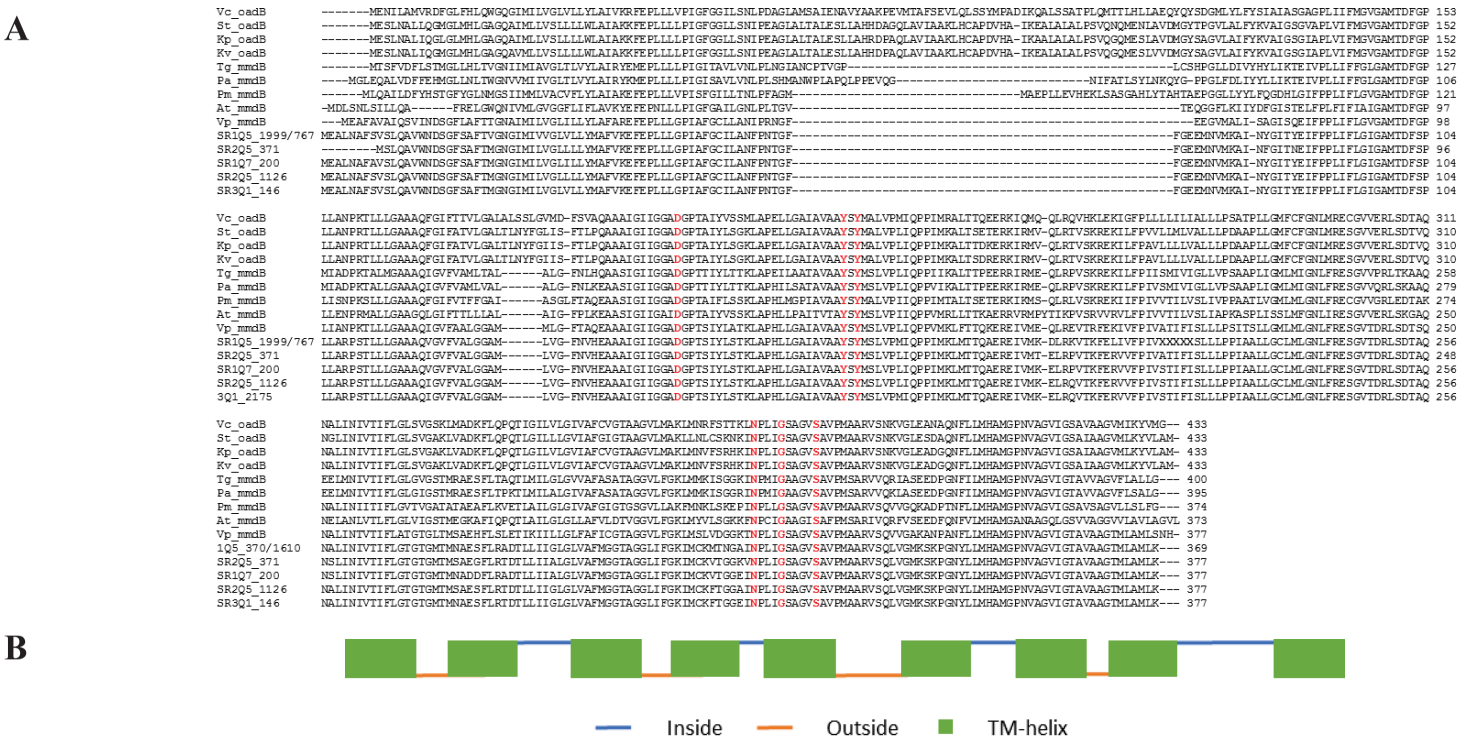


Figure 6.10 A) Alignment of beta subunits of MMCD (mmdB) and OACD (oadB) of *Quinella* with reference sequences. Beta subunit amino acid sequences of MMCD and OACD are reported to have >50% identity (Huder and Dimroth, 1993), similar to what was observed here (all aligned sequences are >43% similar to each other). Key amino acids (red) in the reviewed *Klebsiella pneumoniae* (Jockel et al., 1999; Jockel et al., 2000b; Jockel et al., 2000a) OadB were also found to be conserved in *Quinella* MmdB. The reference sequences use the same species name codes as in Figure 6.9. B) Schematic representation of the *Quinella* MmdB. Nine transmembrane regions were predicted using SPOCTOPUS. TM, transmembrane.

A

```
St_OadA C terminus EAAQPVAKAEKPAASGIYTVEVEGKAFVVKVSDGGDI--SQLTAAVP--------------AASSAPVQAAAP-----AGAGT-PVTAPLAGNIWKVIATEGQTVAEGDVLLILEAMKMETEIRAAQAGTVRGIAVKSGDAVSVGDTLMTLA 130
Kp_OadA C terminus EAAQPVAKAEKPAASGVYTVEVEGKAFVVKVSDGGDV--SQLTAAAP--------------APAPAPAPASAPAAAAPAGAGT-PVTAPLAGTIWKVLASEGQTVAAGEVLLILEAMKMETEIRAAQAGTVRGIAVKAGDAVAVGDTLMTLA 135
Pm_MmdC           M--KN------------FKVTVNGTEYDVAVEEMGGAAVASAPAARP---AAAPAPAAPKPAAAPAPAPAPKTTA-AGAGAGANTVTAPMPGTILNVGCHAGDKVSKGDTLVVLEAMKMENEIMAPHDGVVSEVRVQQGASVNAGDILVVLS 134
Vp_MmdC           M--KK------------FNVTVNGTAYDVEVNEVKAAAPAAAPKAAP---AAAPAPK---AAPAPAPAAAAAPVP-----AGAETVKAPMPGKILSVAVSAGQAVKKGETLLILEAMKMQNEIAAPHDAVVSEVRVSANQTVSTGDDMVVLG 127
SR1Q7_968         MATKK------------FNVTVNGTTYEVEVEEVKAA--GGAPKAAPAPKAAAPAPKAAAPA-APAPAPKAAAPAKVAAGAGEHSIDSPMPGKIVKLVAAEGAAVKAGDVLLILEAMKMQNEITADADGTVKKFNVAAGQSVKAHESLVILG 137
SR3Q1_773         MATKK------------FNVTVNGTTYEVEVEEVKAA--GGAPKAAP-----APAPKAAAPAPAPKAAPAPAAAPKVAAGAGEHSIDAPMPGKVIKLVAAEGAAVKAGDVLLILEAMKMQNEITADADGTVKKFNVAAGQSVKAHESLVILG 133
SR1Q5_488         MATKK------------FNVTVNGTTYEVEVEEVKAA--GGAPKAAP-----APKVA---PAPAPAPKAAPATAPKVAAGAGEHSIDAPMPGKVIKLVAAEGQAVKAGDVLLILEAMKMQNEITADADGTVKKFNVAAGQSVKAHESLVILG 130
SR2Q5_1055        MATKK------------FNVTVNGTTYEVEVEEVKAA--GSAPKAAP-----APAPK---AAPAPAPKAAPAAAPKVAAGAGEHSIDAPMPGKVVKLVAAEGQAVKAGDVLLILEAMKMQNEITADADGTVKKFNVAAGQSVKAHESMVILG 130
```

B

```
Kp_oadC            M-------------TDNAVLLGEGFT----LMCLGMGFVLVFLLLLIFAIRGMSLA-------VNRLFP---EP------PAAPKPAPAAVAPA-DD------FARLKPAIVAA------------IHH------------------HRRLHP 83
St_oadC            M-------------TNAALLLGEGFT----LMFLGMGFVLAFLFLLIFAIRGMSAA-------VNRFFP---EP------APAPKAAPAAAAPVVDD------FTRLKPVIAAA------------IHH------------------HHRLNA 84
Pm_mmdD            MNITELMELFSNPETIKTLETGDLMTGIGVTVVLGMGITVVALIFLMYIIGGMAAIMAEKPKEVKETAAAAPKPEAAA--------APA---PAANN--DEELVAVIAAAVAAQLGTSASNLIIRNVTRSLDTTPAWGRAGIVDQMATRL--- 137
Vp_mmdD            ME--------GQAVTTNPW----------LIMAINMTVVFAVLIALGILMEIVHLIDPTKKKKE---APAATAPVATPTA------TPVAPANASAQN-EDEVVAAIVGAIVA-MGYSSEQIASIRPTA--TSAK-WRLEGRL---SGRG--- 115
SR1Q7_201          MS--------HSPVTTNPL----------IIMMINMTVVFIVLVVLMYLIKLIHIIDPTKPREVPK-SDDEDLPLVAVKSEPSTATPPPEPEPVT-QGVSPEIIAAISAAIAA-YGFAGQVKAVRIINH--EGTP-WRASARI-NTFSRK--- 125
SR2Q5_370          MG--------HSPVTTNPL----------IIMMINMTVVFIVLVVLMYLIKLIHIVDPTKPKEEPAAAPVDDEPLVAVKSA-----APE---PVLDAGPTPEIIAAISAAITA-YGFVGQVKAVRIINH--EGTP-WRSSARF-NTLKRK--- 119
SR3Q1_145          ME--------HSPVTTNPL----------IIMMINMTVVFIVLVSLMFLIKLIHIVDPTKEPEVKE-APADDEPLVAVKS------EPAAIAPVVEEGIKPEVIAAISAAIAA-YGFAGQVKAVHIINH--EGTP-WSNAMRNAQGVMRNGRR 124
SR1Q5_2000         ME--------HSPVTTNPL----------IIMMINMTVVFIVLIVLMYLIKLIHFLDPTKEPEAKE-APEEEEPLVAVKSATASAPEPV---PVVEEGIKPEVIAAITAAIAA-YGFAGQVKAVHIINH--EGTP-WRAAAKFNNNLQRK--- 124
SR2Q5_1127         ME--------HSPVTTNPL----------IIMLINMTVVFLVLVSLIYLIKLIHFVDPTKEKPKPE--VEEDEPLVAVKSAPAAS-APA---PVVEEGINSEVIAAISAAIAA-YGFAGQVKAVHILNH--EGTP-WRAAAKF-NNFPRK--- 121
```

Figure 6.11 Gamma and delta subunits. A) Alignment of MmdC of MMCD and the *C*-termini of OadA of OACD. The conserved lysine residue in the LEAMKM motif (red) is implicated in biotin binding (Huder and Dimroth, 1993; Bott et al., 1997), and was conserved in the *Quinella* sequences. B) Alignment of MmdD of MMCD and OadC of OACD. The alanine-proline linker region (purple) in *Klebsiella pneumoniae* (Kp) and *Salmonella typhimurium* (St) was present in MmdC with different degrees of similarity.

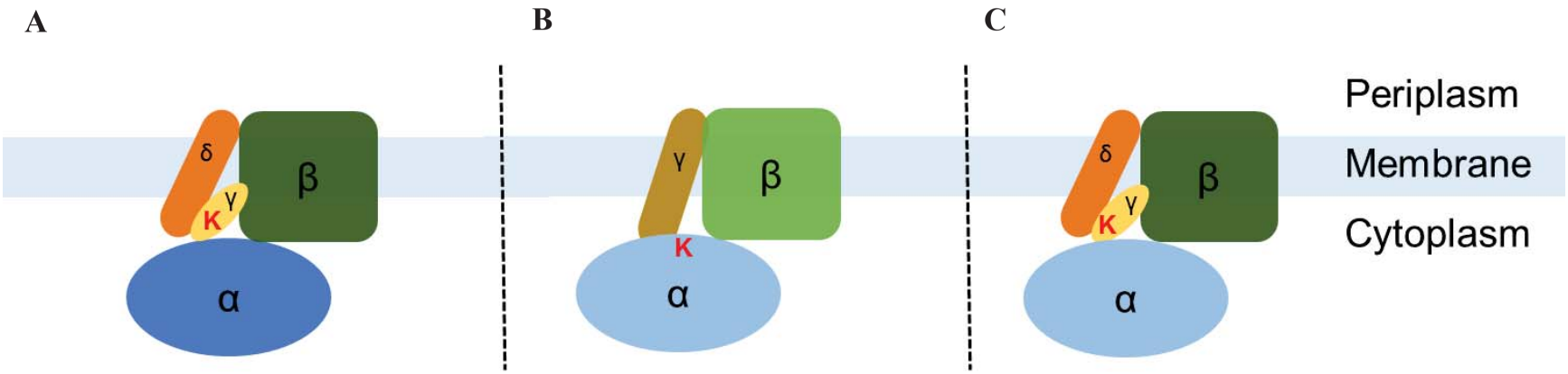


Figure 6.12 MMDC and OADC structures (adapted from (Klimchuk et al., 2016). A) MMCD structure in *Quinella* (this study) and other bacteria (Buckel, 2001). B) Normal OADC structure (Buckel, 2001). C) Speculative functional OADC structure in *Quinella* (this study) using the α subunit of the OACD in a hybrid with the β, γ and δ subunits of MMDC. The K represents the putative biotin-binding lysine residue.

6.2.5.2 Fumarate reductase

Fumarate reductase (quinol:fumarate reductase; QFR) catalyses the reduction of fumarate to succinate coupled to the oxidation of quinol to quinone (Hägerhäll and Hederstedt, 1996). This reaction can also be reversed, coupling quinone reduction to quinol with succinate oxidation to fumarate by succinate dehydrogenase (succinate:quinone reductase; SQR) (Saraste, 1999). These enzymes are structurally and mechanistically similar with very high sequence similarity and cofactor composition, and belong to the succinate:quinone oxidoreductase (SQOR) superfamily (Iverson et al., 1999; Lancaster et al., 1999). QFR is involved in anaerobic respiration (Kröger, 1978; Kröger et al., 1992) whereas SQR participates in aerobic respiration (Saraste, 1999). Under certain conditions they can functionally replace each other (Maklashina et al., 1998). The SQORs are mainly composed of four subunits: two hydrophilic subunits, A and B, and two hydrophobic integral membrane subunits, C and D. In some instances, instead of two, only one comparatively large hydrophobic (Hägerhäll and Hederstedt, 1996) integral membrane subunit is found, which is predicted to have evolved from fusion of the C and D subunits (Hägerhäll and Hederstedt, 1996; Lancaster et al., 1999). SQORs can be classified into five different types based on their hydrophobic domains and heme *b* content (Hägerhäll and Hederstedt, 1996; Lancaster et al., 1999). Type A contain two hydrophobic subunits and two heme, type B one hydrophobic subunit and two heme, type C two hydrophobic subunits and one heme, type D two hydrophobic subunits and no heme, while type E contain two hydrophobic subunits which are very different from those of the other types and no heme (Lancaster, 2003). All four *Quinella* genome bins appeared to have the genes coding for three subunits of QFR (Appendix Table A6.2). These were the two hydrophilic subunits A and B, and one hydrophobic subunit C. These subunits were identified by BLAST analysis of the *Quinella* genomes against several databases (nr database, pfam and Tigerfam) using GAMOLA2 as a search tool, and then confirmed by BLAST of the sequences of the SQOR subunits from *Escherichia coli* and *Wolinella succinogenes* against the complete inferred amino acid sequences of all four *Quinella* genomes. Subunit D was not found by either of these methods, suggesting that *Quinella* SQORs were type B (the only type with one hydrophobic subunit) (Lancaster, 2003).Type B contains SQR from *Bacillus subtilis* and *Paenibacillus macerans* and QFR from *Campylobacter jejuni* and *Helicobacter pylori* and *Wolinella succinogenes*.

6.2.5.2.1 Detailed analysis of *Quinella* QFR

Two crystal structures of QFR are available, one from *Escherichia coli* (Iverson et al., 1999) and another from *Wolinella succinogenes* (Lancaster et al., 1999). These QFR amino acid sequences were used as references to understand the possible QFR from the *Quinella* genomes. All subunit sequences from *Quinella* were aligned with the reference amino acid sequences to identify conserved key amino acid residues.

The hydrophilic flavoprotein A subunits from all four *Quinella* genomes were 31.4% to 31.1% similar to the SQOR subunit of *E. coli* and around 27.9% to 28.5% similar to that from *W. succinogenes*. Fumarate binding sites are present between the FAD-binding domain and capping domain (among four domains) of subunit A (Lancaster, 2003). FAD-binding residue H45 (Iverson et al., 1999; Lancaster et al., 1999) of *Quinella* and *E. coli* and H43 of *W. succinogenes* were conserved in all sequences. Similarly, other important amino acid residues with key functions were also found to be conserved in the predicted *Quinella* enzymes, such as G52 (which accepts a hydrogen bond from FAD), a 'HPT triad' (H233, P234 and T235), and R404 and A409 (in *W. succinogenes*) involved in proton movement. Furthermore, a dicarboxylate-binding site containing two arginine (R301 and R404) and one histidine (H369) found in *W. succinogenes* was also found in the enzymes predicted from the *Quinella* genomes (Figure 6.13). However, it was interesting to find that S45 was conserved in all four *Quinella* enzymes and apparently homologous to S44 in *Wolinella succinogenes* (one of the 11 residue that bind to FAD). This was replaced by threonine (T45 in *E. coli*) in most of the SQORs (Lancaster et al., 1999). Based on these similarities, it appears that the *Quinella* genomes code for subunit A of a QFR.

The second hydrophilic subunit, subunit B, consists of two domains (a *N*-terminal 'plant ferredoxin domain' and a *C*-terminal 'bacterial ferredoxin domain') and three Fe-S clusters (Iverson et al., 1999). Three of the four *Quinella* genome bins contained full length amino acid sequences of subunit B. The fourth, *Quinella* genome bin SR1Q7, contained a sequence coding for only the first 35 residues of subunit B but it was 100% identical to the same region of the subunit B sequences in the other *Quinella* genome bins amino acid sequences, suggested that that strain also contained a QFR subunit B. This short fragment was at the end of a contig, and so is probably not a gene remnant. The amino acid sequences of the *Quinella* subunit B were around 25% similar to the homologous sequences from *W. succinogenes* and *E. coli*. Fumarate reductase contains

three Fe-S clusters (2Fe-2S, 3Fe-4S and 4Fe-4S) which are coordinated with cysteine residues (Iverson et al., 1999; Lancaster et al., 1999). The *Quinella* subunit B sequences contained four conserved cysteine residues that coordinate with 2Fe-2S clusters that are in contact with plant ferredoxin domain and subunit A (Lancaster et al., 1999). The 3Fe-4S cluster coordinates with three cysteine residues whereas the 4Fe-4S is coordinated with four cysteine residues (Iverson et al., 1999; Lancaster et al., 1999), and all of these were conserved in *Quinella* (Figure 6.13). However, like *W. succinogenes*, *Quinella* contains an extra cysteine residue (C158 in *Quinella*) which is replaced by leucine in *E. coli* (Figure 6.13B). The 3Fe-4S and 4Fe-4S are bound by the bacterial ferredoxin domain, which is in contact with the hydrophobic subunit(s) (Iverson et al., 1999; Lancaster et al., 1999).

Genes coding for only one hydrophobic subunit of QFR were found in the *Quinella* genome bins. Sequence alignment of this subunit C with the subunit C from the *E. coli* SQOR (which also contains only one hydrophobic subunit) showed <15.6% identity, suggesting that it is quite different to subunit from the one in *E. coli*. It was also <17.8% similar to the subunit C from *W. succinogenes*. Therefore, subunit C sequences from *Bacillus subtilis*, *Paenibacillus macerans*, *Campylobacter jejuni* and *Helicobacter pylori*, which are B type SQOR (Hägerhäll and Hederstedt, 1996; Hederstedt, 1999; Lancaster, 2003), were included in the analysis. The *Quinella* sequences were 33.8% to 39.3% similar to those of *B. subtilis* and *P. macerans* and only around 15% similar to those of *C. jejuni* and *H. pylori* (Table 6.3), suggesting that *Quinella* QFR does belongs to type B and is more similar to those from other Gram positive bacteria (*B. subtilis* and *P. macerans*) (Table 6.3). Similarly, transmembrane topology analysis using SPOCTOPUS predicted that subunit C of *Quinella* is composed of five membrane spanning helices (Figure 6.14), a characteristic feature of QFR with only one hydrophobic subunit (Lancaster et al., 1999). In contrast, subunits C and D each contain three membrane spanning helices (Iverson et al., 1999) in those fumarate reductases with two hydrophobic subunits.

Histidine residues (H27, H71, H114 and H158 in *Quinella*) that are the axial ligands for proximal and distal hemes (Lancaster et al., 1999) were also conserved in all four predicted *Quinella* proteins (Figure 6.14), suggesting that *Quinella* QFR subunit C binds two heme. This is a further indicator of a type B SQOR (Lancaster, 2003). An essential glutamic acid residue (E66 in *W. succinogenes*) is involved in menaquinol oxidation

(Lancaster and Kröger, 2000) and was also conserved in all four *Quinella* genomes. In summary, all QFR subunits found in the *Quinella* genome bins contained conserved key amino acid residues that match with QFR of type B SQOR (*W. succinogenes*), suggesting that *Quinella* has a fully functional QFR. This suggests that the enzyme functions to reduce fumarate to succinate, rather than to oxidise succinate to fumarate.

Table 6.3 Amino acid sequence similarities of fumarate reductase subunit C of *Quinella* genome bins, and other fumarate reductase subunit C proteins (all B type SQOR unless noted otherwise).

SQOR	Genes in *Quinella* genome bins			
	SR1Q5_1182	SR1Q7_2325	SR2Q5_1217	SR3Q1_1942
B. subtilis	33.8[a]	34.8	34.8	33.8
P. macerans	38.9	37.9	36.0	39.3
C. jejuni	12.8	14.1	13.2	14.1
H. pylori	16.4	15.2	16.1	17.0
W. succinogenes	17.2	15.9	15.9	18.5
E. coli (type D SQOR)	14.9	14.9	14.0	14.0

[a]All values are represented as percentages.

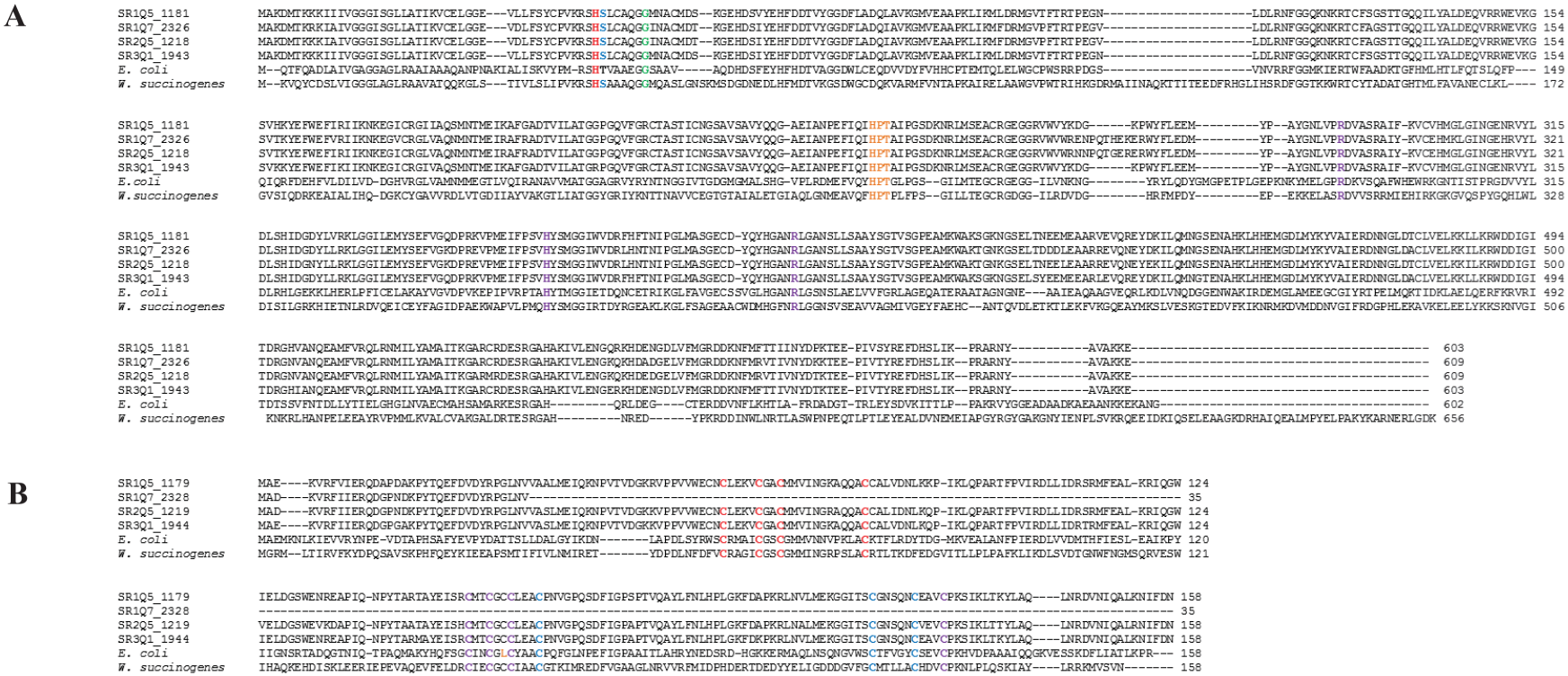


Figure 6.13 Alignments of amino acid sequences of hydrophilic subunits of fumarate reductase from *Quinella* with those from other bacteria. A) Subunit A, showing conserved FAD-binding histidine residues (red), glycines that accept an H-bond from FAD (green), the 'HPT triad' (orange), the dicarboxylate-binding site (purple) and serines that are one of the 11 residue that bind to FAD (blue). B) Subunit B, showing conserved cysteine residues that coordinate with 2Fe-2S clusters (red), with 3Fe-4S clusters (blue) or with 4Fe-4S clusters (purple). Like in *W. succinogenes*, and unlike in *E. coli*, cysteine (C154) is replaced by leucine.

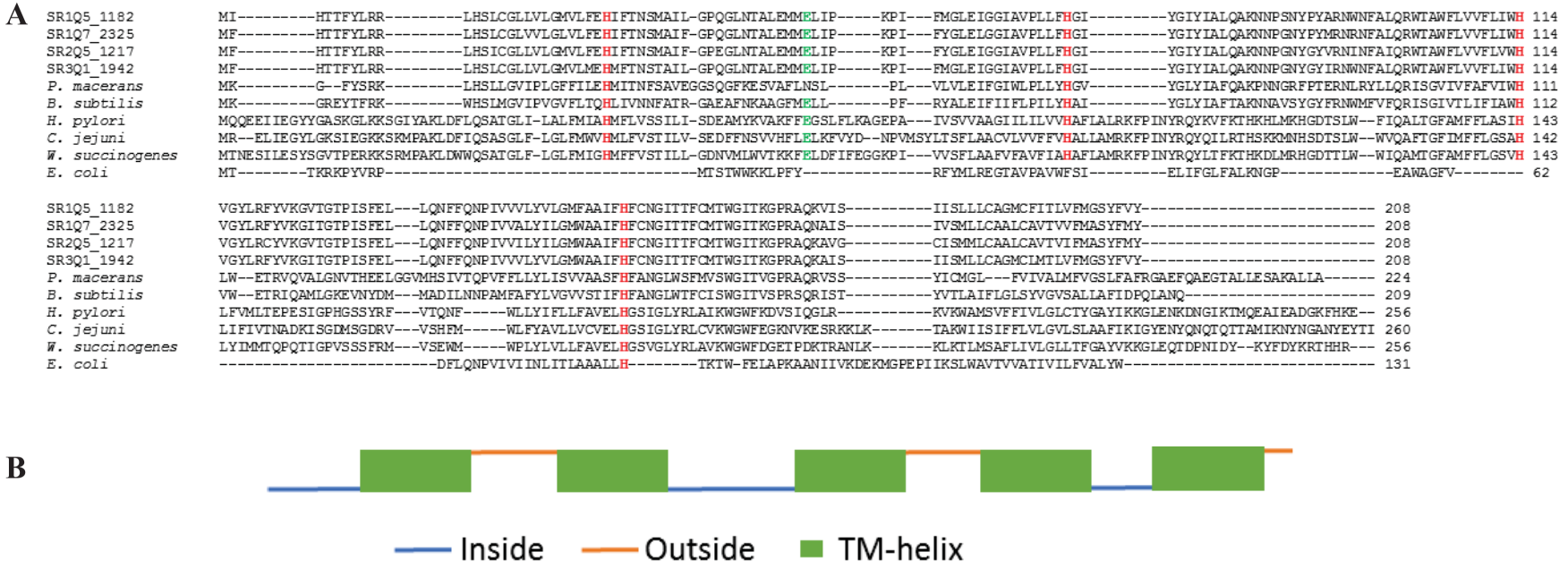


Figure 6.14 Subunit C of fumarate reductase. A) Alignment of amino acid sequences of *Quinella* subunit C with type B SQOR from *P. macerans*, *B. subtilis*, *H. pylori*, *C. jejuni*, *W. succinogenes*, and *E. coli*. Conserved histidine residues (red) provide axial ligands for proximal and distal heme groups. The conserved glutamate (green) is an essential residue, involved in menaquinol oxidation. B) Schematic representation of the fumarate reductase, suggesting that *Quinella* subunit C is composed of five membrane-spanning domains. TM, transmembrane.

6.2.5.3 Hydrogenase (Ni-Fe)

Hydrogen plays a major role in rumen fermentation and regulates the amount of methane that is formed (Janssen, 2010). Some microbial groups use hydrogen (like methanogens and acetogens) as energy sources for growth, while many others produce hydrogen as a by-product (Schwartz et al., 2013). The ability to take up or produce hydrogen is dependent on hydrogenases, enzymes characterised by different co-factors. Initially, hydrogenases were categorised into three major classes (NiFe, FeFe and Fe type) based on the H_2-binding metal centres at the active site (Volbeda et al., 1995; Peters, 1998; Shima et al., 2008). These three classes of hydrogenase have recently (Greening et al., 2016) been further sub-divided into four groups (22 subgroups) of NiFe-type hydrogenases and three groups (6 subgroups) of FeFe-type hydrogenases, whereas there appears to be only one group of Fe-type hydrogenases. These groupings are based on amino acid sequence-based phylogeny, the presence of metal-binding motifs, genetic organisation and reported biochemical characteristics.

6.2.5.3.1 Identification and classification of hydrogenases in *Quinella* genomes

BLAST-based analyses of all four *Quinella* genome bins against different databases (section 2.18) revealed that all of them contained complete sets of genes that code for the large subunit, the small subunit and *b*-type cytochrome of hydrogenases (Appendix Table A6.2). BLAST analyses of these *Quinella* hydrogenases against the hydrogenase database of Greening et al. (2016) suggested that the hydrogenases in all four genome bins were closely related to the hydrogenase of *Selenomonas ruminantium* (NCBI accession no. WP_029541750), which was classified as a NiFe-type oxygen-tolerant membrane-bound H_2-uptake hydrogenase (Greening et al., 2016).

The large subunits of the three major classes of hydrogenases contain different H_2-binding metal centres: L1 and L2 motifs for [NiFe]-hydrogenases(Vignais and Billoud, 2007); P1, P2 and P3 motifs for [FeFe]-hydrogenases (Vignais and Billoud, 2007); and Cys176 in [Fe]-hydrogenases (Shima et al., 2008). L1 and L2 motifs (Figure 6.15) surrounding the metal-ligating cysteine residues were found in the amino acid sequences of all four *Quinella* hydrogenases, which suggested that *Quinella* has [NiFe]-hydrogenases. However, the variation in consensus motifs meant that further subgrouping was not clear. The L1 motif (xxRICGVCTxxH) was identical to that of subgroup 1d (oxygen-tolerant) type uptake hydrogenases, but the L2 motif (SFDPCxACxxH) was identical to that of

subgroup 1e (possible bidirectional hydrogenases) based on the classification of Greening et al. (2016). A phylogenetic analysis was performed using the amino acid sequences of the large subunit of all 1d and 1e type hydrogenase sequences in the Greening et al. (2016) database, using five sequences each from types 1a, 1b, 1c, 1f and 1h/5 (to cover all NiFe-hydrogenase type), and all four *Quinella* hydrogenase sequences. The *Quinella* hydrogenase sequences formed a stable cluster (100% bootstrap value) within a radiation that contained only hydrogenase sequences from group 1d, and in the same radiation as all other group 1d sequences (Figure 6.15). Most of the sequences that clustered with them originated from members of the class *Negativicutes. Quinella* is also a member of this class. Again, the large subunit hydrogenase sequence from *Selenomonas ruminantium* (Wp_029541750.1) was found to be the closest relative of the *Quinella* hydrogenases. Therefore, even though the L2 motif matched with group 1e hydrogenases, the *Quinella* hydrogenase sequences may still be classified as belonging to group 1d, as a NiFe-type membrane-bound H_2-uptake hydrogenase.

6.2.5.3.2 Detailed analyses of *Quinella* hydrogenase subunits

The initial analyses suggested that *Quinella* has a NiFe-type membrane-bound H_2-uptake hydrogenase that may be further categorised as an oxygen-tolerant type based on the classification proposed by Greening et al. (2016). However, the L2 motif was different to that expected for an oxygen-tolerant (1d) type. Recently, Flanagan and Parkin (2016) identified and compared the variation in amino acid sequences involved in binding the metal-cofactors and iron-sulfur (FeS) clusters of the small and large subunits of O_2-tolerant and O_2-sensitive NiFe type membrane-bound hydrogenases. They concluded that conserved cysteine residues play an important role by participating in stabilisation of the unusual proximal FeS cluster in O_2-tolerant and O_2-sensitive sensitive NiFe-type membrane-bound hydrogenases.

Three O_2-tolerant (*E. coli* Hyd-1, *Ralstonia eutropha* and *Aquifex aeolicus*) and three O_2-sensitive (*E. coli* Hyd-2, *Desulfovibrio gigas* and *Desulfovibrio vulgaris*) bacterial hydrogenases were used as references to compare and understand the *Quinella* hydrogenases. When the amino acid sequences of the *Quinella* hydrogenases were aligned and compared with those of the six reference hydrogenases, it was found that amino acid similarity varied from 36.1% to 58.3% but that the key amino acid residues aligned and matched well (Figure 6.16 and 6.17).

The large and small subunits of membrane-bound hydrogenases contain four co-factors: one NiFe catalytic centre in the large subunit and three FeS clusters in the small subunit. The large, or catalytic, subunit is responsible for H_2 oxidation and contains the NiFe active site. When compared with the reference hydrogenases, four cysteine residues (Cys62, Cys65, Cys608 and Cys610) surrounding the NiFe active site were found to be conserved in the proposed large catalytic subunit of the *Quinella* hydrogenases (Figure 6.16). Furthermore, a histidine residue (His236), a metal-binding amino acid that coordinates with the proximal 4Fe4S or 4Fe3S cluster of the small subunit (Volbeda et al., 1995), was also found in the correct location when aligned with the amino acid sequences from the reference hydrogenases (Figure 6.16). The large subunit sequences from the *Quinella* genomes are therefore very similar to those of other O_2-tolerant and O_2-sensitive NiFe type membrane-bound hydrogenases. This was expected, since the large catalytic subunits of O_2-tolerant and O_2-sensitive type hydrogenases are very similar in arrangement and this subunit does not confer the O_2 sensitivity (Goris et al., 2011).

The small subunit contains three FeS clusters, termed the proximal, medial and distal clusters. The proximal FeS cluster accepts electrons from the large subunit, where the hydrogen is oxidised at the NiFe centre, and the electrons are conducted via the medial FeS to the distal FeS cluster (Fritsch et al., 2011). On the basis of crystallography studies conducted by Shomura et al. (2011) and Fritsch et al. (2011), the O_2-sensitive NiFe type membrane-bound hydrogenases, termed 1d group type hydrogenases by Greening et al. (2016), contain standard 4Fe-4S clusters. In contrast, the O_2-tolerant type, termed group 1e by Greening et al. (2016), has a novel proximal 4Fe3S cluster. When the amino acid sequences of the small subunits from *Quinella* were aligned to reference hydrogenase sequences (Figure 6.17A), it was found that the *Quinella* small subunits have four cysteine residues (Cys62, Cys65, Cys160 and Cys194) that should bind a proximal 4Fe-3S cluster instead of the six cysteine residues that bind the 4Fe-4S cluster found in the O_2-tolerant type. In the O_2-tolerant type, there are two extra cysteines (Cys55 and Cys165), and targeted mutation of these two cysteines to glycine can convert an O_2-tolerant hydrogenase to an O_2-sensitive type (Lukey et al., 2011). This analysis therefore suggests that the small subunit of the *Quinella* hydrogenase is similar to those of O_2-sensitive NiFe type membrane-bound uptake hydrogenases, and therefore probably not bidirectional.

Membrane-bound hydrogenases are proposed to be dimers of heterotrimers, i.e., two copies of each of the large and small subunits and a *b*-type cytochrome (cyt*b*) (Shomura et al., 2011; Volbeda et al., 2012). The small subunits of the hydrogenases from the *Quinella* genomes also contained three conserved residues (Arg238, Lys263 and Arg311; see Figure 6.17A) that are proposed to be exposed towards the carboxylate group of the heme of the *b*-type cytochrome associated with NiFe type membrane-bound hydrogenases (Volbeda et al., 2013). All four *Quinella* genomes contained genes coding for cytochrome *b* (cyt*b*), the redox partner of membrane-bound hydrogenases. The amino acid sequence of the *Aquifex aeolicus* protein was the most similar to the *Quinella* amino acid sequences, with an average of 38% amino acid similarity. Overall, cyt*b* amino acid sequences of *Quinella* were found to be 26% identical to that of *E. coli* Hyd-1 and 31% identical to that of *Ralstonia eutropha*. All *Quinella* cyt*b* sequences were >81% similar to each other and the *S. ruminantium* cyt*b* was found to be the closest relative by BLAST analysis against the NCBI non-redundant database. Out of 21 conserved surface amino acid sequences from cyt*b* that have contact with the small subunit of the hydrogenases (Volbeda et al., 2013), 13 were found in cyt*b* of *Quinella* (Figure 6.18A). Transmembrane topology analysis using SPOCTOPUS (Viklund and Elofsson, 2008), suggested that the all four *Quinella* genome bins cyt*b* contain four transmembrane helices (Figure 6.18B), which is reported for other cytb associated with membrane-bound NiFe-hydrogenases (Volbeda et al., 2012).

Cytochrome *b* is unstable when removed from the cytoplasmic membrane (Menon et al., 1991), and to date there is only one crystal structure of a hydrogenase (*E.* coli Hyd-1) with its cytochrome *b* complex available (Volbeda et al., 2013). The amino acid sequence of this enzyme, *E. coli* Hyd-1, plus sequences from *E. coli* Hyd-2, *R. eutropha* and *A. aeolicus* (well annotated and reviewed cyt*b* sequences) were used as references to understand the *Quinella* cyt*b*. The amino acid sequence of the cyt*b* from the *E. coli* Hyd-2 was very different to the other amino acid sequences on the basis of the alignment. Histidine residues (His25, His67, His181 and His195), metal-binding amino acid residues, were conserved in all sequences (Figure 6.18A) except for *E. coli* Hyd-2. At least three (His25, His67, and His195) of the four conserved histidine residues served as ligands for two heme, the carrier for electron movement from the periplasmic to the cytoplasmic side of the membrane (Berks et al., 1995), and mutations in any of these conserved histidines leads to total loss of cytochrome *b* reduction by H_2 (Berks et al.,

1995; Gross et al., 1998; Meek and Arp, 2000). Furthermore, these four conserved histidines are also conserved in quinone-interacting *b*-type cytochromes (Berks et al., 1995; Gross et al., 1998) and His25, His67, and His195 are required for electron transport from H_2 to fumarate and for quinone reactivity in *Wolinella succinogenes* (Gross et al., 1998).

To understand the orientation of *Quinella* hydrogenase, membrane protein topology and signal peptides prediction was done using web-based server of SPOCTOPUS (Viklund and Elofsson, 2008) and Phobius (Käll et al., 2004). The large subunits from all four *Quinella* genome bins were predicted to be non-cytoplasmic, suggesting they are present outside the cell, although no leader sequences were detected. The small subunits appeared to have one *N*-terminal signal peptide region and contained two transmembrane-helices (Figure 6.17B). The position of the proximal FeS cluster of the small subunit that faces towards the large subunit appeared to be adjacent to the signal peptide sequence and so oriented towards the outer surface of the cell, based on the location of the residues that coordinate it as reported by Volbeda et al. (1995). This suggests that the large subunit of the hydrogenase is on the outside of the cell where it is connected to the small subunit, which has three FeS clusters that transfer electrons, ultimately to the fumarate reductase.

In summary, evidence was found for the presence of hydrogenases in all four *Quinella* genome bins. These can be classified as NiFe type membrane-bound H_2-uptake hydrogenases. Based on the inferred presence of a proximal 4Fe3S cluster in the small subunit, it appears that the *Quinella* hydrogenase is an O_2 sensitive type, even though other analyses (overall amino acid similarity and phylogeny) suggested that it an O_2 tolerant type. Confirmation of this will require experimental evidence. Evidence was found to suggest that the hydrogenase is associated with a *b*-type cytochrome that contains ligands for two heme groups, and that the reaction with hydrogen occurs on the outer face of the cell membrane.

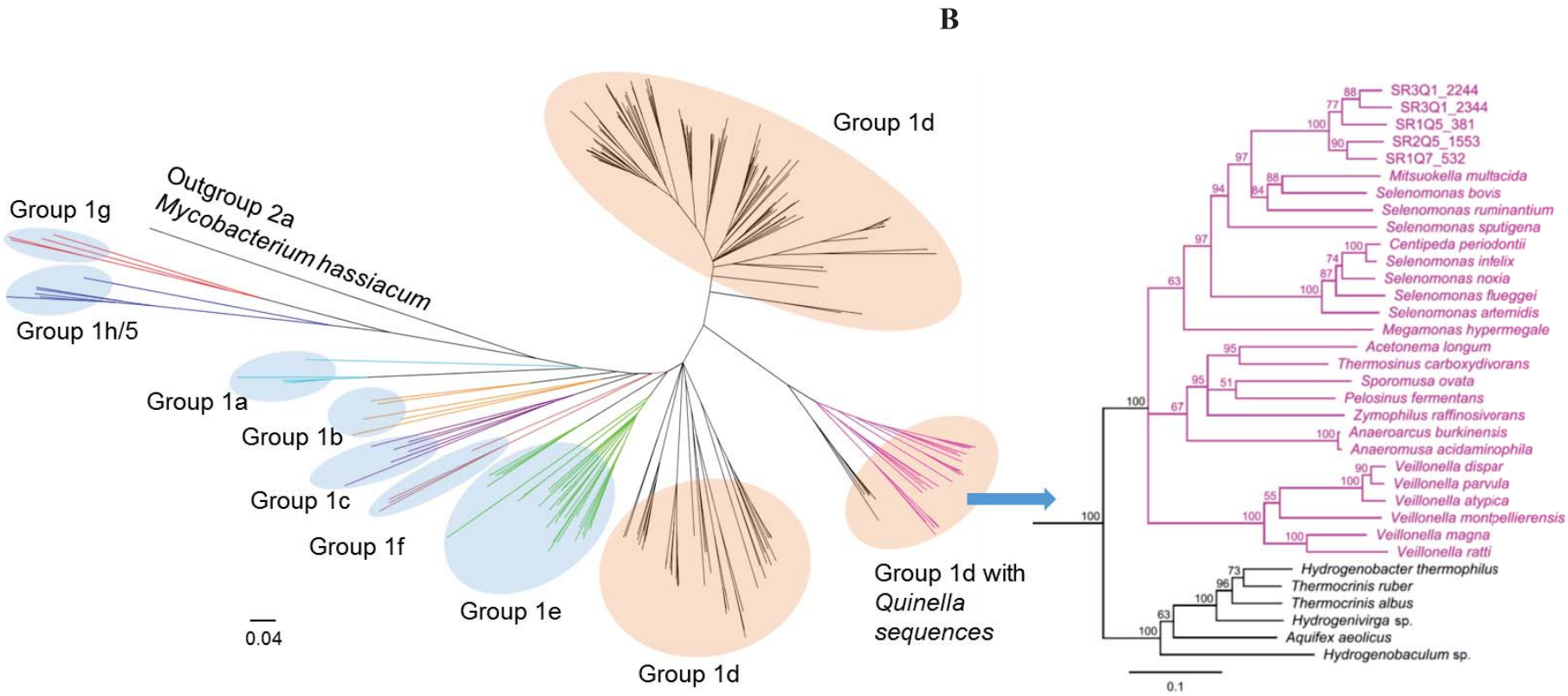


Figure 6.15 Phylogenetic classification of *Quinella* hydrogenases using the amino acid sequences of the large (catalytic) subunit. A) Phylogenetic tree built using all sequences from 1d and 1e type, and five sequences each from 1a, 1b, 1c, 1f and 1h/5 type hydrogenases from the Greening *et al*. (2016) database and 5 sequences from the *Quinella* genome bins. Sequences were aligned using MUSCLE (Edgar, 2004) and Clustal W (Larkin et al., 2007),

and then a phylogenetic tree was constructed using the Neighbor-Joining method with the Jukes-Cantor genetic distance model and 500 bootstrap resamplings. A 2a type hydrogenase sequence from *Mycobacterium hassiacum* was used as the outgroup to root the tree. B) Detailed view of the cluster containing the "*Quinella* sequences" (pink) and adjacent 1d sequences (black). The scale bar represents 0.04 and 0.1 changes per amino acid residue in panels A and B respectively. The numbers at the nodes are the percentage of trees that conserved that node in 500 bootstrap resamplings.

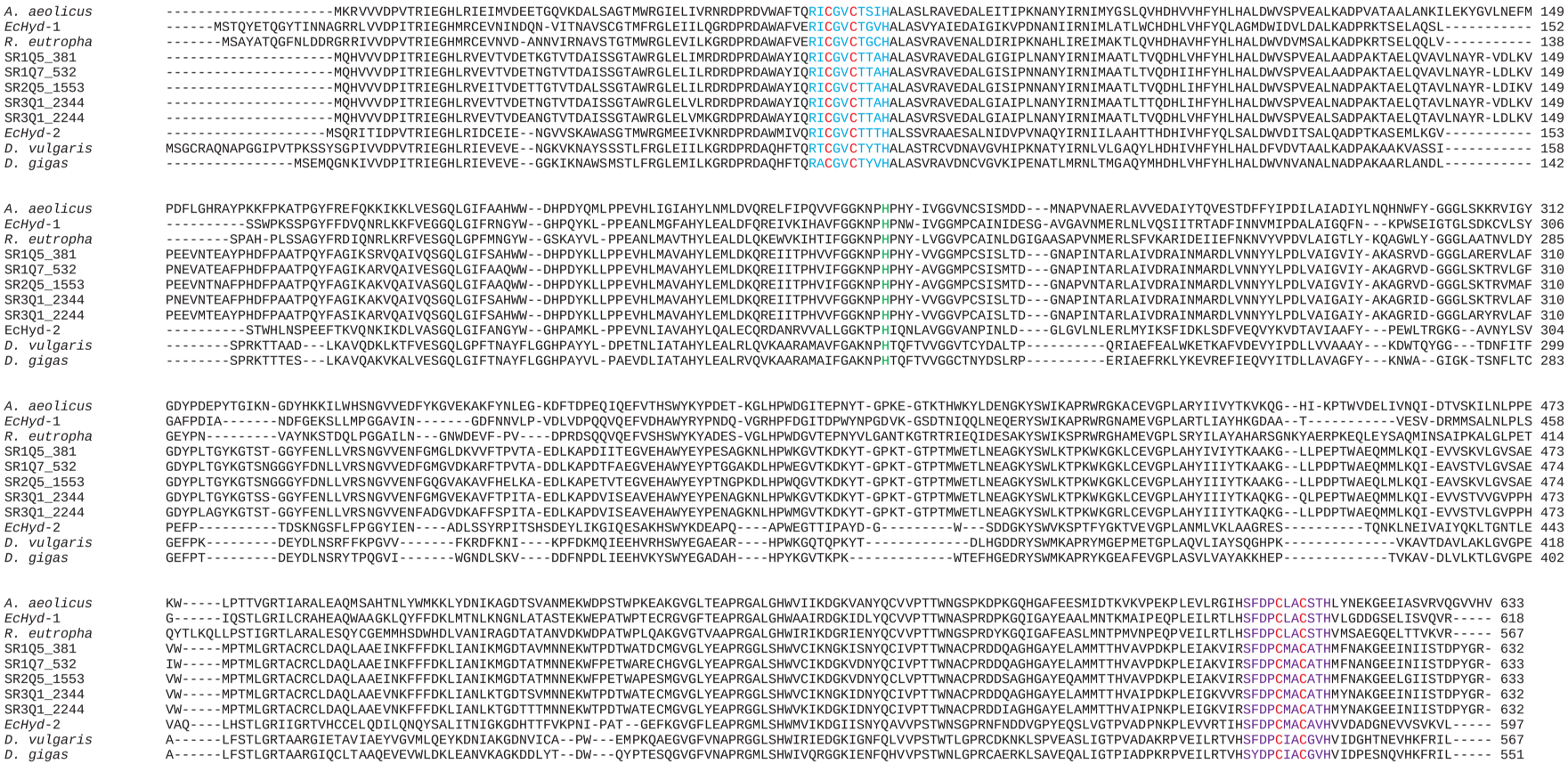


Figure 6.16 Amino acid sequence alignment of the large subunit of the hydrogenase of *Quinella* together with O_2-tolerant (*E. coli* Hyd-1, *R. eutropha* and *A. aeolicus*) and O_2-sensitive (*E. coli* Hyd-2, *D. vulgaris* and *D. gigas*) large subunit sequences. Motif 1 (light blue with conserved cysteines in red)

matches with 1d type hydrogenases while motif 2 (purple with conserve cysteines in red) matches with 1e type hydrogenases as defined by Greening et al. (2016). Conserved metal-binding residue His236 (green) in *Quinella* that coordinates with the proximal 4Fe4S or 4Fe3S cluster of the small subunit (Volbeda et al., 1995).

A

```
A. aeolicus     ----------------------------------------------------------------------------------------------------METFWEVFKRHGVSRRDFLKFATTITGLMGLAPSMVPEVVRAMETKPRVPVLWIHGLECTCCSESFIRSATPLASD 76
EcHyd-1         ----------------------------------------------------------------------------------------------------MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWALENKPRIPVVWIHGLECTCCTESFIRSAHPLAKD 79
R. eutropha     ----------------------------------------------------------------------------------------------------MVETFYEVMRRQGISRRSFLKYCSLTATSLGLGPSFLPQIAHAMETKPRTPVLWLHGLECTCCSESFIRSAHPLAKD 77
SR1Q5_380       ----------------------------------------------------------------------------------------------------MEKRESLWEAYLRRGFSRRSFLKGCVALTSLMGFSTDMFSKVVEAAESKPLPVVIWIHGHECTGCDESFIRSGAPLASD 79
SR1Q7_531       ----------------------------------------------------------------------------------------------------METRESLWEAYLKRGFSRRSFLKGCVALTSLMGLSTDMFSKVVEAAETKPLPVVIWIHGHECTGCDESFIRSGAPLASD 79
SR2Q5_1552      ----------------------------------------------------------------------------------------------------METRESLWEAYLRRGLSRRSFLKGCVALTSLMGLSADMFSKVVEAAENKPLPVVIWIHGHECTGCDESFIRSGAPLASD 79
SR3Q1_2245      ----------------------------------------------------------------------------------------------------MEKRESLWEAYLRRGFSRRSFLKGCVALTSLMGLSTDMFSKVVEAAESKPLPVVIWIHGHECTGCDESFIRSGAPLASD 79
SR3Q1_2502      ----------------------------------------------------------------------------------------------------MEKRESLWEAYLKRGFSRRSFLKGCVALTSLMGFGTDMFSKVVEAAETKPLPVVIWIHGHECTGCDESFIRSGAPLASD 79
SR3Q1_2343      MAISFRRFGNILLLKIFIVKLRYPLTSEKYRRRVDNRRAFFYKWRSTSGRIYKNFGESCCQYRIDIAQRMLKFFGTYYAYHLPIIFEEGRTLEKRESLWEAYLRRGFSRRSFLKGCVALTSLMGLSTDMFSKVVEAAEMKPLPVVIWIHGHECTGCDESFIRSGAPLASD 170
EcHyd-2         ----------------------------------------------------------------------------------------------------MTGDNTLIHS---HGINRRDFMKLCAALAATMGLSSKAAAEMAESVTNPQRPPVIWIGAQECTGCTESLLRATHPTVEN 76
D. vulgaris     ----------------------------------------------------------------------------------------------------MKISIGLGKEGVEERLAERGVSRRDFLKFCTAIAVTMGMGPAFAPEVARALMGPRRPSVVYLHNAECTGCSESVLRAFEPYIDT 84
D. gigas        ----------------------------------------------------------------------------------------------------MKFCTAVAVAMGMGPAFAPKVAEALTAKKRPSVVYLHNAECTGCSESLLRTVDPYVDE 58

A. aeolicus     VVLSMISLEYDDTLSAAAGEAVEKHRERIIKEYWGNYILAVEGNPPLGEDGMYCIIGGRPFVEILKESAEGAKAVIAWGSCASWGCVQAAKPNPTTAVPIDKVIK--DKPIIKVPGCPPIAEVMTGVIMYMVLFDRIPPLDSQGRPKMFYGNRIHDTCYRRSFFNAGQFV 244
EcHyd-1         VILSLISLDYDDTLMAAAGTQAEEVFEDIITQYNGKYILAVEGNPPLGEQGMFCISSGRPFIEKLKRAAAGASAIIAWGTCASWGCVQAARPNPTQATPIDKVIT--DKPIIKVPGCPPIPDVMSAIITYMVTFDRLPDVDRMGRPLMFYGQRIHDKCYRRAHFDAGEFV 247
R. eutropha     VVLSMISLDYDDTLMAAAGHQAEAILEEIMTKYKGNYILAVEGNPPLNQDGMSCIIGGRPFIEQLKYVAKDAKAIISWGSCASWGCVQAAKPNPTQATPVHKVIT--DKPIIKVPGCPPIAEVMTGVITYMLTFDRIPELDRQGRPKMFYSQRIHDKCYRRPHFDAGQFV 245
SR1Q5_380       VLLSSIALEYDHLLSAACGAPFEAHLDETITKYNGQYILAVEGAIATKDSGVYCMSGGHTFTHLLEKAAKGAAAIIAYGSCSAYGGIQAAKPNPTGSAGIERFVG--GKPVVKVPGCPPIPEVMTGVVMHVALFGTLPPVDADGRPKQFYGNRIHDTCYRRPFFDAGMFA 247
SR1Q7_531       VVLSSIALEYDHLLSAACGAPFEAHLDETIAKYNGQYILAVEGAIATKDSGVYCMSGGHTFTHLLEKAAKGAAAIIAYGSCSAYGGIQAAKPNPTGSAGIERFVG--GKPVVKVPGCPPIPEVMTGVVMHVALFGTIPPVDGDGRPKQFYGNRIHDTCYRRPFFDAGMFA 247
SR2Q5_1552      VVLSSIALEYDHLLSAACGAPFEAHLDETIAKYNGQYILCVEGAIATKDSGVYCMSGGHTFTHLLEKAAKGAAAIIAYGSCSAYGGIQAAKPNPTGSAGIERFVG--GKPVVKVPGCPPIPEVMTGVVMHVALFGTIPPVDGDGRPKQFYGNRIHDTCYRRPFFDAGMFA 247
SR3Q1_2245      VVLSSIALEYDHLLSAACGAPFEAHLDETIAKYKGQYILAVEGAVATRDSGVYCMSGGHTFTHLLEKAAAGAAAIIAYGSCAAYGGIQAAKPNPTGSAGIEHFVG--GKPVVKVPGCPPIPEVMTGVVMHVALFGSLPPCDADGRPKQFYGNRIHDTCYRRPFFDAGMFA 247
SR3Q1_2502      VVLSSIALEYDHLLSAACGAPFEAHLDETIAKYNGQYILAVEGAVATKDSGVYCMSGGHTXXXX---------------------------------------------------------------------------------------------------- 144
SR3Q1_2343      VVLSSIALEYDHLLSAACGAPFEAHLDETIAKYNGQYILAVEGAVATRDSGVYCMSGGHTFTHLLEKAAKGAAAIIAYGSCSAYGGIQAAKPNPTGSAGIERFVG--GKPVVKVPGCPPIPEVMTGVVMHVALFGSLPPCDADGRPKPFYGNRIHDTCYRRPFFDAGMFA 338
EcHyd-2         LVLETISLEYHEVLSAAFGHQVEENKHNALEKYKGQYVLVVDGSIPLKDNGIYCMVAGEPIVDHIRKAAEGAAAIIAIGSCSAWGGVAAAGVNPTGAVSLQEVLP--GKTVINIPGCPPNPHNFLATVAHIITYGKPPKLDDKNRPTFAYGRLIHEHCERRPHFDAGRFA 244
D. Vulgaris     LILDTLSLDYHETIMAAAGDAAEAALEQAVNSPHG-FIAVVEGGIPTAANGIYGKVANHTMLDICSRILPKAQAVIAYGTCATFGGVQAAKPNPTGAKGVNDALKHLGVKAINIAGCPPNPYNLVGTIVYYLKNKAAPELDSLNRPTMFFGQTVHEQCPRLPHFDAGEFA 253
D. gigas        LILDVISMDYHETLMAGAGHAVEEALHEAI---KGDFVCVIEGGIPMGDGGYWGKVGRRNMYDICAEVAPKAKAVIAIGTCATYGGVQAAKPNPTGTVGVNEALGKLGVKAINIAGCPPNPMNFVGTVVHLLTKG-MPELDKQGRPVMFFGETVHDNCPRLKHFEAGEFA 244

A. aeolicus     EQFDDEGAKKGWCLYKVGCRGPTTYNSCGNMRWYN-GLSYPIQSGHGCIGCAENNF----------WDN-GPFYERIGGIPVPGIESKADKVGAIAAAAAAGGAIIHGIASKIRKSGEKEE---------- 353
EcHyd-1         QSWDDDAARKGYCLYKMGCKGPTTYNACSSTRWND-GVSFPIQSGHGCLGCAENGF----------WDR-GSFYSRVVDIPQMGTHSTADTVGLTALGVVAAAVGVHAVASAVDQRRRHNQQPTETEHQPGNEDKQA 372
R. eutropha     EEWDDESARKGFCLYKMGCKGPTTYNACSTTRWNE-GTSFPIQSGHGCIGCSEDGF----------WDK-GSFYDRLTGISQFGVEANADKIGGTASVVVGAAVTAHAAASAIKRASKKNETSGSEH---------- 361
SR1Q5_380       ENFDDAGAKAGWCLYKLGCRGPETYNSCGNLRWWQ-GMSYPIQSGAPCIGCSNSNF----------WDE-DPFTDRL---PKYGKLGDVDKIGVGLGVLTAAGVGAHAVASVIQRNQRENLLEEDHHE--------- 361
SR1Q7_531       ETFDDAGAKAGWCLYKLGCRGPETYNSCGNLRWWQ-GMSYPIQSGAPCIGCSNSNF----------WDE-DPFTARL---PQYGKLGDADKIGVGLGVLTAAGVVGHAVASIVQRNQRDNLID-EHHE--------- 360
SR2Q5_1552      ETFDDAGAKAGWCLYKLGCRGPETYNSCGNLRWWQ-GMSYPIQSGAPCIGCSNSNF----------WDE-DPFTARL---PQYGKIGDADKIGTALGVLTAAGVVGHAVASVIQRNQRDNLIDEEHHE--------- 361
SR3Q1_2245      ETFDDAGAKAGWCLYKLGCRGPETYNSCGNLRWWQ-GMSYPIQSGAPCIGCSNSNF----------WDE-DPFTDRL---PKYGKFGDVDKIGAALGVLTAAGVAGHAVASIVQRNTRENLIDEEHHE--------- 361
SR3Q1_2502      ---------------------------------------------------------------------------------------------------- 144
SR3Q1_2343      ETFDDAGAKAGWCLYKLGCRGPETYNSCGNLRWWQ-GMSYPIQSGAPCIGCSNANF----------WDE-DPFTNRL---PQYGKFGDVDKIGAALGVLTAAGVAGHAVASIVQRNQRENLIDEEHH---------- 451
EcHyd-2         KEFGDEGHREGWCLYHLGCKGPETYGNCSTLQFCDVGGVWPVAIGHPCYGCNEEGIGFHKGIHQLANVENQTPRSQKPDVNAKEGGNVSAGAIGLLGGVVGLVAGVSVMAVRELGRQQKKDNADSRGE---------- 373
D. vulgaris     PSFESEEARKGWCLYELGCKGPVTMNNCPKIKFNQ--TNWPVDAGHPCIGCSEPDF----------WDAMTPFYQN---------------------------------------------------------- 318
D. gigas        TSFGSPEAKKGYCLYELGCKGPDTYNNCPKQLFNQ--VNWPVQAGHPCIACSEPNF----------WDLYSPFYSA---------------------------------------------------------- 289
```

B

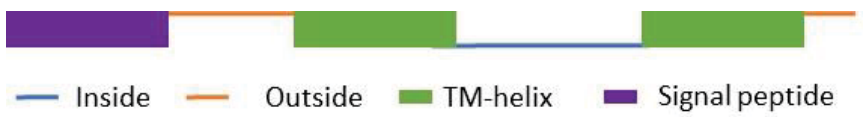


Figure 6.17 A) Amino acid sequence alignment of the small subunit of the hydrogenase of *Quinella* together with O_2-tolerant (*E. coli* Hyd-1, *R. eutropha* and *A. aeolicus*) and O_2-sensitive (*E. coli* Hyd-2, *D. vulgaris* and *D. gigas*) small subunit sequences. Conserved cysteine (red) and glycine (light blue)

residues proximal to the 4Fe4S cluster are indicators of O_2-sensitive type hydrogenases (Lukey et al., 2011). Conserved arginine and lysine residues (orange) are proposed to be exposed towards the carboxylate group of the heme of the *b*-type cytochrome associated with NiFe type membrane-bound hydrogenases (Volbeda et al., 2013). B) Schematic representation of the transmembrane helices and signal peptide in the small subunit of the *Quinella* hydrogenase. TM, transmembrane.

A

```
SR1Q5_382          ------------------MAIKSVKDVKRKEYYIFSPFLRFFHWIMASQIVVLFATGLLITKP--MNVGAVEPVFT---NTSMDLVRDIHFTAAFILCAALILRVY-GFIVNKGDR-LF-PRVWEGHFYEETVDVALHYM 114
SR1Q7_533          ------------------MASKSVKDVQRKEYYIFSPFLRIFHWIMASQIVILFATGLLITKP--MNVAITEPVFS---ITSMDLVRDIHFTSAFILCASLILRMY-GFIINKGDR-LF-PRVWEGHFYEETVDVALHYM 114
SR2Q5_1554         ------------------MAIKSVKDVHRKEYYIFSPFLRIFHWIMASQIVILFATGLLITKP--MNVAITEPVFS---ITSMDLVRDIHFTSAFILCASLILRIY-GFIINKGDR-LF-PRVWEGHFYEETVDVALHYM 114
SR3Q1_2243         ------------------MAIKSVKNVKRKEYYLFSPFLRIFHWIMVVMILVLFGTGLLITKP--MNVGGTEPVFA---GMSVDIIRDVHFLAAFILCAALIMRVY-GFIINKGDR-LF-PRVWEGHFYSETVEVALHYM 114
SR3Q1_2345         ------------------MAIKSVKDVKRKEYYLFSPFLRIFHWIMVVMILVLFGTGLLITKP--MNVGGTEPVFA---GMSVDIIRDIHFLAAFILCAGLIMRVY-GFIINKGDR-LF-PRVWEGHFYSETVEVALHYM 114
EcHyd-1            ------------------MQQKSDNVV---SHYVFEAPVRIWHWLTVLCMAVLMVTGYFIGKP--LPSVSGEATY----LFYMGYIRLIHFSAGMVFTVVLLMRIYWAFVGNRYSRELFIVPVWRKSWWQGVWYEIRWYL 113
A. aeolicus        --------------------------MFKRIFIFSASLRIWHWINFLTIMVLFFTGLYIGTPYFTTSSGYEATYAYDKNITMGFIRFLHFSAGYVLLFSFLFRIFIAFF-RKGDR-LFIPKFWTIEYWEGIKETLKEYL 111
R. eutropha        MSTKMQADRIADATGTDEGAVASGKSI--KATYVYEAPVRLWHWVNALAIVVLAVTGFFIGSP--PATRPGEASA----NFLMGYIRFAHFVAAYIFAIGMLGRIYWATAGNHHSRELFSVPVFTRAYWQEVISMLRWYA 132
S. ruminantium     ----------------------MKQVVRREYYLFSPWLRIFHWIMVVSICVLFATGLMITSPWI--IYSAEPTFS---GMTVDNIRNIHFLAAYLFSAAFILKIY-GFIINPGDR-LF-PHVWKGYFYSNTVDVALHYM 109

SR1Q5_382          LLRGHHASFL-RNPLARGSYAVLYVL-LAIEIFTGFAMYCAANPTGTGY------LLFGWVNTLVGSEMMSHYIHHYVAWFIILFAIGHLYMVIRAEFMEGESEISSMFSGKKILAHTPKDANELD---- 232
SR1Q7_533          LIRPHHASFL-RNPLARGSYAALYLL-LAIEILTGFAMYFMTNPSSTGG------MLFGWVNVLVGSEMMAHYIHHYVAWFIILFAIGHLYMVIRAEFMEGESEISSMFSGKKILAHTPKDANELD---- 232
SR2Q5_1554         LIRPHHASFL-RNPLARGSYAALYLL-LLIEILTGFAMYFMTNPSSTGG------MLFGWVNVLVGSEMMAHYIHHYVAWFIILFAIGHLYMVIRAEFMEGESEISSMFSGKKILEHTPKDANELD---- 232
SR3Q1_2243         LLKSHHAPFL-RNPLARGSYAGLYVL-LAIEILTGFAMYLMANPFSTGG------MLFGWVNTLVGSEMMSHYIHHYVAWIIILFAIGHFYMVVRAEFMEGESEVSSMFSGKKVLPHTPKDANELD---- 232
SR3Q1_2345         LLKSHHSPFL-RNPLARGSYAGLYVL-LAIEILTGFAMYLMANPSSVFG------SLFNWVNVLVGSEMMSHYIHHYVAWIIILFAIGHFYMVVRAEFMEGESEVSSMFSGKKVLPHTPKDANELD---- 232
EcHyd-1            FLAKRPSADIGHNPIAQAAMFGYFLM-SVFMIITGFALYSEHSQYAIFAPFRYVVEFFYWTG---GNSMDIHSWHRLGMWLIGAFVIGHVYMALREDIMSDDTVISTMVNGYRSHKFGKISNKERS---- 235
A. aeolicus        LIGSRK-PMI-RNPLARTAYFFVYLM-IFFMIATGFGMYGLSDPDGFWA------SVFGWVVWSLGGEFEAHYWHHIVAWLIILFFIVHVYFVSREDFVHKNGEVSSMFNGSKFFKKDPVDLKDIK---- 228
R. euturopha       FLSARPSRYVGHNPLARFAMFFIFFLSSVFMILTGFAMYGEGAQMGSWQ-----ERMFGWVIPLLGQSQDVHTWHHLGMWFIVVFVIVHVYAAIREDIMGRQSVVSTMVSGYRTFKD------------ 244
S. ruminantium     FLRSEHASYL-RNPLARCSYAFLYVL-VFIEILTGFGIYYMTDPFGIGG------TLFGWVITVLGGEMMAHLVHHYVAWAIILFAIGHLYMVIRAEFMEGESEVSSMFSGSKILAHEPRDAGHLDGTLK 231
```

B

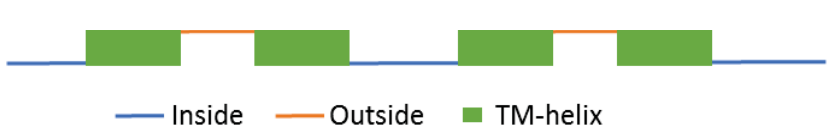


Figure 6.18 A) Amino acid sequences alignment of the cytochrome-*b* subunit of the *Quinella* hydrogenase, together with reference cyt*b* from other hydrogenases. At least three (His43, His90, and His215) of the four conserved histidine residues (red) serve as ligands for two heme groups (Berks et al., 1995). B) Schematic representation showing the four transmembrane helices. TM, transmembrane.

6.2.5.4 Rnf complex

The *Quinella* genome bins were searched for evidence of the *rnf* gene cluster *rnfCDGEAB* that codes for the ferredoxin:NAD^+-oxidoreductase complex and is composed of at least six subunits. This enzyme complex might allow H_2 generation from NADH and ferredoxin via an electron bifurcating mechanism (Buckel and Thauer, 2013). Reviewed rnf complex amino acid sequences from the bacteria *Acetobacterium woodii* and *Rhodobacter capsulatus* were used as queries in a BLAST search against the *Quinella* genome bins. None of the subunits were found in any of the *Quinella* genome bins using this approach. However, the GAMOLA2 search tool suggested that a gene sequence potentially coding for subunit D (SR3Q1_1981; 29% identity with NADH:ubiquinone oxidoreductase of *Anaerococcus* sp. PH9) was present in *Quinella* genome bin SR3Q1, and that genes for subunit E (SR1Q7_1676; 78% identity with putative sugar-specific permease SgaT of *Candidatus* Arthromitus sp. SFB-3) and subunit G (SR1Q7_1737; 49% identity with FMN-binding protein *Selenomonas* sp. CM52) were found in bin SR1Q7. All six subunits are required for a fully functional rnf complex (Biegel et al., 2009), and so it can be assumed that *Quinella* does not contain a fully functional rnf complex. Consequently, the subunits detected may come from assembly or binning errors, or play a different role in these bacteria. This conclusion is supported by the generally complete nature of other multi-subunit enzyme complexes found in these four genome bins.

6.2.5.5 ATP synthase

The membrane-bound ATP synthase catalyses ATP synthesis from ADP and inorganic phosphate driven by the translocation of Na^+ or/and H^+ ions across a biological membrane (Skulachev, 1978; Senior, 1988). All four *Quinella* genome bins contained genes that code for all eight subunits of the bacterial type ATP synthase (F_1F_o). The deduced amino acids of these had the greatest similarity to ATP synthase subunits of *Selenomonas* spp. The F_o part of the bacterial ATP synthase is the integral membrane ion-translocating complex containing α, β, γ, δ and ε subunits whereas the F_1 part is a peripheral membrane unit on the cytoplasmic face and is composed of a, b and c subunits (Engelbrecht and Junge, 1997). Grüber et al. (2014) identified conserved amino acid residues that allow classification of ATP synthases as Na^+- or H^+-binding types in a wide range of bacteria and archaea. They analysed amino acid sequences of subunit c (the ion carrier) and suggested that motif Q/E…E T/S (at positions 32, 65 and 66 in the *Ilyobacter tartaricus*

protein) are conserved in Na^+-binding ATP synthases, while N/I…E/D A were found conserved in the corresponding positions in H^+-binding bacterial ATP synthases. Amino acid sequences from the ATP synthase c subunits from the *Quinella* genome bins were aligned with the reviewed and experimentally verified bacterial sequences (Grüber et al., 2014) from the Na^+-binding c subunits from *Acetobacterium woodii*, *Ilyobacter tartaricus*, and *Propionigenium modestum*, and the H^+-binding c subunits from *Bacillus subtilis* and *Escherichia coli*, to identify the positions of these diagnostic residues. The sequences from the four *Quinella* bins did not contain Q/E at position 32 or T/S at position 66 as would be expected for Na^+-binding c subunits. Instead, at those positions the amino acids were D at position 32 and A at position 66, indicative of a H^+-coupled enzyme. The LxEALxxI motif in the sequences from the *Quinella* genome bins which includes positions 65 and 66, was similar to that from *Bacillus subtilis* (LxEALxxI) (Figure 6.19), which is a H^+-binding ATP synthase (Santana et al., 1994). Furthermore, the overall amino acid sequence was also most similar to that of *Bacillus subtilis* (42.9%) than to the other reviewed reference sequences compared (34.1 to 38.6%). Therefore, on the basis of sequences similarity and conserved amino acid residues, *Quinella* ATP synthases appear to be H^+ binding enzymes.

```
SR1Q5_1786                1 ----MEHAIMVAAALLGAGITMGLAAIGAGVGDGLVTSKFIDGITRQPEAKNTLFTNTL
SR1Q7_892                 1 ----MEHAIMVAAALLGAGITMGLAAIGAGVGDGLVTSKFIDGITRQPEAKNTLFTNTL
SR2Q5_1189                1 ----MEHAIMVAAALLGAGITMGLAAIGAGVGDGLVTSKFIDGITRQPEAKNTLFTNTL
SR3Q1_1280                1 ----MEHAIMVAAALLGAGITMGLAAIGAGVGDGLVTSKFIDGITRQPEAKNTLFTNTL
Bacillus subtilis         1 ------------MNLIAAAIAIGLGALGAGIGNGLIVSRTVEGIARQPEAGKELRTLMF
Escherichia coli          1 -----MENLNMDLLYMAAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFF
Propionigenium modestum   1 MDMVLAKTVVLAASAVGAGAAM-IAGIGPGVGQGYAAGKAVESVARQPEAKGDIISTMV
Ilyobacter tartaricus     1 MDMLFAKTVVLAASAVGAGTAM-IAGIGPGVGQGYAAGKAVESVARQPEAKGDIISTMV
Acetobacterium woodii     1 ---MEGLDFIKACSAIGAGIAM-IAGVGPGIGQGFAAGKGAEAVGRQPEAQSDIIRTML

SR1Q5_1786               56 ISVGLIEAMAIIATVVALIMLYANPLL---
SR1Q7_892                56 ISVGLIEAMAIIATVVALIMLYANPLL---
SR2Q5_1189               56 ISVGLIEAMAIIATVVALIMLYANPLL---
SR3Q1_1280               56 ISVGLIEAMAIIATVVALIMLYANPLLG--
Bacillus subtilis        48 MGIALVEALPIIAVVIAFLAFFG-------
Escherichia coli         55 IVMGLVDAIPMIAVGLGLYVMFAVA-----
Propionigenium modestum  59 LGQAIAESTGIYSLVIALILLYANPFVGLL
Ilyobacter tartaricus    59 LGQAVAESTGIYSLVIALILLYANPFVGLL
Acetobacterium woodii    56 LGAAVAETTGIYGLIVALILLFANPFF---
```

Figure 6.19 Alignment of amino acid sequences from subunit c of ATP synthase. The diagnostic residue at position 32 of the *Quinella* sequences was not the **Q**/E expected for Na^+-specific ATP synthases (Grüber et al., 2014). The motif coloured red contains the conserved amino acid residue **E** at position 65 found in nearly all sequences except *E. coli*, and the residue at position 66 which is **S/T** in bacterial Na^+-specific ATP synthases and **A** in bacterial H^+-specific ATP synthases (Grüber et al., 2014). Sequence numbering is based on the *Ilyobacter tartaricus* protein.

6.2.5.6 Na^+/H^+ antiporter

All four *Quinella* genome bins also contained genes that code for a Na^+/H^+ antiporter. So far, seven structurally different antiporter genes have been found in bacteria (Ito et al., 1997). BLAST was used to compare the amino acid sequences from the putative Na^+/H^+ antiporter genes in the *Quinella* genome bins to reviewed sequences of all seven different types of antiporters. This comparison, showed that the *Quinella* genome bins contained sequences most similar to a putative *nhaC* from *Selenomonas ruminantium* (72% to 74% identity). There were no matches to validated antiporters. The Na^+/H^+ antiporter assists in maintaining pH homeostasis and may also help lowering cytoplasmic Na^+ concentration (Ito et al., 1997). However, in *Quinella*, it could be speculated that the Na^+/H^+ antiporter plays a role in converting the Na^+ gradient generated by the methylmalonyl-CoA decarboxylase (section 6.2.4.2) into a H^+ gradient that can be used by the H^+-translocating ATP synthase to generate ATP (section 6.2.5.5). This would be especially true if oxaloacetate is formed from phosphoenolpyruvate by a phosphoenolpyruvate carboxykinase rather than by a reversed oxaloacetate decarboxylase (Figure 6.1). In the case of the former, the Na^+ gradient generated by the methylmalonyl-CoA decarboxylase would not be needed to drive the synthesis of oxaloacetate, and could be converted to a H^+ gradient by the antiporter.

6.2.6 Amino acid and flagella synthesis in *Quinella*

The annotated *Quinella* genome bins were further queried against the KEGG database (Kanehisa et al., 2016) for identification of the genes that encode for the enzymes involved in amino acid synthesis pathways. *Quinella* genes were plotted on the amino acid synthesis pathways, and this analysis suggested that *Quinella* is able to synthesise all 20 amino acids. In some instances, a few genes from a pathway were missing from one genome bin but found in others. For example, 3-phosphoshikimate 1-carboxyvinyltransferase (aroA), an enzyme involved in phenylalanine, tyrosine and tryptophan biosynthesis, was absent in genome bin 2Q5 but present in the other three *Quinella* genomic bins (Appendix Figure A6.1). This finding suggesting that these isolated gene sequences might have been missed during sequencing, in assembly or in the binning process.

A similar trend was seen in the genes involved in the assembly of flagella. Overall analysis suggested that *Quinella* may able to make flagella (Appendix Figure A6.2).

Electron microscopy of preparations of *Quinella* revealed short surface structures, although it was unclear whether these were damaged or unusual flagella, some other outer membrane structure like fimbrae, or even preparation artifacts (Section 5.2.5). Orpin (1972) observed motility in *Quinella* but the presence of flagella was not reported. Further work is needed to determine if these genes encode for flagella proteins and synthesis, and whether they are expressed in *Quinella*.

6.3 Conclusions

The analysis of the four genome bins allows a preliminary metabolic scheme for *Quinella* to be constructed (Figure 6.20). CAZymes analysis of all four *Quinella* genome suggested that *Quinella* may not degrade polysaccharides and so are dependent on other rumen microbes to breakdown the polysaccharide components of feed, and then use the breakdown products for its growth. However, *Quinella* may be able to use cellobiose and cellodextrins released by other rumen microbes using a β-glucosidase. Genes coding for enzymes in the phosphotransferase system (PTS) for glucose, sorbitol, fructose, maltose, mannose, galactitol, glucitol and ascorbate transport were present in *Quinella* genome bins indicating that it uses smaller carbohydrates released by the polysaccharide degrading microbes. These carbohydrates will be converted to pyruvate by glycolysis. All *Quinella* genome bins also contained a gene for L-lactate dehydrogenase, suggesting that it can produce lactate as an end product (Figure 6.21A), use lactate as a substrate (Figure 6.21D), or even both. Previous work reported that *Quinella* may produce lactate (Brough et al., 1970), and this genomic analysis supports that. The use of lactate will require further studies, preferably with a culture of this bacterium.

All four *Quinella* genomic bins contained genes that code for PFOR, which converts pyruvate to acetyl-CoA. Acetyl-CoA appears to be converted to acetate using succinate CoA-transferase, with concurrent conversion of succinate to succinyl-CoA. Succinyl-CoA synthetase is then postulated to generate ATP from ADP with the release of succinate and CoA. This reaction is not reported for bacteria, but is found in helminths, protists and fungi (Marvin-Sikkema et al., 1993; Rivière et al., 2004; Van Grinsven et al., 2008). Interestingly, *S. ruminantium*, a close relative of *Quinella*, also may have this succinyl-CoA cycle for making ATP based on genes found in its genome. This cycle could be studied in *S. ruminantium*.

If *Quinella* makes acetate then it should also make propionate to balance the electrons released in conversion of glucose to pyruvate and pyruvate to acetyl-CoA. This balance would result in the formation of acetate and propionate in a ratio of 1 to 2 (Figure 6.21B). All the enzymes needed to form propionate were present in all four *Quinella* genome bins. Formation of oxaloacetate is a key step in this pathway, either from phosphoenolpyruvate via pyruvate or from directly from phosphoenolpyruvate. Both pathways seem possible in *Quinella*. If there is a functional OACD in *Quinella*, it seems to have an unusual structure. The OACD enzyme in *Quinella* may be a hybrid of α subunit of OACD, and the β, γ and δ subunits of MMCD. All other key enzymes in the propionate formation pathway for example, the Na^+-translocating methylmalonyl-CoA decarboxylase and the fumarate reductase are present. Overall, these analyses indicate that *Quinella* uses the randomizing pathway of acetate and propionate formation. A study of sheep rumens dominated by *Quinella* indicated that this bacterium may be able to form propionate (Vicini et al., 1987), and this study supports that.

Formation of lactate (Figure 6.21A) or of acetate and propionate (Figure 6.21B) will not result in production of hydrogen, one of the major precursors of CH_4 formation, explaining how *Quinella* might be functioning inside the low CH_4 emitting sheep rumen. Interestingly, it seems that *Quinella* might encodes enzymes that could allow uptake of exogenous hydrogen using a NiFe membrane bound uptake hydrogenase, which would transfer electrons to cytochrome *b*, and then further to a fumarate reductase. The presence of this NiFe membrane bound uptake hydrogenase in all four *Quinella* genome bins suggested that *Quinella* might even use H_2 for its growth. This would allow the production of more propionate at the expense of acetate (Figure 6.21C). However, the presence of an active hydrogenase need to be checked in a pure *Quinella* culture, which will also confirm the directionality of the hydrogenase (hydrogen uptake, hydrogen producing, or bidirectional).

Overall, from pyruvate, *Quinella* may carry out some of three different possible pathways of end product formation. These are fermentation to form lactate (Figure 6.21A), fermentation to form acetate plus propionate (Figure 6.21B) or fermentation of glucose plus hydrogen to form propionate (Figure 6.21C). The first two of these are standard and well-known metabolic schemes, but the third, the use of hydrogen plus glucose, is not. These different pathways of lactate, propionate and acetate formation result in the production of different amounts of ATP. Lactate formation from glucose (Figure 6.21A)

yields 2 ATP while formation of propionate and acetate (Figure 6.21B), or just propionate (Figure 6.21C) produce approximately 2.7 - 4.4 ATP (Table 6.4). If lactate is used as a substrate then electrons from the lactate to pyruvate and pyruvate to acetyl-CoA conversions may be used in the propionate pathway (Figure 6.21D), yielding 0.33 to 0.66 ATP (Table 6.4). Lactate could, however, theoretically be used together with hydrogen to form only propionate (Figure 6.21E) with formation of up to 1.2 ATP (Table 6.4). Multiple of these pathways might operate at the same time, producing a mix of products. For example, simultaneous formation of lactate, acetate and propionate from sugars seems feasible, as does the use of hydrogen plus sugars with variable amounts of hydrogen to yield different ratios of acetate to propionate. It does seem, from the analyses conducted here, that hydrogen formation is not supported, but this will require confirmation, preferably using a culture of *Quinella*.

So far, attempts to isolate *Quinella* performed by Orpin (1972) and in this project have been unsuccessful, but the information gathered using *Quinella* genome bins analysis may play useful role in future isolation attempts. Identification of an uptake hydrogenase suggests that including hydrogen together with sugars may lead to successful cultivation of *Quinella*. By using methanogen inhibitors, the use of the hydrogen by methanogens could be eliminated. The activity of *Quinella* in the rumen still needs to be confirmed by metatranscriptomic analyses of low CH_4 emitting sheep containing the Q-type community, much like Kamke et al. (2016) have analysed gene expression in sheep with S-type communities. The genome bins generated here will help with mapping transcripts because they provide genomic data for *Quinella*.

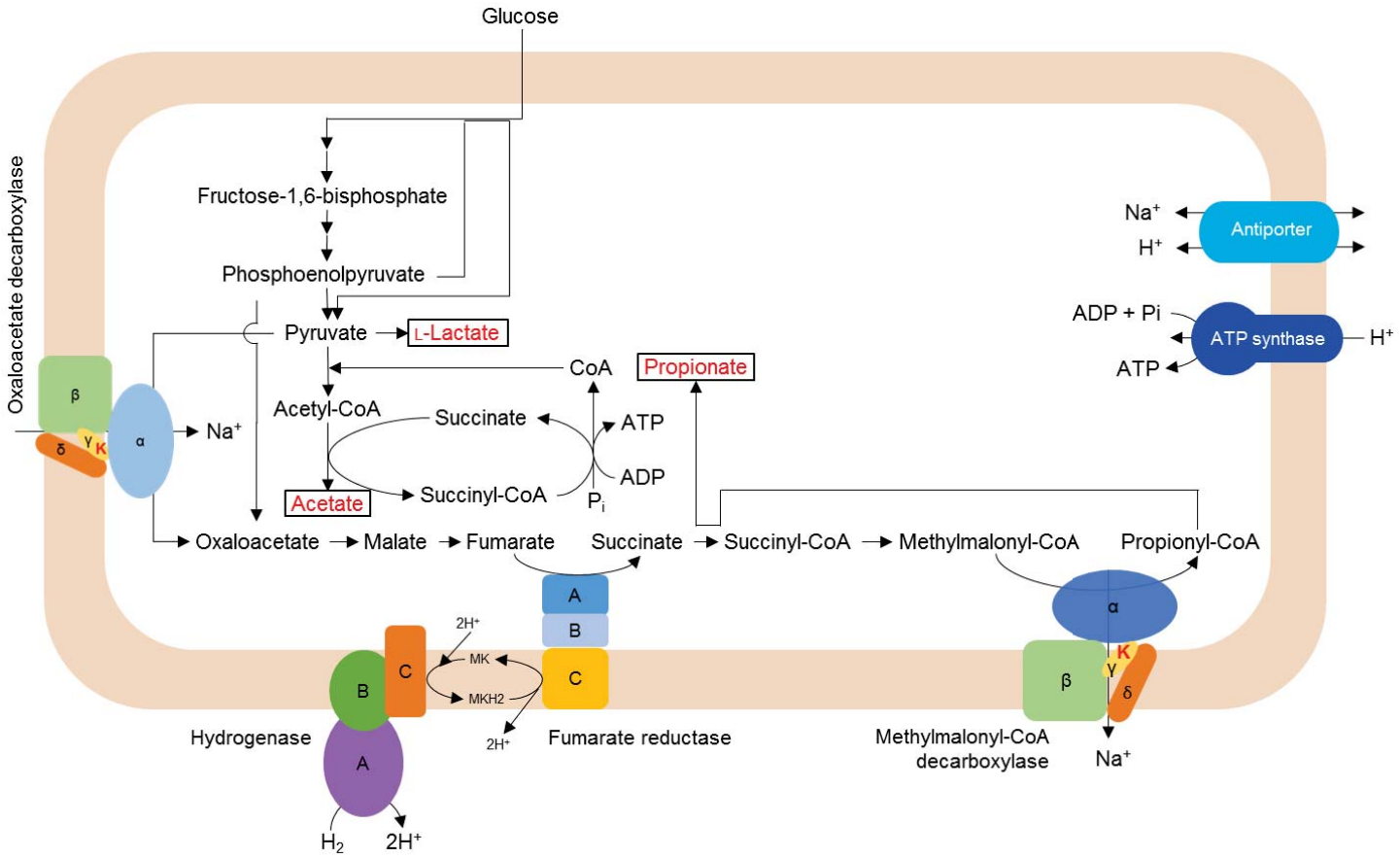


Figure 6.20 Schematic showing glucose fermentation pathway of *Quinella* based on four *Quinella* genomic bins and the analyses presented in this chapter.

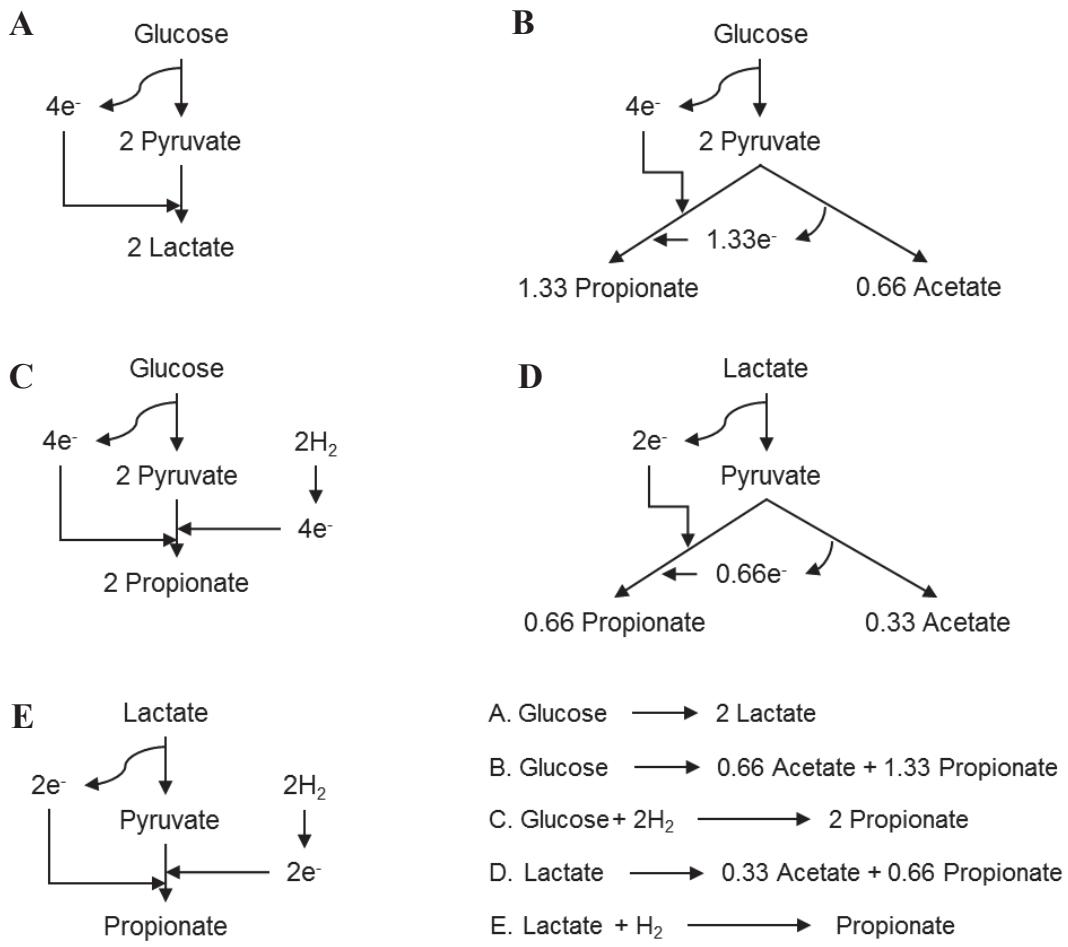


Figure 6.21 Possible end product formation by *Quinella* from glucose and lactate utilisation. There are three possible schemes for using glucose as a substrate. A) Two lactate can be produced from one glucose. B) When glucose is fermented without hydrogen uptake, then it will produce 0.66 acetate together with 1.33 propionate. C) If hydrogen is used via an uptake hydrogenase together with glucose, then two propionate can be produced. There are two pathways for using lactate as a substrate. D) Lactate can be fermented to 0.33 acetate and 0.66 propionate without hydrogenase activity, while E) one propionate will be produced if hydrogen is used by the uptake hydrogenase when using lactate. e^- represents electrons transferred by NAD, ferredoxin, cytochrome *b*, or other electron carriers.

Table 6.4 Steps and enzymes involved in ATP formation and consumption, following different pathways (Figure 6.21) for end product formation by *Quinella*.

Steps	Enzymes	ATP balances per substrate for different pathways				
		Glucose to 2 lactate	Glucose to 1.33 propionate + 0.66 acetate	Glucose + 2 H_2 to 2 propionate	Lactate to 0.66 propionate + 0.33 acetate	Lactate + H_2 to propionate
Glucose + Phosphoenolpyruvate → Glucose 6-phosphate	Phosphotransferase	0[a]	0[a]	0[a]	–[b]	–
Fructose-6-phosphate → Fructose-1, 6-bisphosphate	Phosphofructokinase	–1.0	–1.0	–1.0	–	–
1, 3-bisphosphoglycerate → 3-Phosphoglycerate	3-Phosphoglycerate kinase	2.0	2.0	2.0	–	–
Phosphoenolpyruvate → Pyruvate	Pyruvate kinase	1.0[a]	1.0[a]	1.0[a]	–	–
Pyruvate → Oxaloacetate	Oxaloacetate decarboxylase	–	–0.4[c]	–0.6[c]	–0.2[c]	–0.3[c]
Phosphoenolpyruvate → Oxaloacetate	Phosphoenolpyruvate carboxykinase (ATP-forming)		1.0	1.0		
Fumarate → Succinate	Fumarate reductase	–	0.8?[d]	2.4?[e]	0.4?[d]	1.2?[e]

Steps	Enzymes	ATP balances per substrate for different pathways				
		Glucose to 2 lactate	Glucose to 1.33 propionate + 0.66 acetate	Glucose + 2 H_2 to 2 propionate	Lactate to 0.66 propionate + 0.33 acetate	Lactate + H_2 to propionate
Methylmalonyl-CoA → Propionyl-CoA	Methylmalonyl-CoA decarboxylase	–	0.4[c]	0.6[c]	0.2[c]	0.3[c]
Acetyl-CoA → Acetate	Succinate CoA-transferase	–	0.66	–	0.33	–
Total ATP gain		**2.0**	**2.66 to 3.06**	**2.0 to 5.0**	**0.33 to 0.73**	**0 to 1.2**

[a]One phosphoenolpyruvate is used in the phosphotransferase step, leaving only one to generate ATP at the pyruvate kinase step.

[b]–, not needed in this pathway.

[c]Na^+ pumping step, written as ATP equivalents, assuming 10 Na^+ are used to generate 3 ATP (Meier et al., 2011). This would not be active if oxaloacetate is formed by phosphoenolpuyruvate carboxykinase.

[d]Uncertain, because enzyme for transfer of NADH from fermentation to the fumarate reductase was not able to be identified. Assuming a quinone cycle is present, this could translocate 2 H^+ at the fumarate reductase step, which could be converted to ATP via the ATP synthase, assuming 10 H^+ are used to generate 3 ATP (Meier et al., 2011).

[e]Potentially 4 H^+ translocated/generated at the fumarate reductase step, which could be converted to ATP via the ATP synthase, assuming 10 H^+ are used to generate 3 ATP (Meier et al., 2011).

Chapter 7 Summary, conclusions and ideas for further research

7.1 Rationale

Since 11000 to 9000 BC, when sheep were the first animals known to be domesticated for food use (Meadows, 2014), agriculture practices have kept improving. In countries like New Zealand, where a major part of the economy is ruled by agriculture, it is important to continuously monitor and improve agriculture practices and strategies. Ruminant-derived products contribute about 36% of the total value of NZ's exports. At the same time they also contribute towards greenhouse gas emissions, especially CH_4 gas, which accounts for 31.5% of NZ's total GHG emission, almost 6 times the global average CH_4 emissions from ruminants (Clark, 2009). In this process not only is CH_4 being released in the environment but also it is an approximately 8% loss of metabolic energy to the ruminants (Lassey et al. 1997). Researchers around the globe have been working on different strategies to reduce the CH_4 emissions from ruminants. Breeding low CH_4 ruminants has been a major part of that effort.

It has been shown that the CH_4 emission trait in sheep is a heritable and reproducible process (Pinares-Patiño et al., 2013b). Methane gas is emitted as by-product of rumen fermentation which is driven by microbes, mainly bacteria and methanogens present in the rumen. Rumen bacteria play a major role in feed fermentation and the formation of hydrogen (H_2) or formate, which are converted to CH_4 by the action of methanogens. Different fermentation pathways lead to different ratios of the major volatile fatty acids (acetate, propionate and butyrate) formed in the fermentation with different amounts of H_2 or formate being formed per unit of feed fermented. This leads to the hypothesises that the metabolism of dominant rumen bacteria associated with low CH_4 emitting sheep should explain the lower CH_4 yield, for example, by producing less H_2 or formate than bacteria associated with high CH_4 emitting sheep. To understand the influence of bacterial community on CH_4 emissions, and to see whether abundant bacterial groups present in low CH_4 emitting ruminants are involved in less H_2 or formate production, detailed study of the bacterial groups was needed. Kittelmann et al. (2014) identified two bacterial ruminotypes, Q and S, associated with low CH_4 emissions. Ruminotype Q showed a higher relative abundance of *Quinella* spp. while ruminotype S was mainly characterised

by *Sharpea* spp. None of these bacteria has been studied in detail before in context to their fermentation pathways. This was followed up by a metagenomic and metatranscriptomic study conducted by Kamke et al. (2016), who inferred that the rumen microbial community of *Sharpea*-enriched sheep produced lactate and butyrate as major end products which leads to lower CH_4 emissions from those sheep than in high CH_4 emitting sheep that have a rumen microbial community with fewer *Sharpea* that results in more H_2 production. To date, no work has been conducted on the mechanism in the Q-type community characterised by *Quinella* spp.

The aim of this thesis was to conduct a deeper examination of bacterial taxa associated with high and low CH_4 yields and provide a more accurate identification of bacterial groups associated with low CH_4 emitting sheep rumens. Strains of *Sharpea* and its close relative *Kandleria*, and concentrated *Quinella* samples were then used to understand how these low-CH_4 associated bacterial groups ultimately lead to lower CH_4 emissions.

7.2 Summary of results

Analysis of short reads of 16S rRNA gene sequences (provided by (Kittelmann et al., 2014) suggested that *Quinella* and *Sharpea* were associated with two low bacterial CH_4 community types (Q and S). In addition, there was a high CH_4-associated community, designated H-type, which appeared to be broadly similar to the Q- and S-types, but with *Quinella* or *Sharpea*. These data were re-examined in the work described here. First, the data were reanalysed using a more up-to-date bacterial taxonomic scheme. The significance of *Quinella* and *Sharpea* in the Q- and S-type communities was confirmed using different statistical tests, and then the data were interrogated using sPLS-DA (using representative sequences of OTUs at 97% sequence identity as input) to identify any additional taxa that differentiated these community types from each other. In addition to *Quinella* and *Sharpea*, seven other taxa were found to be associated with low CH_4 samples while ten taxa were found to be associated with high-CH_4 type samples (Table 3.3). It was notable that, in this OTU-level analysis, all 24 OTUs that separated the samples containing a Q-type community from other samples were assigned to the genus *Quinella*. In contrast, in samples containing the S-type community, of 16 discriminating OTUs, the seven most discriminating OTUs were affiliated to *Sharpea* (6 OTUs) and to the genus *Kandleria* (a closely related genus to *Sharpea*; 1 OTU).

To understand the species level diversity of these bacterial groups in sheep, first the phylogeny of these bacteria was analysed, using high-quality 16S rRNA gene sequences (both pre-existing and from newly formed clone libraries). The analysis suggested that *Quinella* may be represented by at least eight species level taxa (currently there is just one recognised species; *Quinella* ovalis) whereas *Sharpea* and *Kandleria* were only represented by one species each, *S. azabuensis* and the *K. vitulina* respectively. When this taxonomy was used to assess the *Quinella*, *Sharpea* and *Kandleria* diversity in sheep rumens by using 16S rRNA short reads generated by 454 pyrosequencing, *Quinella* was found to be more diverse than *Sharpea* and *Kandleria*. Sequences assigned to *Quinella* were grouped in 6 of the 8 species defined earlier, and 5 new clusters of short-read sequences that might be part of these species or new species. In contrast, short read sequences affiliated with the genus *Sharpea* formed an additional cluster (containing short reads only) as well as grouping with the type species, *S. azabuensis*. Sequences belonging to the genus *Kandleria* all grouped with the type species, *K. vitulina*. The association of *Sharpea* with low CH_4 sheep was explained by Kamke et al. (2016) but the role of *Quinella* in the rumen of low CH_4 sheep is still not understood. However, the volatile fatty acid profile in rumen samples of the sheep with the different community types provide a clue. The propionate to acetate ratios in Q-type community samples were greater than other community types, indicating that *Quinella* might be involved in propionate formation which would lead to less CH_4 formation. These finding are needed to be confirm by conducting more detailed studies of *Quinella*.

To understand the physiology of *Sharpea* and *Kandleria* associated with low CH_4 sheep rumen samples, four strains of each were used for analysis. In substrate utilisation tests it was found that most of the tested strains can grow equally well with glucose, fructose, galactose, sucrose, cellobiose and raffinose. D-Lactate was produced as the major end product with small amounts of formate, acetate and ethanol also detected. In co-culture experiments it was demonstrated that *Sharpea* and *Kandleria* didn't change their fermentation pattern when they grew with *M. olleyae* (a methanogen that can use both H_2 and/or formate), and lactate was still produced as the major end product. It was speculated that in co-culture, *Sharpea* and *Kandleria* might change their fermentation pattern from homolactic to mixed acid fermentation which would resulted in a decrease in production of lactate and an increase in production of acetate and perhaps ethanol, but this did not happened. This was a further confirmation of the Kamke et al. (2016) hypothesis, that in

the complex rumen environment *Sharpea* and *Kandleria* produce lactate as their major end product. Lactate is further converted to butyrate Kamke et al. (2016), resulting in less CH_4 production in comparison to the direct fermentation of carbohydrates to butyrate because the net H_2 yield is lower in the *Sharpea-Megasphaera* system.

Quinella, the diagnostic bacterium associated with the Q-type community, has been known for more than a century. However, no culture is available and its physiology is also not very clear. Orpin (1972) tried to isolate it but was not able to maintain it as pure culture. The first step towards understanding *Quinella* was to confirm its physical presence in sheep rumen samples in the study described in this thesis, as the association of *Quinella* with the Q-community type samples were based on short length 16S rRNA gene sequences generated by using 454 pyrosequencing. Hybridisation of a *Quinella*-specific FISH probe, designed in this study, suggested that *Quinella* cells are large and oval, which matches previously reported observations (Quin, 1943; Wicken and Howard, 1967b; Orpin, 1972) using standard microscopy methods. This was further confirmed using SEM and TEM. *Quinella* cells were 3 to 5 µm long and 1 to 2 µm in diameter. After an unsuccessful attempt to isolate a culture of *Quinella*, a cell concentration method was developed. Microscopy and clone library preparation of concentrated *Quinella* cells confirmed the marked increase in its relative abundance in these preparations. Metagenome sequencing was performed on DNA extracted from these concentrated *Quinella* samples. Assembly and genome binning performed on metagenomic sequences was successful, as out of 13 *Quinella* genomic bins, four appeared to be >90% complete and as little as 0.20% contaminated. Phylogenetic analysis of full length 16S rRNA gene sequences representing these genomic bins suggested that each of them may represent a different *Quinella* species. This finding was also supported by difference in G+C% content and coverage of these *Quinella* genomic bins.

On the basis of CAZyme analysis it appeared that none of these genome bins have genes for polysaccharides degradation, suggesting that *Quinella* might be dependent on other rumen microbes to degrade the polysaccharide components of feed. However, β-glucosidase and a range of PTS genes were found in the *Quinella* genome bins, which indicates that it uses cellobiose, cellodextrins, and other simple sugars to grow. Detailed analysis of genes for key enzymes involved in end product formation pathways suggested that *Quinella* may produce lactate perhaps use it as well. The L-lactate dehydrogenases of all four *Quinella* genome bins were highly similar to their homologue of *Selenomonas*

ruminantium, a close relative of *Quinella* that has been reported to produce and use lactate (Wallace, 1978; Asanuma and Hino, 2005). This somewhat strengthens the prediction of *Quinella* both being a lactate producer and user.

Quinella also possessed the randomizing pathway for propionate production. This finding is consistent with the VFA profiles of rumen samples containing Q-type communities, where greater ratios of propionate to acetate were observed than in samples containing the other community types. However, it appeared that *Quinella* uses succinyl CoA-transferase and succinyl-CoA synthetase to catalyse the succinate-dependent conversion of acetyl-CoA to acetate, an ATP yielding process, which hasn't been reported in bacteria previously but is present in helminths, protists and fungi. Notably, this might be happening in *S. ruminantium* as well, as the gene annotated as acetyl-CoA hydrolase in *S. ruminantium* genome might actually be a succinyl CoA-transferase. This conclusion was made on the basis of presence of a key amino acid that determines whether the enzyme functions as a hydrolase (D447: *Saccharomyces cerevisiae*) or transferase (E431: *Selenomonas ruminantium*; E432: *Quinella*). Furthermore, an unusual OACD structure was postulated in *Quinella*, by creating a hybrid of it unusual OACD α subunit with β, γ and δ subunits of MMDC.

Unexpectedly, all four *Quinella* genomic bins contained genes that appeared to code for NiFe hydrogen-uptake hydrogenases, which could feed electrons into the propionate-forming pathway. This suggests that *Quinella* might be able to take up hydrogen produced by other rumen microbes, which would result in increased propionate formation by *Quinella*. These findings explain how *Quinella* might be involved in lower CH_4 formation in the Q- type sheep rumen community. Overall, it appeared that there are five possible pathways for product formation, three of them using sugars as substrates while the other two use lactate. All of these pathways not only result in no H_2 production, but also the propionate formation may function as an electron sink and take up the available H_2 in the rumen when *Quinella* is growing with sugars and lactate. Overall, on the basis of gene sequences present in *Quinella* genome bins, it appeared that *Quinella* may produce lactate, acetate plus propionate, or just propionate as major end products, or some mix of these three schemes because all pathways could operate simultaneously with compromising electron balances.

These analyses also provided useful information which may help in future with attempts to isolate *Quinella*. The presence of a membrane bound uptake hydrogenase in all *Quinella* genome bins suggested that *Quinella* might grow well in presence of glucose and hydrogen. However, this could also lead to growth of some other hydrogen utilisers like methanogens as contaminants which could be prevented by using methanogen inhibitors.

7.3 Ideas for further research

Complete genome sequences of *Sharpea* and *Kandleria* are available through the Hungate1000 genome project (Creevey et al., 2014). However, even though the physiology of *Sharpea* and *Kandleria* has been studied in this project as well as by Kamke et al. (2016), detailed analysis of these genomes may reveal new features of these bacteria. Conducting CAZyme analysis may help in understanding whether they have the polysaccharide-degrading potential or, like *Quinella*, they rely on simple sugars released by other microbes for their growth. Small amounts of formate, acetate and ethanol were detected as fermentation products of *Sharpea* and *Kandleria*, and the roles of these products needs to be better understood. Furthermore, 16S rRNA gene based phylogenetic analysis of *Sharpea* and *Kandleria* did not show much diversity, so it would be interesting to conduct a comparative genomic analysis of these bacteria to find out whether there are any unique features at the species or strain level.

Quinella were found to be diverse in the present study based on almost full-length 16S rRNA gene sequences. Short read sequences of 16S rRNA genes indicated that there may be even more diversity in the genus *Quinella*. This could be investigated by generating more full length 16S rRNA gene clone sequences using rumen fluid samples from sheep that show higher relative abundance of *Quinella*. This could be best explored by sampling from a much wider range of ruminants on different diets, especially to understand if different species have any host or diet preferences.

This thesis was more focused on getting a first understanding of the physiology of *Quinella*. However, by using these genome bins many other analysis can be conducted to understand different features of *Quinella*. It would be interesting to study how *Quinella* maintains its remarkable large cell size. This might be revealed by identifying any unique gene(s) or pathways involve in cell cycle and cell division (Marshall et al., 2012).

Similarly, detailed analysis of *Quinella* genome bins may help in understanding different cell surface types observed in SEM and TEM images of *Quinella*.

The succinyl CoA-transferase and succinyl-CoA synthetase cycle of acetate and ATP formation needs further investigation, as it has not been reported in bacteria before. This can be done by either using pure cultures of *Quinella* or using *S. ruminantium*, which appeared to use the same cycle of acetate formation. Cloning and expression of succinyl CoA-transferase and succinyl-CoA synthetase gene from *Quinella* and *S. ruminantium* in *E. coli* may able provide answers.

Isolation attempt of *Quinella* could be attempted in presence of glucose and hydrogen together with methanogen inhibitors. The chances isolating pure *Quinella* cells may be increased by using rumen fluid samples from sheep with large *Quinella* populations. If successful, isolated *Quinella* cultures may be then used to assemble complete *Quinella* genomes and conduct comparative genome analyses. *Quinella* isolates may further be used for *in vitro* fermentation experiments to confirm the finding of this study. These include the use and production of lactate, and the uptake of hydrogen in the presence of sugars to increase the flow to the propionate-forming pathway

This study deduced *Quinella*'s physiology on the basis of genes detected in genome bins. Some of the conclusions could be tested by conducting a metatranscriptomics study of low CH_4 emitting sheep containing the Q-type community. This can be done in a similar way as performed by Kamke et al. (2016) who studied sheep that had S-type communities. The gene transcripts in those rumen samples could be mapped onto the *Quinella* genome bins to identify and estimate the transcript abundance. It would be interesting to determine to degree to which the lactate dehydrogenase and the uptake hydrogenase are expressed in these sheep.

To date, two low CH_4 associated bacterial community types, or ruminotypes (Q and S) have been identified in the rumens of sheep fed lucerne pellets. In this study, the relative abundance of *Quinella* 16S rRNA genes was as high as 74.4% of the total bacterial 16S rRNA genes detected. It would be interesting to repeat this experiment using high and low CH_4 emitting sheep fed on pasture and other diets. This will provide information whether community association with CH_4 yield are consistent or not across diets. It would also be interesting to see if the community types are stable through the lifespan of an animal. Studies conducted by Wallace et al. (2015) and Danielsson et al. (2017) observed

slight increases in the relative abundance of members of the family *Succinivibrionaceae* in low CH_4 emitting cattle using eight and 21 high and low CH_4 emitters respectively. This type of study needs to be conducted on a large scale using more animals and different ruminant species. Members of this family are also implicated in low CH_4 emissions from wallabies, producing succinate (Pope et al., 2011). Succinate is an intermediate in the randomising pathway of propionate formation, and so mechanisms similar to that proposed for the Q-type community may function in these animals but be carried out by phylogenetically different bacteria. It is not known to what extent the mechanisms proposed in this thesis for low CH_4 emissions in the Q and S type communities are applicable to other low CH_4 emitting animals across different diets, but hypotheses have been generated that can be tested.

References

Adams, D.W., and Errington, J. (2009) Bacterial cell division: Assembly, maintenance and disassembly of the Z ring. Nature Reviews Microbiology 7: 642-653.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z. et al. (2014) Binning metagenomic contigs by coverage and composition. Nature Methods 11: 1144-1146.

Alonso, J., Schimpl, M., and Van Aalten, D.M.F. (2014) O-GlcNAcase: Promiscuous hexosaminidase or key regulator of O-GlcNAc signaling? Journal of Biological Chemistry 289: 34433-34439.

Altermann, E. (2012) Tracing lifestyle adaptation in prokaryotic genomes. Frontiers in Microbiology 3: 48.

Altermann, E., Lu, J., and McCulloch, A. (2017) GAMOLA2, a comprehensive software package for the annotation and curation of draft and complete microbial genomes. Frontiers in Microbiology 8: 346

Amann, R.I., Krumholz, L., and Stahl, D.A. (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. Journal of Bacteriology 172: 762-770.

An, D., Dong, X., and Dong, Z. (2005) Prokaryote diversity in the rumen of yak *Bos grunniens* and Jinnan cattle *Bos taurus* estimated by 16S rDNA homology analyses. Anaerobe 11: 207-215.

Asanuma, N., and Hino, T. (2005) Ability to utilize lactate and related enzymes of a ruminal bacterium, *Selenomonas ruminantium*. Animal Science Journal 76: 345-352.

Asanuma, N., Ishiwata, M., Yoshii, T., Kikuchi, M., Nishina, Y., and Hino, T. (2005) Characterization and transcription of the genes involved in butyrate production in *Butyrivibrio fibrisolvens* type I and II strains. Current Microbiology 51: 91-94.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455-477.

Belenguer, A., Toral, P.G., Frutos, P., and Hervás, G. (2010) Changes in the rumen bacterial community in response to sunflower oil and fish oil supplements in the diet of dairy sheep. Journal of Dairy Science 93: 3275-3286.

Benning, M.M., Haller, T., Gerlt, J.A., and Holden, H.M. (2000) New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from *Escherichia coli*. Biochemistry 39: 4630-4639.

Berks, B.C., Page, M.D., Richardson, D.J., Reilly, A., Cavill, A., Outen, F., and Ferguson, S.J. (1995) Sequence analysis of subunits of the membrane-bound nitrate reductase from a denitrifying bacterium: the integral membrane subunit provides a prototype for the dihaem electron-carrying arm of a redox loop. Molecular Microbiology 15: 319-331.

Biegel, E., Schmidt, S., and Müller, V. (2009) Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*. Environmental Microbiology 11: 1438-1443.

Bladen, H.A., Bryant, M.P., and Doetsch, R.N. (1961) A study of bacterial species from the rumen which produce ammonia from protein hydrolyzate. Applied Microbiology 9: 175-180.

Boone, D.R., Johnson, R.L., and Liu, Y. (1989) Diffusion of the interspecies electron carriers H_2 and formate in methanogenic ecosystems and its implications in the measurement of K_m for H_2 or formate uptake. Applied and Environmental Microbiology 55: 1735-1741.

Bott, M., Pfister, K., Burda, P., Kalbermatter, O., Woehlke, G., and Dimroth, P. (1997) Methylmalonyl-CoA decarboxylase from *Propionigenium modestum*. European Journal of Biochemistry 250: 590-599.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A.J. et al. (2016) Uniprotkb/swiss-prot, the manually annotated section of the uniprot knowledgebase: How to use the entry view. In: Edwards D. (ed) Plant Bioinformatics. Methods in Molecular Biology, vol 1374. New York, NY, USA: Humana Press, pp. 23-54.

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G.W. (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. Nat Meth 9: 425-426.

Brough, E.B., Reid, T.C., and Howard, B.H. (1970) The biochemistry of the rumen bacterium "Quin's oval". New Zealand Journal of Science 13:570-575.

Brown, A.T., and Patterson, C.E. (1973) Ethanol production and alcohol dehydrogenase activity in *Streptococcus mutans*. Archives of Oral Biology 18: 127-131.

Brulc, J.M., Antonopoulos, D.A., Miller, M.E., Wilson, M.K., Yannarell, A.C., Dinsdale, E.A. et al. (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proceedings of the National Academy of Sciences USA 106: 1948-1953.

Bryant, M.P., Small, N., Bouma, C., and Robinson, I. (1958) Studies on the composition of the ruminal flora and fauna of young calves. Journal of Dairy Science 41: 1747-1767.

Bryant, M.P., Wolin, E.A., Wolin, M.J., and Wolfe, R.S. (1967) *Methanobacillus omelianskii*, a symbiotic association of two species of bacteria. Archiv für Mikrobiologie 59: 20-31.

Buckel, W. (2001) Sodium ion-translocating decarboxylases. Biochimica et Biophysica Acta 1505: 15-27.

Buckel, W., and Thauer, R.K. (2013) Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Biochimica et Biophysica Acta 1827: 94-113.

Buddle, B.M., Denis, M., Attwood, G.T., Altermann, E., Janssen, P.H., Ronimus, R.S. et al. (2011) Strategies to reduce methane emissions from farmed ruminants grazing on pasture. Veterinary Journal 188: 11-17.

Caldwell, D.R., and Bryant, M.P. (1966) Medium without rumen fluid for nonselective enumeration and isolation of rumen bacteria. Applied Microbiology 14: 794-801.

Callaway, E., and Martin, S. (1997) Effects of a *Saccharomyces cerevisiae* culture on ruminal bacteria that utilize lactate and digest cellulose. Journal of Dairy Science 80: 2035-2044.

Calsamiglia, S., Busquet, M., Cardozo, P.W., Castillejos, L., and Ferret, A. (2007) Invited review: Essential oils as modifiers of rumen microbial fermentation. Journal of Dairy Science 90: 2580-2595.

Cao, K.A.L., Costello, M.E., Lakis, V.A., Bartolo, F., Chua, X.Y., Brazeilles, R., and Rondeau, P. (2016) MixMC: A multivariate statistical framework to gain insight into microbial communities. PLoS One 11: e0160169.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335-336.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A. et al. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Research 42: D459-D471.

Cerrilla, M.E.O., and Martínez, G.M. (2003) Starch digestion and glucose metabolism in the ruminant: A review. Interciencia 28: 380-386.

Chen, X., Schreiber, K., Appel, J., Makowka, A., Fähnrich, B., Roettger, M. et al. (2016) The Entner–Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. Proceedings of the National Academy of Sciences USA 113: 5441-5446.

Chesson, A., and Forsberg, C.W. (1997) Polysaccharide degradation by rumen microorganisms. In: The Rumen Microbial Ecosystem. Hobson, P.N., and Stewart, C.S. (eds). Dordrecht, The Netherlands: Springer, pp. 329-381.

Chin, K.-J., and Janssen, P.H. (2002) Propionate formation by *Opitutus terrae* in pure culture and in mixed culture with a hydrogenotrophic methanogen and implications for carbon fluxes in anoxic rice paddy soil. Applied and Environmental Microbiology 68: 2089-2092.

Clark, H. (2009) Methane emissions from ruminant livestock; are they important and can we reduce them? Proceedings of the New Zealand Grassland Association 71: 73-76.

Clarke, R.T.J. (1979) Niche in pasture-fed ruminants for the large rumen bacteria *Oscillospira*, *Lampropedia*, and Quin's and Eadie's ovals. Applied and Environmental Microbiology 37: 654-657.

Cloud-Hansen, K.A., Peterson, S.B., Stabb, E.V., Goldman, W.E., McFall-Ngai, M.J., and Handelsman, J. (2006) Breaching the great wall: Peptidoglycan and microbial interactions. Nature Reviews Microbiology 4: 710-716.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S. et al. (2003) The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Research 31: 442-443.

Coleman, G.S. (1985) The cellulase content of 15 species of entodiniomorphid protozoa, mixed bacteria and plant debris isolated from the ovine rumen. Journal of Agricultural Science 104: 349-360.

Conrad, R., Schink, B., and Phelps, T.J. (1986) Thermodynamics of H_2-consuming and H_2-producing metabolic reactions in diverse methanogenic environments under in situ conditions. FEMS Microbiology Letters 38: 353-360.

Cottle, D.J., Nolan, J.V., and Wiedemann, S.G. (2011) Ruminant enteric methane mitigation: a review. Animal Production Science 51: 491-514.

Crable, B.R., Plugge, C.M., McInerney, M.J., and Stams, A.J.M. (2011) Formate formation and formate conversion in biological fuels production. Enzyme Research 2011: 532536.

Creevey, C.J., Kelly, W.J., Henderson, G., and Leahy, S.C. (2014) Determining the culturability of the rumen bacterial microbiome. Microbial Biotechnology 7: 467–479.

Cronín, C.N., Nolan, D.P., and Paul Voorheis, H. (1989) The enzymes of the classical pentose phosphate pathway display differential activities in procyclic and bloodstream forms of *Trypanosoma brucei*. FEBS Letters 244: 26-30.

Dahinden, P., Pos, K.M., and Dimroth, P. (2005) Identification of a domain in the α-subunit of the oxaloacetate decarboxylase Na^+ pump that accomplishes complex formation with the γ-subunit. FEBS Journal 272: 846-855.

Danielsson, R., Dicksved, J., Sun, L., Gonda, H., Müller, B., Schnürer, A., and Bertilsson, J. (2017) Methane production in dairy cows correlates with rumen methanogenic and bacterial community structure. Frontiers in Microbiology 8: 226.

de Haas, Y., Pszczola, M., Soyeurt, H., Wall, E., and Lassen, J. (2017) Invited review: Phenotypes to genetically reduce greenhouse gas emissions in dairying. Journal of Dairy Science 100: 855-870.

Deng, W., Wanapat, M., Ma, S., Chen, J., Xi, D., He, T. et al. (2007) Phylogenetic analysis of 16S rDNA sequences manifest rumen bacterial diversity in gayals (*Bos frontalis*) fed fresh bamboo leaves and twigs (*Sinarumdinaria*). Asian-Australasian Journal of Animal Sciences 20: 1057-1066.

Denger, K., and Schink, B. (1992) Energy conservation by succinate decarboxylation in *Veillonella parvula*. Microbiology 138: 967-971.

Denman, S. (2016) The application of 'omic' technologies to understand low methane animal gut systems. Unpublished conference presentation, 6th GGAA conference. Melbourne, Australia.

Denman, S.E., Martinez, F.G., Shinkai, T., Mitsumori, M., and McSweeney, C.S. (2015) Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog. Frontiers in Microbiology 6: 1087.

Di Bella, J.M., Bao, Y., Gloor, G.B., Burton, J.P., and Reid, G. (2013) High throughput sequencing methods and analysis for microbiome research. Journal of Microbiological Methods 95: 401-414.

Di Berardino, M., and Dimroth, P. (1996) Aspartate 203 of the oxaloacetate decarboxylase beta-subunit catalyses both the chemical and vectorial reaction of the Na^+ pump. EMBO Journal 15: 1842-1849.

Dimroth, P., and Thomer, A. (1983) Subunit composition of oxaloacetate decarboxylase and characterization of the α chain as carboxyltransferase. European Journal of Biochemistry 137: 107-112.

Dimroth, P., Kaim, G., and Matthey, U. (1998) The motor of the ATP synthase. Biochimica et Biophysica Acta 1365: 87-92.

Dimroth, P., Jockel, P., and Schmid, M. (2001) Coupling mechanism of the oxaloacetate decarboxylase Na^+ pump. Biochimica et Biophysica Acta 1505: 1-14.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27: 2194-2200.

Elbourne, Liam D H., Tetu, S.G., Hassan, K.A., and Paulsen, I.T. (2017) TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Research 45: D320-D324.

Engelbrecht, S., and Junge, W. (1997) ATP synthase: A tentative structural model. FEBS Letters 414: 485-491.

Everett, K.D.E., Bush, R.M., and Andersen, A.A. (1999) Emended description of the order *Chlamydiales*, proposal of *Parachlamydiaceae* fam. nov. and *Simkaniaceae* fam. nov., each containing one monotypic genus, revised taxonomy of the family *Chlamydiaceae*, including a new genus and five new species, and standards for the identification of organisms. International Journal of Systematic and Evolutionary Microbiology 49: 415-440.

Ferry, J. (2015) Acetate metabolism in anaerobes from the domain archaea. Life 5: 1454-1571.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L. et al. (2016) The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Research 44: D279-D285.

Flanagan, Lindsey A., and Parkin, A. (2016) Electrochemical insights into the mechanism of NiFe membrane-bound hydrogenases. Biochemical Society Transactions 44: 315-328.

Forsberg, C.W., Cheng, K.J., and White, B.A. (1997) Polysaccharide degradation in the rumen and large intestine. In: Gastrointestinal Microbiology: Volume 1 Gastrointestinal Ecosystems and Fermentations. Mackie, R.I., and White, B.A. (eds). Boston, MA, USA: Springer, pp. 319-379.

Fraser, M.E., James, M.N.G., Bridger, W.A., and Wolodko, W.T. (1999) A detailed structural description of *Escherichia coli* succinyl-CoA synthetase. Journal of Molecular Biology 285: 1633-1653.

Fritsch, J., Scheerer, P., Frielingsdorf, S., Kroschinsky, S., Friedrich, B., Lenz, O., and Spahn, C.M.T. (2011) The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. Nature 479: 249-252.

González, I., Cao, K.A.L., Davis, M.J., and Déjean, S. (2012) Visualising associations between paired 'omics' data sets. BioData Mining 5. 19.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R. et al. (2014) Human genetics shape the gut microbiome. Cell 159: 789-799.

Goopy, J.P., Donaldson, A., Hegarty, R., Vercoe, P.E., Haynes, F., Barnett, M., and Oddy, V.H. (2014) Low-methane yield sheep have smaller rumens and shorter rumen retention time. British Journal of Nutrition 111: 578-585.

Goopy, J.P., Robinson, D.L., Woodgate, R.T., Donaldson, A.J., Oddy, V.H., Vercoe, P.E., and Hegarty, R.S. (2016) Estimates of repeatability and heritability of methane production in sheep using portable accumulation chambers. Animal Production Science 56: 116-122.

Goris, T., Wait, A.F., Saggu, M., Fritsch, J., Heidary, N., Stein, M. et al. (2011) A unique iron-sulfur cluster is crucial for oxygen tolerance of a [NiFe]-hydrogenase. Nature Chemical Biology 7: 310-318.

Gottschalk, G. (1986) Bacterial Metabolism, 2nd edition. New York, NY, USA: Springer-Verlag.

Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B. et al. (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. ISME Journal 10: 761-777.

Greuter, D., Loy, A., Horn, M., and Rattei, T. (2016) probeBase—an online resource for rRNA-targeted oligonucleotide probes and primers: new features 2016. Nucleic Acids Research 44: D586-D589.

Gronow, S., Welnitz, S., Lapidus, A., Nolan, M., Ivanova, N., Glavina Del Rio, T. et al. (2010) Complete genome sequence of *Veillonella parvula* type strain ($Te3^T$). Standards in Genomic Science 2: 57-65.

Gross, R., Simon, J., Lancaster, C.R.D., and Kröger, A. (1998) Identification of histidine residues in *Wolinella succinogenes* hydrogenase that are essential for menaquinone reduction by H_2. Molecular Microbiology 30: 639-646.

Grüber, G., Manimekalai, M.S.S., Mayer, F., and Müller, V. (2014) ATP synthases from archaea: The beauty of a molecular motor. Biochimica et Biophysica Acta 1837: 940-952.

Gruninger, R.J., Sensen, C.W., McAllister, T.A., and Forster, R.J. (2014) Diversity of rumen bacteria in Canadian cervids. PLoS One 9: e89682.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013) QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29: 1072-1075.

Haft, D.H., Selengut, J.D., and White, O. (2003) The TIGRFAMs database of protein families. Nucleic Acids Research 31: 371-373.

Hägerhäll, C., and Hederstedt, L. (1996) A structural moDAl for the membrane-integral domain of succinate:quinone oxidoreductases. FEBS Letters 389: 25-31.

Hayes, B.J., Donoghue, K.A., Reich, C.M., Mason, B.A., Bird-Gardiner, T., Herd, R.M., and Arthur, P.F. (2016) Genomic heritabilities and genomic estimated breeding values for methane traits in Angus cattle. Journal of Animal Science 94: 902-908.

He, X., McLean, J.S., Edlund, A., Yooseph, S., Hall, A.P., Liu, S.Y. et al. (2015) Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proceedings of the National Academy of Sciences USA 112: 244-249.

Hederstedt, L. (1999) Respiration without O_2. Science 284: 1941-1942.

Hegarty, R.S., Alcock, D., Robinson, D.L., Goopy, J.P., and Vercoe, P.E. (2010) Nutritional and flock management options to reduce methane output and methane per unit product from sheep enterprises. Animal Production Science 50: 1026-1033.

Heider, J. (2001) A new family of CoA-transferases. FEBS Letters 509: 345-349.

Henderson, G., Cook, G.M., and Ronimus, R.S. (2016) Enzyme- and gene-based approaches for developing methanogen-specific compounds to control ruminant methane emissions: a review. Animal Production Science 2016: AN15757.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Global Rumen Census Collaborators, and Janssen, P.H. (2015) Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Scientific Reports 5: 14567.

Henderson, G., Yilmaz, P., Forster, R.J., Kelly, W.J., Leahy, S.C., Kumar, S. et al. (2017) Improved taxonomic assignment of rumen bacterial 16S rRNA sequences using a revised SILVA taxonomic framework. PLoS One, submitted.

Henikoff, S., and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences USA 89: 10915-10919.

Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochemical Journal 293: 781-788.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G. et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463-467.

Hilpert, W., and Dimroth, P. (1983) Purification and characterization of a new sodium-transport decarboxylase. European Journal of Biochemistry 132: 579-587.

Hilpert, W., and Dimroth, P. (1991) On the mechanism of sodium ion translocation by methylmalonyl-CoA decarboxylase from *Veillonella alcalescens*. European Journal of Biochemistry 195: 79-86.

Hobson, P.N., and Stewart, C.S. (1997) The Rumen Microbial Ecosystem, 2nd edition. New York, NY, USA: Chapman & Hall.

Hoffmann, A., Hilpert, W., and Dimroth, P. (1989) The carboxyltransferase activity of the sodium-ion-translocating methylmalonyl-CoA decarboxylase of *Veillonella alcalescens*. European Journal of Biochemistry 179: 645-650.

Horgan, R.P., and Kenny, L.C. (2011) 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. Obstetrician and Gynaecologist 13: 189-195.

Hristov, A., Ahvenjarvi, S., McAllister, T., and Huhtanen, P. (2003) Composition and digestive tract retention time of ruminal particles with functional specific gravity greater or less than 1.02. Journal of Animal Science 81: 2639-2648.

Hristov, A.N., Oh, J., Giallongo, F., Frederick, T.W., Harper, M.T., and Weeks, H.L. (2015) An inhibitor persistently decreased enteric methane emission from dairy cows with no negative effect on milk production. Proceedings of the National Academy of Science USA 112: E5218.

Hristov, A.N., Ott, T., Tricarico, J., Rotz, A., Waghorn, G., Adesogan, A. et al. (2013) Special topics - Mitigation of methane and nitrous oxide emissions from animal operations: III. A review of animal management mitigation options. Journal of Animal Science 91: 5095-5113.

Hua, Z.S., Han, Y.J., Chen, L.X., Liu, J., Hu, M., Li, S.J. et al. (2015) Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. ISME Journal 9: 1280-1294.

Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20: 2317-2319.

Huder, J.B., and Dimroth, P. (1993) Sequence of the sodium ion pump methylmalonyl-CoA decarboxylase from *Veillonella parvula*. Journal of Biological Chemistry 268: 24564-24571.

Huder, J.B., and Dimroth, P. (1995) Expression of the sodium ion pump methylmalonyl-coenzyme A-decarboxylase from *Veillonella parvula* and of mutated enzyme specimens in *Escherichia coli*. Journal of Bacteriology 177: 3623-3630.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J. et al. (2016) A new view of the tree of life. Nature Microbiology 2016: 16048.

Hugenholtz, P., Tyson, G.W., and Blackall, L.L. (2001) Design and evaluation of16 rRNA-targeted oligonucleotide probes for fluorescence in situ hybridisation In:

Gene Probes: Principles and Protocols. Aquino de Muro, M. and Rapley, R. (eds). London, UK: Humana Press, pp. 29–42.

Huhtanen, C., and Gall, L. (1953a) Rumen organisms I.: Curved rods and a related rod type. Journal of Bacteriology 65: 548.

Huhtanen, C., and Gall, L. (1953b) Rumen organisms II.: Two lactate utilizers and six miscellaneous types. Journal of Bacteriology 65: 554.

Hungate, R. (1950) The anaerobic mesophilic cellulolytic bacteria. Bacteriological Reviews 14: 1-149.

Hungate, R. (1975) The rumen microbial ecosystem. Annual Review of Ecology and Systematics 6: 39-66.

Hungate, R., Dougherty, R., Bryant, M., and Cello, R. (1952) Microbiological and physiological changes associated with acute indigestion in sheep. Cornell Veterinarian 42: 423-449.

Hungate, R., Smith, W., Bauchop, T., Yu, I., and Rabinowitz, J. (1970) Formate as an intermediate in the bovine rumen fermentation. Journal of Bacteriology 102: 389-397.

Hungate, R.E. (1947) Studies on cellulose fermentation. III. The culture and isolation of cellulose-decomposing bacteria from the rumen of cattle. Journal of Bacteriology 53. 631-645.

Hungate, R.E. (1966) The Rumen and its Microbes. New York, NY, USA: Academic Press.

Huse, S.M., Welch, D.M., Morrison, H.G., and Sogin, M.L. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology 12: 1889-1898.

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.

Iannotti, E.L., Kafkewitz, D., Wolin, M.J., and Bryant, M.P. (1973) Glucose fermentation products of *Ruminococcus albus* grown in continuous culture with

Vibrio succinogenes: changes caused by interspecies transfer of H_2. Journal of Bacteriology 114: 1231-1240.

Imelfort, M., Parks, D., Woodcroft, B.J., Dennis, P., Hugenholtz, P., and Tyson, G.W. (2014) GroopM: An automated tool for the recovery of population genomes from related metagenomes. PeerJ 2014.

International, VSN. (2015) GenStat for Windows. Hemel Hempstead, UK: VSN International.

Ito, M., Guffanti, A.A., Zemsky, J., Ivey, D.M., and Krulwich, T.A. (1997) Role of the nhaC-encoded Na^+/H^+ antiporter of alkaliphilic *Bacillus firmus* OF4. Journal of Bacteriology 179: 3851-3857.

Iverson, T.M., Luna-Chavez, C., Cecchini, G., and Rees, D.C. (1999) Structure of the *Escherichia coli* fumarate reductase respiratory complex. Science 284: 1961-1966.

Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V. (2012) Untangling genomes from metagenomes: Revealing an uncultured class of marine euryarchaeota. Science 335: 587-590.

Janssen, P.H. (2010) Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. Animal Feed Science and Technology 160: 1-22.

Janssen, P.H., and Kirs, M. (2008) Structure of the archaeal community of the rumen. Applied and Environmental Microbiology 74: 3619-3625.

Joblin, K.N. (1999) Ruminal acetogens and their potential to lower ruminant methane emissions. Crop and Pasture Science 50: 1307-1314.

Jockel, P., Di Berardino, M., and Dimroth, P. (1999) Membrane topology of the β-subunit of the oxaloacetate decarboxylase Na^+ pump from *Klebsiella pneumoniae*. Biochemistry 38: 13461-13472.

Jockel, P., Schmid, M., Steuber, J., and Dimroth, P. (2000a) A molecular coupling mechanism for the oxaloacetate decarboxylase Na^+ pump as inferred from mutational analysis. Biochemistry 39: 2307-2315.

Jockel, P., Schmid, M., Choinowski, T., and Dimroth, P. (2000b) Essential role of tyrosine 229 of the oxaloacetate decarboxylase β-subunit in the energy coupling mechanism of the Na^+ pump. Biochemistry 39: 4320-4326.

Johnson, B.C., Hamilton, T.S., Robinson, W., and Garey, J. (1944) On the mechanism of non-protein-nitrogen utilization by ruminants. Journal of Animal Science 3: 287-298.

Jonker, A., Sandoval, S., Boma, P., Hickey, S., McEwan, J.C., Janssen, P.H., and Rowe, S. (2017a) Sheep selected for divergent methane yield on lucerne pellets also express the same trait when fed fresh pasture. British Society of Animal Sciences Annual Conference, University of Chester, UK, 26 - 27 April 2017, abstract 106.

Jonker, A., Hickey, S., McEwan, J., Piares-Patiño, C., Olinga, S., Dias, A. et al. (2017b) Sheep from low methane yield selection lines created on alfalfa pellets also have lower methane yield under pastoral farming conditions. Journal of Animal Science. doi:10.2527/jas2017-1709.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004) A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology 338: 1027-1036.

Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S. et al. (2016) Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a *Sharpea*-enriched microbiome characterised by lactic acid formation and utilisation. Microbiome 4: 56.

Kamra, D.N. (2005) Rumen microbial ecosystem. Current Science 89: 124-135.

Kanegasaki, S., and Takahashi, H. (1967) Function of growth factors for rumen microorganisms. I. Nutritional characteristics of *Selenomonas ruminantium*. Journal of Bacteriology 93: 456-463.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology 428: 726-731.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45: D353-D361.

Kaneko, J., Yamada-Narita, S., Abe, N., Onodera, T., Kan, E., Kojima, S. et al. (2015) Complete genome sequence of *Selenomonas ruminantium* subsp. *lactilytica* will accelerate further understanding of the nature of the class *Negativicutes*. FEMS Microbiology Letters 362: fnv050.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3: e1165.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S., and Janssen, P.H. (2011) Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. Journal of Microbiological Methods 84:52-60.

Kerepesi, C., Bánky, D., and Grolmusz, V. (2014) AmphoraNet: The webserver implementation of the AMPHORA2 metagenomic workflow suite. Gene 533: 538-540.

Kim, M., Morrison, M., and Yu, Z. (2011) Phylogenetic diversity of bacterial communities in bovine rumen as affected by diets and microenvironments. Folia Microbiologica 56: 453-458.

Kim, M., Oh, H.S., Park, S.C., and Chun, J. (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 64: 346-351.

King, E.E., Smith, R.P., St Pierre, B., and Wright, A.D.G. (2011) Differences in the rumen methanogen populations of lactating Jersey and Holstein dairy cows under the same diet regimen. Applied and Environmental Microbiology 77: 5682-5687.

Kittelmann, S., Pinares-Patiño, C.S., Seedorf, H., Kirk, M.R., Ganesh, S., McEwan, J.C., and Janssen, P.H. (2014) Two different bacterial community types are linked with the low-methane emission trait in sheep. PLoS One 9: e103171.

Klieve, A.V., and Swain, R.A. (1993) Estimation of ruminal bacteriophage numbers by pulsed-field gel electrophoresis and laser densitometry. Applied and Environmental Microbiology 59: 2299-2303.

Klimchuk, O.I., Dibrova, D.V., and Mulkidjanian, A.Y. (2016) Phylogenomic analysis identifies a sodium-translocating decarboxylating oxidoreductase in thermotogae. Biochemistry (Moscow) 81: 481-490.

Knappe, J., Blaschkowski, H.P., Grobner, P., and Schmttt, T. (1974) Pyruvate formate-lyase of *Escherichia coli*: the acetyl-enzyme intermediate. European Journal of Biochemistry 50: 253-263.

Koike, S., Handa, Y., Goto, H., Sakai, K., Miyagawa, E., Matsui, H. et al. (2010) Molecular monitoring and isolation of previously uncultured bacterial strains from the sheep rumen. Applied and Environmental Microbiology 76: 1887-1894.

Krause, D.O., Nagaraja, T.G., Wright, A.D., and Callaway, T.R. (2013) Rumen microbiology: leading the way in microbial ecology. Journal of Animal Science 91: 331-341.

Krause, D.O., Denman, S.E., Mackie, R.I., Morrison, M., Rae, A.L., Attwood, G.T., and McSweeney, C.S. (2003) Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. FEMS Microbiology Reviews 27: 663-693.

Kröger, A. (1978) Fumarate as terminal acceptor of phosphorylative electron transport. Biochimica et Biophysica Acta 505: 129-145.

Kröger, A., Geisler, V., Lemma, E., Theis, F., and Lenger, R. (1992) Bacterial fumarate respiration. Archives of Microbiology 158: 311-314.

Krumholz, L.R., Bryant, M.P., Brulla, W.J., Vicini, J.L., Clark, J.H., and Stahl, D.A. (1993) Proposal of *Quinella ovalis* gen. nov., sp. nov., based on phylogenetic analysis. . International Journal of Systematic Bacteriology 43: 393-296.

Kruskal, W.H., Wallis, W.A. (1952) Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 47:583-621.

Kumar, G.K., Bahler, C.R., Wood, H.G., and Merrifield, R.B. (1982) The amino acid sequences of the biotinyl subunit essential for the association of transcarboxylase. Journal of Biological Chemistry 257: 13828-13834.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33: 1870-1874.

Kumar, S., Choudhury, P.K., Carro, M.D., Griffith, G.W., Dagar, S.S., Puniya, M. et al. (2014) New aspects and strategies for methane mitigation from ruminants. Applied and Microbiology and Biotechnology 98: 31-44.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environmental Microbiology 12: 118-123.

Lancaster, C.R.D. (2003) *Wolinella succinogenes* quinol:fumarate reductase and its comparison to *E. col*i succinate:quinone reductase. FEBS Letters 555: 21-28.

Lancaster, C.R.D., and Kröger, A. (2000) Succinate: quinone oxidoreductases: new insights from X-ray crystal structures. Biochimica et Biophysica Acta 1459: 422-431.

Lancaster, C.R.D., Kröger, A., Auer, M., and Michel, H. (1999) Structure of fumarate reductase from *Wolinella succinogenes* at 2.2Å resolution. Nature 402: 377-385.

Lapage, S.P., Sneath, P.H.A., Lessel, E.F., Skerman, V.B.D., Seeliger, H.P.R., and Clark, W.A. (1992) International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision. Washington, DC, USA: ASM Press.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

Lassen, J., Poulsen, N.A., Larsen, M.K., and Buitenhuis, A.J. (2016) Genetic and genomic relationship between methane production measured in breath and fatty acid content in milk samples from Danish Holsteins. Animal Production Science 56: 298-303.

Lassey, K.R., Ulyatt, M.J., Martin, R.J., Walker, C.F., and Shelton, I.D. (1997) Methane emissions measured directly from grazing livestock in New Zealand. Atmospheric Environment 31: 2905-2914.

Laussermair, E., Schwarz, E., Oesterhelt, D., Reinke, H., Beyreuther, K., and Dimroth, P. (1989) The sodium ion translocating oxaloacetate decarboxylase of *Klebsiella pneumoniae*. Sequence of the integral membrane-bound subunits β and γ. Journal of Biological Chemistry 264: 14710-14715.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P., Ramey, R.R., Bircher, J.S. et al. (2008) Evolution of mammals and their gut microbes. Science 320: 1647-1651.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015) MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31: 1674-1676.

Li, F., and Guan, L.L. (2017) Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle. Applied and Environmental Microbiology 83: e00061-17.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research 13: 2178-2189.

Li, Z., Wright, A.D.G., Si, H., Wang, X., Qian, W., Zhang, Z., and Li, G. (2016) Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. Environmental Microbiology Reports 8: 1016-1023.

Lin, H.H., and Liao, Y.C. (2016) Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. Scientific Reports 6: 24175.

Lloyd, J.S., Weston, A., and Cardy, D.L.N. (2006) Method for detecting nucleic acid target sequences involving in vitro transcription from an RNA polymerase promoter. Patent WO 1999037805 A1. URL https://encrypted.google.com/patents/WO1999037805A1. Accessed 29 August 2017.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 42: D490-D495.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, A. et al. (2004) ARB: A software environment for sequence data. Nucleic Acids Research 32: 1363-1371.

Lukey, M.J., Roessler, M.M., Parkin, A., Evans, R.M., Davies, R.A., Lenz, O. et al. (2011) Oxygen-tolerant [NiFe]-hydrogenases: The individual and collective importance of supernumerary cysteines at the proximal Fe-S cluster. Journal of the American Chemical Society 133: 16881-16892.

Mack, M., and Buckel, W. (1997) Conversion of glutaconate CoA-transferase from *Acidaminococcus fermentans* into an acyl-CoA hydrolase by site-directed mutagenesis. FEBS Letters 405: 209-212.

Mackie, R.I., Aminov, R.I., Hu, W., Klieve, A.V., Ouwerkerk, D., Sundset, M.A., and Kamagata, Y. (2003) Ecology of uncultivated *Oscillospira* species in the rumen of cattle, sheep, and reindeer as assessed by microscopy and molecular approaches. Applied and Environmental Microbiology 69: 6808-6815.

Maklashina, E., Berthold, D.A., and Cecchini, G. (1998) Anaerobic expression of *Escherichia coli* succinate dehydrogenase: functional replacement of fumarate reductase in the respiratory chain during anaerobic growth. Journal of Bacteriology 180: 5989-5996.

Malmuthuge, N., Li, M., Fries, P., Griebel, P.J., and Guan, L.L. (2012) Regional and age dependent changes in gene expression of Toll-like receptors and key antimicrobial defence molecules throughout the gastrointestinal tract of dairy calves. Veterinary Immunology and Immunopathology 146: 18-26.

Mao, S., Zhang, M., Liu, J., and Zhu, W. (2015) Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. Scientific Reports 5: 16116.

Marchandin, H., Teyssier, C., Campos, J., Jean-Pierre, H., Roger, F., Gay, B. et al. (2010) *Negativicoccus succinicivorans* gen. nov., sp. nov., isolated from human clinical samples, emended description of the family *Veillonellaceae* and description of *Negativicutes* classis nov., *Selenomonadales* ord. nov. and *Acidaminococcaceae* fam. nov. in the bacterial phylum *Firmicutes*. International Journal of Systematic and Evolutionary Microbiology 60: 1271-1279.

Marshall, W.F., Young, K.D., Swaffer, M., Wood, E., Nurse, P., Kimura, A. et al. (2012) What determines cell size? BMC Biology 10:101.

Martin, C., Morgavi, D.P., and Doreau, M. (2010) Methane mitigation in ruminants: from microbe to the farm scale. Animal 4: 351-365.

Marvin-Sikkema, F.D., Pedro Gomes, T.M., Grivet, J.P., Gottschal, J.C., and Prins, R.A. (1993) Characterization of hydrogenosomes and their role in glucose metabolism of *Neocallimastix* sp. L2. Archives of Microbiology 160: 388-396.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A. et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME Journal 6: 610-618.

McInerney, M.J., Struchtemeyer, C.G., Sieber, J., Mouttaki, H., Stams, A.J.M., Schink, B. et al. (2008) Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. Annals of the New York Academy of Sciences 1125: 58-72.

Meadows, J.R.S. (2014) Sheep: Domestication. In Encyclopedia of Global Archaeology. Smith, C. (ed). New York, NY, USA: Springer, pp. 6597-6600.

Meek, L., and Arp, D.J. (2000) The hydrogenase cytochrome b heme ligands of *Azotobacter vinelandii* are required for full H_2 oxidation capability. Journal of Bacteriology 182: 3429-3436.

Meier, T., Faraldo-Gómez, J., and Börsch, M. (2011) ATP synthase – A paradigmatic molecular machine. In: Molecular Machines in Biology. Frank J (ed). Cambridge, UK: Cambridge University Press, pp. 208-238.

Menon, N.K., Robbins, J., Wendt, J.C., Shanmugam, K.T., and Przybyla, A.E. (1991) Mutational analysis and characterization of the *Escherichia coli* hya operon, which encodes [NiFe] hydrogenase 1. Journal of Bacteriology 173: 4851-4861.

Michel, T.A., and Macy, J.M. (1990) Purification of an enzyme responsible for acetate formation from acetyl coenzyme A in *Selenomonas ruminatium*. FEMS Microbiology Letters 68: 189-194.

Miller, T.L., and Wolin, M.J. (1985) *Methanosphaera stadtmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. Archives of Microbiology 141: 116-122.

Mitsumori, M., and Sun, W. (2008) Control of rumen microbial fermentation for mitigating methane emissions from the rumen. Asian-Australasian Journal of Animal Science 21: 144-154.

Morita, H., Shiratori, C., Murakami, M., Takami, H., Toh, H., Kato, Y. et al. (2008) *Sharpea azabuensis* gen. nov., sp. nov., a Gram-positive, strictly anaerobic bacterium isolated from the faeces of thoroughbred horses. International Journal of Systematic and Evolutionary Microbiology 58: 2682-2686.

Morvan, B., Bonnemoy, F., Fonty, G., and Gouet, P. (1996) Quantitative determination of H_2-utilizing acetogenic and sulfate-reducing bacteria and methanogenic archaea from digestive tract of different mammals. Current Microbiology 32: 129-133.

Moss, A.R., Jouany, J.-P., and Newbold, J. (2000) Methane production by ruminants: its contribution to global warming. Annales de Zootechnie 49: 231-254.

Moumen, A., Azizi, G., Chekroun, K.B., and Baghour, M. (2016) The effects of livestock methane emission on the global warming: A review. International Journal of Global Warming 9: 229-253.

Murray, E.L., and Conway, T. (2005) Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of *Escherichia coli*. Journal of Bacteriology 187: 991-1000.

Neves, A.R., Pool, W.A., Kok, J., Kuipers, O.P., and Santos, H. (2005) Overview on sugar metabolism and its control in *Lactococcus lactis* – The input from in vivo NMR. FEMS Microbiology Reviews 29: 531-554.

Newbold, C.J., de la Fuente, G., Belanche, A., Ramos-Morales, E., and McEwan, N.R. (2015) The role of ciliate protozoa in the rumen. Frontiers in Microbiology 6: 1313.

Ng, F., Kittelmann, S., Patchett, M.L., Attwood, G.T., Janssen, P.H., Rakonjac, J., and Gagic, D. (2016) An adhesin from hydrogen-utilizing rumen methanogen *Methanobrevibacter ruminantium* M1 binds a broad range of hydrogen-producing microorganisms. Environmental Microbiology 18: 3010-3021.

Niderkorn, V., Baumont, R., Le Morvan, A., and Macheboeuf, D. (2011) Occurrence of associative effects between grasses and legumes in binary mixtures on in vitro rumen fermentation characteristics. Journal of Animal Science 89: 1138-1145.

NIWA, N.Z. (2016). Methane. URL https://www.niwa.co.nz/atmosphere/our-data/trace-gas-plots/methane. Accessed 17 July 2017.

Noel, S. (2013) Cultivation and community composition analysis of plant-adherent rumen bacteria. PhD thesis, Massey University, Palmerston North, New Zealand.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A. et al. (2013) Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. Journal of Computational Biology 20: 714-737.

Nyonyo, T., Shinkai, T., Tajima, A., and Mitsumori, M. (2013) Effect of media composition, including gelling agents, on isolation of previously uncultured rumen bacteria. Letters in Applied Microbiology 56: 63-70.

Orpin, C.G. (1972) The culture in vitro of the rumen bacterium Quin's oval. Journal of General Microbiology 73: 523-530.

Orpin, C.G. (1975) Studies on the rumen flagellate *Neocallimastix frontalis*. Journal of General Microbiology 91: 249-262.

Paillard, D., McKain, N., Chaudhary, L.C., Walker, N.D., Pizette, F., Koppova, I. et al. (2007) Relation between phylogenetic position, lipid metabolism and butyrate production by different *Butyrivibrio*-like bacteria from the rumen. Antonie Van Leeuwenhoek 91: 417-422.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25: 1043-1055.

Patra, A.K., and Yu, Z. (2012) Effects of essential oils on methane production and fermentation by, and abundance and diversity of, rumen microbial populations. Applied and Environmental Microbiology 78: 4271-4280.

Peters, J.W. (1998) X-ray crystal structure of the Fe-only hydrogenase (Cpl) from *Clostridium pasteurianum* to 1.8 angstrom resolution. Science 282: 1853-1858.

Pieulle, L., Guigliarelli, B., Asso, M., Dole, F., Bernadac, A., and Hatchikian, E.C. (1995) Isolation and characterization of the pyruvate-ferredoxin oxidoreductase from the sulfate-reducing bacterium *Desulfovibrio africanus*. Biochimica et Biophysica Acta 1250: 49-59.

Pinares-Patio, C.S., Ulyatt, M.J., Lassey, K.R., Barry, T.N., and Holmes, C.W. (2003) Persistence of differences between sheep in methane emission under generous grazing conditions. Journal of Agricultural Science 140: 227-233.

Pinares-Patiño, C., McEwan, J., Dodds, K., Cárdenas, E., Hegarty, R., Koolaard, J., and Clark, H. (2011a) Repeatability of methane emissions from sheep. Animal Feed Science and Technology 166: 210-218.

Pinares-Patiño, C.S., Ebrahimi, S.H., McEwan, J.C., Clark, H., and Luo, D. (2011b) Is rumen retention time implicated in sheep differences in methane emission? Proceedings of the New Zealand Society of Animal Production 71: 219-222.

Pinares-Patiño, C., Kjestrup, H., MacLean, S., Sandoval, E., Molano, G., Harland, R. et al. (2013a) Methane emission from sheep is related to concentrations of rumen volatile fatty acids. In: Energy and protein metabolism and nutrition in sustainable animal production, vol 134. Oltjen J.W., Kebreab E., Lapierre H. (eds). Wageningen, The Netherlands: Wageningen Academic Publishers, pp. 495-496.

Pinares-Patiño, C.S., Hickey, S.M., Young, E.A., Dodds, K.G., MacLean, S., Molano, G. et al. (2013b) Heritability estimates of methane emissions from sheep. Animal 7: 316-321.

Pitta, D., Barry, T., Lopez-Villalobos, N., and Attwood, G. (2009) Effect of willow supplementation upon plasma amino acid concentration in ewes grazing drought pastures of low nutritive value. Animal Feed Science and Technology 148: 183-191.

Pope, P.B., Smith, W., Denman, S.E., Tringe, S.G., Barry, K., Hugenholtz, P. et al. (2011) Isolation of *Succinivibrionaceae* implicated in low methane emissions from Tammar wallabies. Science 333: 646-648.

Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9: e93827.

Poulsen, M., Schwab, C., Jensen, B.B., Engberg, R.M., Spang, A., Canibe, N. et al. (2013) Methylotrophic methanogenic *Thermoplasmata* implicated in reduced methane emissions from bovine rumen. Nature Communications 4: 1428.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35: 7188-7196.

Purdom, M.R. (1963) Micromanipulation in the examination of rumen bacteria. Nature 198: 307-308.

Qi, M., Wang, P., O'Toole, N., Barboza, P.S., Ungerfeld, E., Leigh, M.B. et al. (2011) Snapshot of the eukaryotic gene expression in muskoxen rumen - A metatranscriptomic approach. PLoS One 6: e20521.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41: D590-D596.

Quin, J.I. (1943) Studies on the alimentary tract of merino sheep in South Africa VII.-Fermentation in the forestomach of sheep. Onderstepoort Journal of Veterinary Sciences and Animal Industry 18: 91-112.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M. et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods 6: 639-641.

Rajendhran, J., and Gunasekaran, P. (2011) Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. Microbiological Research 166: 99-110.

Ramette, A. (2007) Multivariate analyses in microbial ecology. FEMS Microbiology Ecology 62: 142-160.

Rea, S., Bowman, J.P., Popovski, S., Pimm, C., and Wright, A.-D.G. (2007) *Methanobrevibacter millerae* sp. nov. and *Methanobrevibacter olleyae* sp. nov.,

methanogens from the ovine and bovine rumen that can utilize formate for growth. International Journal of Systematic and Evolutionary Microbiology 57: 450-456.

Rees, E.M.R., Lloyd, D., and Williams, A.G. (1995) The effects of co-cultivation with the acetogen *Acetitomaculum ruminis* on the fermentative metabolism of the rumen fungi *Neocallimastix patriciarum* and *Neocallimastix* sp. strain L2. FEMS Microbiology Letters 133: 175-180.

Rivière, L., Van Weelden, S.W.H., Glass, P., Vegh, P., Coustou, V., Biran, M. et al. (2004) Acetyl:succinate CoA-transferase in procyclic *Trypanosoma brucei*. Gene identification and role in carbohydrate metabolism. Journal of Biological Chemistry 279: 45337-45346.

Roehe, R., Dewhurst, R.J., Duthie, C.-A., Rooke, J.A., McKain, N., Ross, D.W. et al. (2016) Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. PLoS Genetics 12: e1005846.

Rooke, J.A., Wallace, R.J., Duthie, C.-A., McKain, N., de Souza, S.M., Hyslop, J.J. et al. (2014) Hydrogen and methane emissions from beef cattle and their rumen microbial community vary with diet, time after feeding and genotype. British Journal of Nutrition 112: 1-10.

Ross, E.M., Moate, P.J., Marett, L., Cocks, B.G., and Hayes, B.J. (2013) Investigating the effect of two methane-mitigating diets on the rumen microbiome using massively parallel sequencing. Journal of Dairy Science 96: 6030-6046.

Ross, E.M., Moate, P.J., Bath, C.R., Davidson, S.E., Sawbridge, T.I., Guthridge, K.M. et al. (2012) High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. BMC Genetics 13: 53.

Rotaru, A.-E., Shrestha, P.M., Liu, F., Ueki, T., Nevin, K., Summers, Z.M., and Lovley, D.R. (2012) Interspecies electron transfer via hydrogen and formate rather than direct electrical connections in cocultures of *Pelobacter carbinolicus* and *Geobacter sulfurreducens*. Applied and Environmental Microbiology 78: 7645-7651.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: Sequence visualization and annotation. Bioinformatics 16: 944-945.

Salvetti, E., Felis, G.E., Dellaglio, F., Castioni, A., Torriani, S., and Lawson, P.A. (2011) Reclassification of *Lactobacillus catenaformis* (Eggerth 1935) Moore and Holdeman 1970 and *Lactobacillus vitulinus* Sharpe et al. 1973 as *Eggerthia catenaformis* gen. nov., comb. nov. and *Kandleria vitulina* gen. nov., comb. nov., respectively. International Journal of Systematic and Evolutionary Microbiology 61: 2520-2524.

Samols, D., Thornton, C.G., Murtif, V.L., Kumar, G.K., Haase, F.C., and Wood, H.G. (1988) Evolutionary conservation among biotin enzymes. Journal of Biological Chemistry 263: 6461-6464.

Santana, M., Ionescu, M.S., Vertes, A., Longin, R., Kunst, F., Danchin, A., and Glaser, P. (1994) *Bacillus subtilis* F0F1 ATPase: DNA sequence of the atp operon and characterization of atp mutants. Journal of Bacteriology 176: 6802-6811.

Saraste, M. (1999) Oxidative phosphorylation at the fin de siècle. Science 283: 1488-1493.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K. et al. (2011) Database resources of the national center for biotechnology information. Nucleic Acids Research 39: D38-D51.

Scheurwater, E., Reid, C.W., and Clarke, A.J. (2008) Lytic transglycosylases: Bacterial space-making autolysins. International Journal of Biochemistry and Cell Biology 40: 586-591.

Schramm, A., Fuchs, B.M., Nielsen, J.L., Tonolla, M., and Stahl, D.A. (2002) Fluorescence in situ hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. Environmental Microbiology 4: 713-720.

Schurig, H., Beaucamp, N., Ostendorp, R., Jaenicke, R., Adler, E., and Knowles, J.R. (1995) Phosphoglycerate kinase and triosephosphate isomerase from the hyperthermophilic bacterium *Thermotoga maritima* form a covalent bifunctional enzyme complex. EMBO Journal 14: 442-451.

Schwartz, E., Fritsch, J., and Friedrich, B. (2013) H_2-metabolizing prokaryotes. In The Prokaryotes: Prokaryotic Physiology and Biochemistry. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Germany: Springer pp. 119-199.

Schwarz, E., Oesterhelt, D., Reinke, H., Beyreuther, K., and Dimroth, P. (1988) The sodium ion translocating oxalacetate decarboxylase of *Klebsiella pneumoniae*. Sequence of the biotin-containing α-subunit and relationship to other biotin-containing enzymes. Journal of Biological Chemistry 263: 9640-9645.

Seedorf, H., Kittelmann, S., Henderson, G., and Janssen, P.H. (2014) RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. PeerJ 2: e494.

Senior, A.E. (1988) ATP synthesis by oxidative phosphorylation. Physiological Reviews 68: 177-231.

Shabat, S.K., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., and Berg Miller, M.E. (2016) Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. ISME Journal 10: 2958-2972.

Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., and Banfield, J.F. (2013) Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Research 23: 111-120.

Sharp, R., Ziemer, C.J., Stern, M.D., and Stahl, D.A. (1998) Taxon-specific associations between protozoal and methanogen populations in the rumen and a model rumen system. FEMS Microbiology Ecology 26: 71-78.

Sharpe, M.E., Latham, M.J., Garvie, E.I., Zirngibl, J., and Kandler, O. (1973) Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp.nov. and *Lactobacillus vitulinus* sp.nov. Microbiology 77: 37-49.

Shi, W., Moon, C.D., Leahy, S.C., Kang, D., Froula, J., Kittelmann, S. et al. (2014) Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Research 24: 1517-1525.

Shima, S., Pilak, O., Vogt, S., Schick, M., Stagni, M.S., Meyer-Klaucke, W. et al. (2008) The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. Science 321: 572-575.

Shomura, Y., Yoon, K.-S., Nishihara, H., and Higuchi, Y. (2011) Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase. Nature 479: 253-256.

Skulachev, V.P. (1978) Membrane-linked energy buffering as the biological function of Na^+/K^+ gradient. FEBS Letters 87: 171-179.

Smith, A.M., Heisler, L.E., St.Onge, R.P., Farias-Hesson, E., Wallace, I.M., Bodeau, J. et al. (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Research 38: e142-e142.

Sneath, P.H.A., and Sokal, R.R. (1962) Numerical taxonomy. Nature 193: 855-860.

Solden, L.M., Hoyt, D.W., Collins, W.B., Plank, J.E., Daly, R.A., Hildebrand, E. et al. (2017) New roles in hemicellulosic sugar fermentation for the uncultivated *Bacteroidetes* family BS11. ISME Journal 11: 691-703.

Solomon, F., and Jencks, W.P. (1969) Identification of an enzyme-gamma-glutamyl coenzyme A intermediate from coenzyme A transferase. Journal of Biological Chemistry 244: 1079-1081.

Stackebrandt, E., and Goebel, B.M. (1994) Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology 44: 846-849.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:149.

Statistics N.Z. (2016). Imports and exports: 2016. URL http://www.stats.govt.nz/browse_for_stats/industry_sectors/imports_and_exports.aspx. Accessed 27 February 2017.

Studer, R., Dahinden, P., Wang, W.-W., Auchli, Y., Li, X.-D., and Dimroth, P. (2007) Crystal structure of the carboxyltransferase domain of the oxaloacetate decarboxylase Na^+ pump from *Vibrio cholerae*. Journal of Molecular Biology 367: 547-557.

Svartstrom, O., Alneberg, J., Terrapon, N., Lombard, V., de Bruijn, I., Malmsten, J. et al. (2017) Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. ISME Journal 17: 1751-1762.

Tapio, I., Snelling, T.J., Strozzi, F., and Wallace, R.J. (2017) The ruminal microbiome associated with methane emissions from ruminant livestock. Journal of Animal Science and Biotechnology 8: 7.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S. et al. (2001) The COG database: New developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Research 29: 22-28.

Tchieu, J.H., Norris, V., Edwards, J.S., and Saier M.H, Jr. (2001) The complete phosphotransferase system in *Escherichia coli*. Journal of Molecular Microbiology and Biotechnology 3: 329-346.

Teusink, B., Bachmann, H., and Molenaar, D. (2011) Systems biology of lactic acid bacteria: a critical review. Microbial Cell Factories 10: S11-S11.

Thomas, T.D., Ellwood, D.C., and Longyear, V.M.C. (1979) Change from homo- to heterolactic fermentation by *Streptococcus lactis* resulting from glucose limitation in anaerobic chemostat cultures. Journal of Bacteriology 138: 109-117.

Tielens, A.G.M., van Grinsven, K.W.A., Henze, K., van Hellemond, J.J., and Martin, W. (2010) Acetate formation in the energy metabolism of parasitic helminths and protists. International Journal for Parasitology 40: 387-397.

Towbin, H., Staehelin, T., and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proceedings of the National Academy of Sciences USA 76: 4350-4354.

Tschech, A., and Pfennig, N. (1984) Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*. Archives of Microbiology 137: 163-167.

Tymensen, L.D., and McAllister, T.A. (2012) Community structure analysis of methanogens associated with rumen protozoa reveals bias in universal archaeal primers. Applied and Environmental Microbiology 78: 4051-4056.

Tymensen, L.D., Beauchemin, K.A., and McAllister, T.A. (2012) Structures of free-living and protozoa-associated methanogen communities in the bovine rumen differ according to comparative analysis of 16S rRNA and mcrA genes. Microbiology 158: 1808-1817.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M. et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37-43.

van der Westhuizen, G.C.A., Oxford, A.E., and Quin, J.I. (1950) Studies on the alimentary tract of merino sheep in South Africa. XVI. On the identity of *Schizosaccharomyces ovis*. Part I. Some yeast-like organisms isolated from the rumen contents of sheep fed on a lucerne diet. Onderstepoort Journal of Veterinary Science and Animal Industry 24: 119-124.

Van Grinsven, K.W.A., Rosnowsky, S., Van Weelden, S.W.H., Pütz, S., Van Der Giezen, M., Martin, W. et al. (2008) Acetate:succinate CoA-transferase in the hydrogenosomes of *Trichomonas vaginalis*: Identification and characterization. Journal of Biological Chemistry 283: 1411-1418.

van Lingen, H.J., Plugge, C.M., Fadel, J.G., Kebreab, E., Bannink, A., and Dijkstra, J. (2016) Thermodynamic driving force of hydrogen on rumen microbial metabolism: A theoretical investigation. PLoS One 11: e0161362.

Van Nevel, C., and Demeyer, D. (1997) Manipulation of rumen fermentation. The rumen microbial ecosystem. In: The Rumen Microbial Ecosystem, 2nd edition. Hobson, P.N., and Stewart, C.S. (eds). New York, NY, USA: Chapman & Hall. pp. 523-632.

Van Soest, P.J. (1994) Nutritional Ecology of the Ruminant. Ithaca, NY, USA: Cornell University Press.

Vaughan, T., Ryan, J., and Czaplewski, N. (2011) Mammalogy, 5th edition. Sudbury, MA, USA: Jones & Bartlett Publishers.

Veira, D.M. (1986) The role of ciliate protozoa in nutrition of the ruminant. Journal of Animal Science 63: 1547-1560.

Venable, J.H., and Coggeshall, R. (1965) A simplified lead citrate stain for use in electron microscopy. Journal of Cell Biology 25: 407-408.

Vicini, J.L., Brulla, W.J., Davis, C.L., and Bryant, M.P. (1987) Quin's oval and other microbiota in the rumens of molasses fed sheep. Applied and Environmental Microbiology 53: 1273-1276.

Vignais, P.M., and Billoud, B. (2007) Occurrence, classification, and biological function of hydrogenases: An overview. Chemical Reviews 107: 4206-4272.

Viklund, H., and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 24: 1662-1668.

Volbeda, A., Charon, M.-H., Piras, C., Hatchikian, E.C., Frey, M., and Fontecilla-Camps, J.C. (1995) Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. Nature 373: 580-587.

Volbeda, A., Amara, P., Darnault, C., Mouesca, J.-M., Parkin, A., Roessler, M.M. et al. (2012) X-ray crystallographic and computational studies of the O_2-tolerant [NiFe]-hydrogenase 1 from *Escherichia coli*. Proceedings of the National Academy of Sciences USA 109: 5305-5310.

Volbeda, A., Darnault, C., Parkin, A., Sargent, F., Armstrong, Fraser A., and Fontecilla-Camps, Juan C. (2013) Crystal structure of the O_2-tolerant membrane-bound hydrogenase 1 from *Escherichia coli* in complex with its cognate cytochrome b. Structure 21: 184-190.

Waghorn, G., and Dewhurst, R. (2007) Feed efficiency in cattle—The contribution of rumen function. In: Meeting the challenges for pasture-based dairying. Proceedings of the 3rd Dairy Science Symposium, University of Melbourne. Chapman D.F., Clark D.A., Macmillan K.L., and Nation D.P. (eds). Melbourne, Australia: National Dairy Alliance, pp. 111-123.

Wall, E., Simm, G., and Moran, D. (2010) Developing breeding schemes to assist mitigation of greenhouse gas emissions. Animal 4: 366-376.

Wallace, R.J. (1978) Control of lactate production by *Selenomonas ruminantium*: homotropic activation of lactate dehydrogenase by pyruvate. Journal of General Microbiology 107: 45-52.

Wallace, R.J., Rooke, J.A., McKain, N., Duthie, C.A., Hyslop, J.J., Ross, D.W. et al. (2015) The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics 16: 839.

Waygood, E.B., and Sanwal, B.D. (1974) The control of pyruvate kinases of *Escherichia coli*. I. Physicochemical and regulatory properties of the enzyme activated by fructose 1,6 diphosphate. Journal of Biological Chemistry 249: 265-274.

Wedlock, D.N., Pedersen, G., Denis, M., Buddle, B.M., Dey, D., and Janssen, P.H. (2010) Development of a vaccine to mitigate greenhouse gas emissions in agriculture: Vaccination of sheep with methanogen fractions induces antibodies that block methane production in vitro. New Zealand Veterinary Journal 58: 29-36.

Weimer, P.J., Stevenson, D.M., Mantovani, H.C., and Man, S.L.C. (2010) Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents. Journal of Dairy Science 93: 5902-5912.

Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D. et al. (1989) A phylogenetic analysis of the mycoplasmas: Basis for their classification. Journal of Bacteriology 171: 6455-6467.

Welkie, D.G., Stevenson, D.M., and Weimer, P.J. (2010) ARISA analysis of ruminal bacterial community dynamics in lactating dairy cows during the feeding cycle. Anaerobe 16: 94-100.

Wicken, A.J., and Howard, B.H. (1967) On the taxonomic status of "Quin's oval" organisms. Journal of General Microbiology 47: 207-211.

Widdel, F., Kohring, G.W., and Mayer, F. (1983) Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. III. Characterization of the filamentous gliding *Desulfonema limicola* gen. nov. sp. nov., and *Desulfonema magnum* sp. nov. Archives of Microbiology 134: 286-294.

Woehlke, G., Wifling, K., and Dimroth, P. (1992) Sequence of the sodium ion pump oxaloacetate decarboxylase from *Salmonella typhimurium*. Journal of Biological Chemistry 267: 22798-22803.

Wolin, M. (1976) Interactions between H_2 producing and methane producing species. In: Microbial Formation and Utilization of Gases. Schlegel, H., Gottschalk, G., and Pfennig, N. (eds). Gottingen, Germany: Verlag Goltze, pp 141-150.

Wolin, M.J. (1979) The rumen fermentation: A model for microbial interactions in anaerobic ecosystems. Advances in Microbial Ecology 3: 49-77.

Woodcock, H.M., and Lapage, G. (1913) Memoirs: On a remarkable new type of protistan parasite. Quarterly Journal of Microscopical Sciences 59: 431-457.

Wright, A., Kennedy, P., O'Neill, C., Toovey, A., Popovski, S., Rea, S. et al. (2004) Reducing methane emissions in sheep by immunization against rumen methanogens. Vaccine 22: 3976-3985.

Wu, M., and Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28: 1033-1034.

Wu, W.-M., Jain, M.K., Hickey, R.F., and Zeikus, J.G. (1996) Perturbation of syntrophic isobutyrate and butyrate degradation with formate and hydrogen. Biotechnology and Bioengineering 52: 404-411.

Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2: 26.

Yáñez-Ruiz, D.R., Abecia, L., and Newbold, C.J. (2015) Manipulating rumen microbiome and fermentation through interventions during early life: a review. Frontiers in Microbiology 6. 1133.

Yang, S., Ma, S., Chen, J., Mao, H., He, Y., Xi, D. et al. (2010) Bacterial diversity in the rumen of Gayals (Bos frontalis), Swamp buffaloes (*Bubalus bubalis*) and Holstein cow as revealed by cloned 16S rRNA gene sequences. Molecular Biology Reports 37: 2063-2073.

Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K.H. et al. (2008) The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. Systematic and Applied Microbiology 31: 241-250.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13: 134.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) DbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 40: W445-W451.

Yokoyama, M., and Johnson, K. (1988) Microbiology of the rumen and intestine. In: The Ruminant Animal: Digestive Physiology and Nutrition. Church, D.C. (ed) Englewood Cliffs, NJ, USA: Prentice-Hall, pp. 125-144.

Yutin, N., and Galperin, M.Y. (2013) A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. Environmental Microbiology 15: 2631-2641.

Zerbino, D.R., and Birney, E. (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18: 821-829.

Ziemer, C.J., Sharp, R., Stern, M.D., Cotta, M.A., Whitehead, T.R., and Stahl, D.A. (2000) Comparison of microbial populations in model and natural rumens using 16S ribosomal RNA-targeted probes. Environmental Microbiology 2: 632-643.

Appendices

Table A3.1. Details of the 88 taxa with a relative abundance of >1% in any of the samples, forming the reduced taxa dataset.

Taxa	Average relative abundance (%)			
	Avg	In community types		
		Q	S	H
Quinella	11.00	31.90	1.30	0.58
Prevotella 1	20.79	14.97	22.79	25.32
Fibrobacter	4.12	2.86	3.52	7.53
Sharpea	2.76	0.77	1.35	9.17
Rikenellaceae RC9 gut group	7.52	5.29	8.63	8.45
Kandleria	0.64	0.29	0.18	2.28
Prevotellaceae UCG 001	2.15	1.52	2.82	1.58
Xylanibacter	1.76	1.21	2.20	1.62
Oscillaspira	0.42	0.46	0.41	0.40
Christensellaceae R7 group	4.51	4.38	5.49	2.43
Erysipelotrichaceae UCG 002	0.06	0.00	0.00	0.29
Anaeroplasma	2.28	2.02	2.22	2.82
Acetitomaculum	2.11	1.57	2.33	2.47
Lachnospiraceae NK3A20 group	2.31	1.80	2.63	2.39
Ruminococcaceae NK4A214 group	2.61	1.99	3.34	1.92
Bacteroidales S24_7	1.90	1.34	2.16	2.21
Ruminococcaceae UCG 005	1.10	0.87	1.50	0.54
Bacteroidales BS11 gut group	2.63	2.44	3.00	2.10
Prevotella 7	0.10	0.03	0.05	0.31
Coriobacteriaceae UCG 001	0.31	0.07	0.09	1.18
Lachnospiraceae gauvreauii group	1.07	0.82	1.02	1.60
Ruminococcus 1	2.07	1.81	2.34	1.88
candidate division SR1	0.34	0.19	0.52	0.15
Erysipelotrichaceae UCG 004	0.69	0.67	0.77	0.53
Thalassospira	1.17	0.86	1.67	0.49
Dorea	0.66	0.27	0.72	1.10

Taxa	Average relative abundance (%)			
	Avg	In community types		
		Q	S	H
Lachnospiraceae NK4A136 group	0.90	0.82	1.04	0.70
Synergistaceae UCG 001	0.12	0.34	0.02	0.01
Prevotellaceae NK3B31 group	0.60	0.44	0.69	0.64
Lachnospiraceae AC2044 group	0.91	0.77	1.21	0.44
Opitutae vadinHA64	0.04	0.02	0.06	0.00
Cyanobacteria 4C0d 2	0.67	0.58	0.87	0.36
Ruminococcaceae UCG 002	0.46	0.43	0.62	0.12
Roseburia	0.31	0.25	0.36	0.27
Saccharofermentans	0.86	0.77	0.97	0.77
Lachnospiraceae colinum group	0.33	0.27	0.48	0.08
Ruminococcaceae UCG 010	0.83	0.77	1.00	0.53
Bacteroidales RF16	0.50	0.39	0.61	0.40
candidate division TM7	0.25	0.24	0.29	0.18
Blautia	0.68	0.69	0.76	0.51
Eubacterium ruminantium group	0.28	0.25	0.32	0.25
Oribacterium	0.65	0.50	0.77	0.62
Ruminococcus 2	0.27	0.23	0.34	0.20
Ruminococcaceae UCG 008	0.19	0.07	0.13	0.49
Lachnospiraceae ND3007 group	0.47	0.41	0.58	0.29
Syntrophococcus	0.46	0.35	0.39	0.76
Victivallis	0.10	0.07	0.15	0.02
Rhodospirillaceae UCG 001	0.10	0.10	0.15	0.02
Porphyromonadaceae UCG 001	0.13	0.18	0.13	0.05
Ruminococcaceae UCG 001	0.11	0.09	0.15	0.06
Siraeum group	0.29	0.29	0.39	0.06
Coprostanoligenes group	0.48	0.42	0.61	0.26
Succiniclasticum	0.66	0.62	0.78	0.46
Butyrivibrio 2	0.59	0.39	0.77	0.49
Clostridiales 1 Family XIII AD3011 group	0.41	0.39	0.47	0.28
Prevotellaceae UCG 003	0.45	0.35	0.56	0.36

Taxa	Average relative abundance (%)			
	Avg	In community types		
		Q	S	H
*Victivallaceae U*CG 001	0.19	0.16	0.25	0.11
Ruminococcaceae UCG 014	0.38	0.33	0.45	0.28
Prevotella 11	0.30	0.22	0.32	0.37
Clostridiales 1 UCG 001	0.34	0.34	0.41	0.21
Treponema 2	0.29	0.23	0.36	0.22
Lachnospiraceae XPB1014 group	0.29	0.25	0.37	0.14
Ruminococcaceae viride group	0.14	0.06	0.12	0.30
Atopobium	0.39	0.28	0.45	0.42
Lachnospiraceae phytofermentans group	0.12	0.09	0.17	0.06
Pseudobutyrivibrio	0.36	0.28	0.46	0.25
Lachnospiraceae uniforme group	0.10	0.06	0.11	0.15
Prevotellaceae UCG 004	0.28	0.26	0.30	0.25
Lachnospiraceae hallii group	0.42	0.36	0.47	0.41
Lachnospiraceae herbivorans group	0.15	0.10	0.19	0.13
Marvinbryantia	0.51	0.41	0.61	0.41
Endomicrobium	0.09	0.05	0.15	0.03
Prevotella YAB2003 group	0.07	0.03	0.05	0.20
Lachnospiraceae UCG 008	0.38	0.32	0.44	0.35
Aneraolineaceae UCG 001	0.21	0.22	0.24	0.12
Prevotella 2	0.04	0.02	0.06	0.04
Lachnospiraceae NK4B4 group	0.10	0.11	0.11	0.07
Coprococcus 1	0.17	0.16	0.17	0.16
Clostridiales 1 Family XIII nodatum group	0.14	0.12	0.15	0.15
Ruminococcaceae UCG 009	0.06	0.06	0.07	0.04
Lachnospiraceae oxidoreducens group	0.16	0.17	0.19	0.08
Lachnospiraceae UCG 011	0.14	0.10	0.15	0.15
Lachnospiraceae FCS020 group	0.16	0.13	0.19	0.15
Captivaceae	0.01	0.01	0.01	0.03
Moryella	0.21	0.18	0.26	0.18
Lachnospiraceae aminophilum group	0.19	0.15	0.22	0.19

Taxa	Average relative abundance (%)			
	Avg	In community types		
		Q	S	H
Erysipelotrichaceae UCG 001	0.09	0.08	0.08	0.13
Parabacteroides	0.02	0.01	0.02	0.06
Coriobacteriaceae UCG 004	0.09	0.09	0.10	0.08

Table A3.2 Summary of statistical analysis and average relative abundance (%) of taxa represent in Table 3.2. Top 20 taxa with higher relative abundance (%) in either of the dataset in any of the community type.

Taxa	Average relative abundance in all samples (reduced taxa) dataset			Average relative abundance extreme samples dataset			Spearman's ranks correlation with methane yield								Kruskal Wallis test using different datasets	
							Using Q and H-community type samples				Using S and H-community type samples				In all samples	In extreme samples
	Q-type	H-type	S-type	Q-type	H-type	S-type	In all samples		In extreme samples		In all samples		In extreme samples			
							R-value	p-value	R-value	p-value	R-value	p-value	R-value	p-value	p-value	p-value
Kandleria	0.29	0.18	2.28	0.40	0.15	2.64	−0.11	1.55E−01	−0.15	4.12E−02	−0.36	4.79E−06	−0.44	1.49E−08	3.21E−15	2.08E−08
Quinella	31.90	1.30	0.58	30.45	1.49	0.45	−0.40	1.33E−08	−0.43	9.14E−10	0.08	3.07E−01	0.13	1.02E−01	0	2.08E−08
Fibrobacter	2.86	3.52	7.53	3.06	3.66	7.88	−0.01	9.36E−01	−0.03	6.42E−01	−0.31	1.19E−04	−0.39	5.54E−07	4.45E−11	2.08E−08
Coriobacteriaceae UCG 001	0.07	0.09	1.18	0.08	0.10	1.15	−0.06	4.15E−01	−0.07	3.16E−01	−0.30	1.99E−04	−0.39	5.70E−07	3.33E−16	2.08E−08
Sharpea	0.77	1.35	9.17	0.76	1.65	9.58	0.13	8.62E−02	0.16	2.74E−02	−0.15	5.77E−02	−0.22	5.08E−03	0	2.08E−08
Anaeroplasma	2.02	2.22	2.82	2.17	2.19	2.89	−0.04	6.00E−01	−0.06	4.06E−01	−0.12	1.39E−01	−0.12	1.43E−01	0.052	2.08E−08
Lachnospiraceae NK4A136 group	0.82	1.04	0.70	0.88	0.89	0.71	−0.10	1.61E−01	−0.09	2.18E−01	−0.04	6.31E−01	0.04	6.08E−01	0.0022	2.08E−08
Ruminococcus 1	1.81	2.34	1.88	1.88	2.13	1.89	−0.06	4.36E−01	−0.03	6.92E−01	−0.09	2.51E−01	0.00	9.88E−01	0.00078	2.08E−08
Saccharofermentans	0.77	0.97	0.77	0.82	0.91	0.75	−0.02	7.48E−01	−0.05	4.96E−01	−0.03	6.73E−01	0.03	6.79E−01	0.0096	2.08E−08
Syntrophococcus	0.35	0.39	0.76	0.06	0.14	0.02	0.02	7.69E−01	−0.02	7.41E−01	−0.01	8.83E−01	0.01	8.94E−01	2.08E−08	2.08E−08
Prevotellaceae UCG 001	1.52	2.82	1.58	1.46	2.66	1.54	0.07	3.51E−01	0.04	5.82E−01	0.08	3.26E−01	0.10	2.12E−01	6.03E−08	2.08E−08
Dorea	0.27	0.72	1.10	0.29	0.77	0.95	0.13	7.26E−02	0.09	2.41E−01	0.01	9.02E−01	−0.03	6.97E−01	2.08E−08	2.08E−08
Bacteroidales S24_7	1.34	2.16	2.21	1.42	2.32	2.16	0.14	6.56E−02	0.12	1.12E−01	−0.11	1.81E−01	−0.13	9.60E−02	2.08E−08	2.08E−08
Lachnospiraceae NK3A20 group	1.80	2.63	2.39	1.87	2.69	2.27	0.15	4.28E−02	0.14	6.53E−02	0.06	4.85E−01	0.04	6.34E−01	2.08E−08	2.08E−08

Taxa	Average relative abundance in all samples (reduced taxa) dataset			Average relative abundance extreme samples dataset			Spearman's ranks correlation with methane yield								Kruskal Wallis test using different datasets	
							Using Q and H-community type samples				Using S and H-community type samples				In all samples	In extreme samples
	Q-type	H-type	S-type	Q-type	H-type	S-type	In all samples		In extreme samples		In all samples		In extreme samples			
							R-value	p-value	R-value	p-value	R-value	p-value	R-value	p-value	p-value	p-value
Rikenellaceae RC9 gut group	5.29	8.63	8.45	5.42	8.69	8.40	0.16	3.17E−02	0.13	8.93E−02	0.00	9.95E−01	−0.02	7.81E−01	2.08E−08	2.08E−08
Bacteroidales BS11 gut group	2.44	3.00	2.10	2.50	2.77	2.14	0.08	3.00E−01	0.01	8.92E−01	0.13	1.02E−01	0.16	4.52E−02	0.00043	2.08E−08
Acetitomaculum	1.57	2.33	2.47	1.72	2.32	2.37	0.17	2.05E−02	0.17	2.26E−02	0.08	3.09E−01	0.10	2.06E−01	0.0016	2.08E−08
Lachnospiraceae gauvreauii group	0.82	1.02	1.60	0.86	1.02	1.64	0.13	7.80E−02	0.20	7.29E−03	−0.14	9.06E−02	−0.15	5.77E−02	0.00015	2.08E−08
Thalassospira	0.86	1.67	0.49	0.79	1.54	0.51	−0.01	8.45E−01	−0.04	5.57E−01	0.13	9.76E−02	0.21	1.02E−02	4.66E−12	2.08E−08
Xylanibacter	1.21	2.20	1.62	1.25	2.53	1.56	0.25	7.08E−04	0.24	1.18E−03	0.20	1.08E−02	0.20	1.42E−02	0.0053	2.08E−08
Prevotella 1	14.97	22.79	25.32	14.78	23.84	25.18	0.25	5.16E−04	0.22	2.39E−03	0.02	7.96E−01	−0.03	7.22E−01	7.1E−11	2.08E−08
Lachnospiraceae AC2044 group	0.77	1.21	0.44	0.79	1.14	0.44	−0.08	2.59E−01	−0.05	4.68E−01	0.17	3.95E−02	0.26	9.51E−04	2.25E−07	2.08E−08
Christensellaceae R7 group	4.38	5.49	2.43	4.40	5.05	2.39	0.10	1.71E−01	0.14	5.96E−02	0.23	4.36E−03	0.39	7.25E−07	7.25E−12	2.08E−08
Cyanobacteria 4C0d 2	0.58	0.87	0.36	0.61	0.95	0.38	0.31	2.36E−05	0.34	2.13E−06	0.40	2.04E−07	0.46	1.46E−09	7.59E−08	2.08E−08
Ruminococcaceae NK4A214 group	1.99	3.34	1.92	2.06	3.56	1.93	0.48	6.18E−12	0.49	2.95E−12	0.42	5.43E−08	0.44	1.44E−08	8.48E−09	2.08E−08
Ruminococcaceae UCG 005	0.87	1.50	0.54	0.86	1.40	0.54	0.28	1.39E−04	0.33	4.14E−06	0.41	1.27E−07	0.53	1.13E−12	2.075E−08	2.08E−08

Table A5.1 Functional classification of the predicted genes in the four *Quinella* genome bins based on the clusters of orthologous proteins (COGs) database (Tatusov et al., 2001a).

COG categories	Description	SR1Q5		SR1Q7		SR2Q5		SR3Q1	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
C	Energy production and conversion	91	4.4	110	4.5	90	5.2	115	4.6
D	Cell cycle control, cell division, chromosome partitioning	51	2.5	51	2.1	41	2.4	52	2.1
E	Amino acid transport and metabolism	157	7.6	182	7.4	147	8.5	184	7.3
F	Nucleotide transport and metabolism	50	2.4	67	2.7	52	3.0	56	2.2
G	Carbohydrate transport and metabolism	110	5.3	158	6.5	120	6.9	159	6.3
H	Coenzyme transport and metabolism	111	5.4	116	4.7	100	5.8	111	4.4
I	Lipid transport and metabolism	54	2.6	71	2.9	55	3.2	80	3.2
J	Translation, ribosomal structure and biogenesis	171	8.3	198	8.1	172	9.9	226	9.0
K	Transcription	83	4.0	91	3.7	66	3.8	109	4.3
L	Replication, recombination and repair	83	4.0	92	3.8	60	3.5	89	3.5
M	Cell wall/membrane/envelope biogenesis	172	8.3	234	9.6	163	9.4	221	8.8
N	Cell motility	54	2.6	74	3.0	68	3.9	72	2.9
O	Posttranslational modification, protein turnover, chaperones	100	4.8	103	4.2	77	4.5	96	3.8
P	Inorganic ion transport and metabolism	58	2.8	63	2.6	45	2.6	78	3.1
Q	Secondary metabolites biosynthesis, transport and catabolism	51	2.5	73	3.0	32	1.9	77	3.1
R	General function prediction only	196	9.5	221	9.0	142	8.2	230	9.1
S	Function unknown	99	4.8	122	5.0	84	4.9	120	4.8
T	Signal transduction mechanisms	119	5.8	135	5.5	103	6.0	140	5.6
U	Intracellular trafficking, secretion, and vesicular transport	44	2.1	39	1.6	41	2.4	51	2.0
V	Defence mechanisms	68	3.3	88	3.6	42	2.4	85	3.4
W	Extracellular structures	20	1.0	17	0.7	16	0.9	28	1.1
X	Phage-derived proteins	33	1.6	29	1.2	4	0.2	28	1.1
Z	Cytoskeleton	1	0.05	0	0	0	0	1	0.04

COG categories	Description	SR1Q5		SR1Q7		SR2Q5		SR3Q1	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
	No COG category assigned	298	14.4	344	14.1	185	10.7	369	14.6
	Total number of genes[a]	2067		2445		1728		2521	

[a]Total gene numbers do not add up, as some genes were assigned to more than one COG category and so counted more than once.

Table A6.1 CAZymes found in *Quinella* genome bins. CAZymes with E-values < 1e-18 and coverage > 0.35 (Yin et al., 2012) were considered as good matches and are shown in this table.

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR2Q5_1268	AA4	Glycolate oxidase, subunit GlcD
SR1Q5_1058	CE10	Carboxylesterase
SR1Q7_161	CE10	Esterase, PHB depolymerase family
SR3Q1_1207	CE10	Esterase, PHB depolymerase family
SR1Q7_2225	CE10	Hypothetical protein
SR2Q5_1666	CE10	Hypothetical protein
SR2Q5_568	CE10	Hypothetical protein
SR2Q5_577	CE10	Hypothetical protein
SR3Q1_1020	CE10	Hypothetical protein
SR3Q1_1255	CE10	Hypothetical protein
SR3Q1_1456	CE10	Hypothetical protein
SR3Q1_1021	CE10	Polyketide synthase-associated domain
SR3Q1_2021	CE10	Putative carboxylesterase
SR1Q5_1059	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q5_1060	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q5_1061	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q5_1062	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q5_1859	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q5_1860	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q7_1258	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR3Q1_1018	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR3Q1_1019	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR3Q1_2386	CE10	Tat (twin-arginine translocation) pathway signal sequence
SR1Q5_1783	CE11	UDP-3-*O*-[3-hydroxymyristoyl] *N*-acetylglucosamine deacetylase, LpxC
SR1Q7_242	CE11	UDP-3-*O*-[3-hydroxymyristoyl] *N*-acetylglucosamine deacetylase, LpxC
SR2Q5_109	CE11	UDP-3-*O*-[3-hydroxymyristoyl] *N*-acetylglucosamine deacetylase, LpxC
SR3Q1_1227	CE11	UDP-3-*O*-[3-hydroxymyristoyl] *N*-acetylglucosamine deacetylase, LpxC
SR1Q5_1463	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR1Q5_1518	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR1Q5_850	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR1Q7_1593	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR1Q7_2096	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR1Q7_503	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR2Q5_126	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR2Q5_1278	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR2Q5_935	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR3Q1_1283	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR3Q1_1994	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR3Q1_943	CE4	Poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase pgab, pgab
SR1Q5_1093	CE4	Polysaccharide deacetylase family sporulation protein pdab, pdab
SR2Q5_1068	CE4	Polysaccharide deacetylase family sporulation protein pdab, pdab
SR3Q1_923	CE4	Polysaccharide deacetylase family sporulation protein pdab, pdab
SR3Q1_371	CE4	Putative urate catabolism protein
SR1Q5_2001	CE9	*N*-acetylglucosamine-6-phosphate deacetylase, naga
SR1Q7_202	CE9	*N*-acetylglucosamine-6-phosphate deacetylase, naga
SR2Q5_369	CE9	*N*-acetylglucosamine-6-phosphate deacetylase, naga
SR3Q1_743	CE9	*N*-acetylglucosamine-6-phosphate deacetylase, naga
SR1Q5_297	GH1	β-galactosidase
SR1Q7_1166	GH1	β-galactosidase
SR1Q7_71	GH1	β-galactosidase
SR2Q5_701	GH1	β-galactosidase
SR3Q1_11	GH1	β-galactosidase
SR2Q5_1651	GH13	1,4-α-glucan branching enzyme, glgb
SR3Q1_1695	GH13	1,4-α-glucan branching enzyme, glgb
SR3Q1_2497	GH13	1,4-α-glucan branching enzyme, glgb
SR3Q1_445	GH13	1,4-α-glucan branching enzyme, glgb
SR1Q5_1565	GH13	1,4-α-glucan branching enzyme, glgb
SR1Q7_1628	GH13	1,4-α-glucan branching enzyme, glgb
SR1Q7_209	GH13	1,4-α-glucan branching enzyme, glgb
SR2Q5_1470	GH13	1,4-α-glucan branching enzyme, glgb
SR2Q5_1500	GH13	α-phosphotrehalase, trec
SR3Q1_1971	GH13	α-phosphotrehalase, trec
SR1Q7_1085	GH13	α-phosphotrehalase, trec
SR1Q5_127	GH13	α-phosphotrehalase, trec
SR2Q5_1718	GH23	Lytic murein transglycosylase B, mltb
SR1Q5_1599	GH23	Lytic transglycosylase
SR1Q7_2199	GH23	Lytic transglycosylase

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR2Q5_1168	GH23	Lytic transglycosylase
SR3Q1_342	GH23	Lytic transglycosylase
SR1Q5_965	GH23	Transglycosylase
SR1Q7_1734	GH23	Transglycosylase
SR2Q5_680	GH23	Transglycosylase
SR3Q1_1032	GH23	Transglycosylase
SR1Q7_77	GH23	Transglycosylase SLT domain protein
SR1Q5_1231	GH25	Hypothetical protein
SR3Q1_2221	GH25	Hypothetical protein
SR3Q1_925	GH25	Hypothetical protein
SR3Q1_1046	GH25	Phosphoglycolate phosphatase, bacterial, gph
SR1Q5_759	GH3	β-glucosidase-related glycosidases
SR1Q7_451	GH3	β -glucosidase-related glycosidases
SR2Q5_479	GH3	β -glucosidase-related glycosidases
SR3Q1_185	GH3	β -glucosidase-related glycosidases
SR1Q5_1015	GH32	Sucrose-6-phosphate hydrolase
SR1Q5_52	GH32	Sucrose-6-phosphate hydrolase
SR1Q5_723	GH32	Sucrose-6-phosphate hydrolase
SR1Q5_871	GH32	Sucrose-6-phosphate hydrolase
SR1Q7_1180	GH32	Sucrose-6-phosphate hydrolase
SR1Q7_1646	GH32	Sucrose-6-phosphate hydrolase
SR1Q7_186	GH32	Sucrose-6-phosphate hydrolase
SR1Q7_796	GH32	Sucrose-6-phosphate hydrolase
SR2Q5_1090	GH32	Sucrose-6-phosphate hydrolase
SR2Q5_590	GH32	Sucrose-6-phosphate hydrolase
SR3Q1_280	GH32	Sucrose-6-phosphate hydrolase
SR3Q1_485	GH32	Sucrose-6-phosphate hydrolase
SR3Q1_733	GH32	Sucrose-6-phosphate hydrolase
SR3Q1_1579	GH4	6-phospho-α-glucosidase
SR1Q5_1094	GH77	4-α-glucanotransferase, malq
SR1Q7_2201	GH77	4-α-glucanotransferase, malq
SR2Q5_1065	GH77	4-α-glucanotransferase, malq
SR3Q1_1217	GH77	4-α-glucanotransferase, malq
SR3Q1_2458	GH77	4-α-glucanotransferase, malq
SR1Q7_1014	GH84	Hypothetical protein

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR1Q5_1310	GH84	Putative *O*- GlcNAcase
SR2Q5_292	GH84	Putative *O*-GlcNAcase
SR3Q1_2032	GH84	Putative *O*-GlcNAcase
SR1Q7_60	GT11	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_61	GT11	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_5	GT17	β-1,4-*N*-acetylgalactosaminyltransferase
SR1Q5_827	GT17	Hypothetical protein
SR1Q7_1137	GT17	Hypothetical protein
SR1Q7_1141	GT17	Hypothetical protein
SR1Q7_1427	GT17	Hypothetical protein
SR2Q5_1396	GT17	Hypothetical protein
SR1Q5_1742	GT19	Lipid-A-disaccharide synthase, lpxb
SR1Q7_692	GT19	Lipid-A-disaccharide synthase, lpxb
SR2Q5_978	GT19	Lipid-A-disaccharide synthase, lpxb
SR3Q1_759	GT19	Lipid-A-disaccharide synthase, lpxb
SR2Q5_1672	GT2	Colanic acid biosynthesis glycosyltransferase wcaa, wcaa
SR3Q1_324	GT2	Colanic acid biosynthesis glycosyltransferase wcaa, wcaa
SR1Q7_1724	GT2	Glycosyltransferase domain
SR3Q1_291	GT2	Glycosyltransferase domain
SR1Q7_1906	GT2	Mycofactocin system glycosyltransferase, mftf
SR1Q7_2396	GT2	Mycofactocin system glycosyltransferase, mftf
SR2Q5_337	GT2	Mycofactocin system glycosyltransferase, mftf
SR2Q5_811	GT2	Mycofactocin system glycosyltransferase, mftf
SR2Q5_813	GT2	Mycofactocin system glycosyltransferase, mftf
SR3Q1_1204	GT2	Mycofactocin system glycosyltransferase, mftf
SR1Q5_1122	GT2	*N*-acetyl-α-D-glucosaminyl L-malate synthase bsha, bsha
SR3Q1_549	GT2	*N*-acetyl-α-D-glucosaminyl L-malate synthase bsha, bsha
SR1Q5_101	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_1101	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_1602	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_1745	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_209	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_265	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_266	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_578	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR1Q5_579	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_656	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_986	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1197	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1410	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1693	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1782	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1783	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1784	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1786	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1974	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_2395	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_690	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_819	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_820	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_978	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_1075	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_1116	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_1118	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_1120	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_222	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_418	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_645	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_646	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_809	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_812	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR2Q5_841	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_1652	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_1996	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_216	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_225	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_755	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_756	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_84	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_86	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR3Q1_89	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR3Q1_917	GT2	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q7_1785	GT2	Putative glycosyltransferase, exosortase G- associated
SR3Q1_1225	GT2	Putative glycosyltransferase, exosortase G- associated
SR3Q1_215	GT2	Putative glycosyltransferase, exosortase G- associated
SR1Q7_768	GT2	Sugar transferase, PEP-CTERM/epsh1 system associated
SR2Q5_6	GT2	Sugar transferase, PEP-CTERM/epsh1 system associated
SR2Q5_818	GT2	Transferase 2, rsam/selenodomain-associated
SR1Q5_1255	GT26	Glycosyltransferase, wecb/taga/cpsf family
SR1Q7_1963	GT26	Glycosyltransferase, wecb/taga/cpsf family
SR2Q5_1414	GT26	Glycosyltransferase, wecb/taga/cpsf family
SR3Q1_1268	GT26	Glycosyltransferase, wecb/taga/cpsf family
SR1Q5_1278	GT28	Murg transferase
SR1Q5_718	GT28	Murg transferase
SR1Q7_1960	GT28	Murg transferase
SR1Q7_846	GT28	Murg transferase
SR2Q5_1515	GT28	Murg transferase
SR3Q1_1265	GT28	Murg transferase
SR1Q7_1977	GT30	Tetraacyldisaccharide 4'-kinase
SR1Q7_331	GT30	Tetraacyldisaccharide 4'-kinase
SR1Q5_69	GT30	Tetraacyldisaccharide 4'-kinase, lpxk
SR1Q7_22	GT30	Tetraacyldisaccharide 4'-kinase, lpxk
SR2Q5_1349	GT30	Tetraacyldisaccharide 4'-kinase, lpxk
SR3Q1_505	GT30	Tetraacyldisaccharide 4'-kinase, lpxk
SR3Q1_1203	GT32	Hypothetical protein
SR1Q7_210	GT35	Glycogen/starch/α-glucan phosphorylases, glgp
SR2Q5_1471	GT35	Glycogen/starch/α-glucan phosphorylases, glgp
SR3Q1_1696	GT35	Glycogen/starch/α-glucan phosphorylases, glgp
SR1Q7_1061	GT4	Accessory Sec system glycosylation protein gtfa, gtfa
SR1Q5_1054	GT4	*N*-acetyl-α-D-glucosaminyl L-malate synthase bsha, bsha
SR1Q5_1216	GT4	*N*-acetyl-α-D-glucosaminyl L-malate synthase bsha, bsha
SR1Q7_1350	GT4	*N*-acetyl-α-D-glucosaminyl L-malate synthase bsha, bsha
SR1Q5_68	GT4	Sugar transferase, PEP-CTERM/epsh1 system associated
SR1Q7_604	GT4	Sugar transferase, PEP-CTERM/epsh1 system associated
SR2Q5_820	GT4	Sugar transferase, PEP-CTERM/epsh1 system associated

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR3Q1_1169	GT4	Sugar transferase, PEP-CTERM/epsh1 system associated
SR1Q5_82	GT41	Hypothetical protein
SR1Q5_84	GT41	Hypothetical protein
SR1Q7_1076	GT41	Hypothetical protein
SR1Q7_258	GT41	Hypothetical protein
SR2Q5_1369	GT41	Hypothetical protein
SR3Q1_1043	GT41	Hypothetical protein
SR3Q1_1182	GT41	Hypothetical protein
SR3Q1_1183	GT41	Hypothetical protein
SR1Q5_1234	GT41	Hypothetical protein HMPREF9166
SR1Q5_1018	GT41	Peptide *S*-glycosyltransferase, suns family, suns
SR1Q7_962	GT41	Peptide *S*-glycosyltransferase, suns family, suns
SR2Q5_907	GT41	Peptide *S*-glycosyltransferase, suns family, suns
SR3Q1_487	GT41	Peptide *S*-glycosyltransferase, suns family, suns
SR1Q7_792	GT41	Poly-β-1,6 *N*-acetyl-D-glucosamine export porin pgaa, pgaa
SR1Q7_2089	GT41	Putative PEP-CTERM system TPR-repeat lipoprotein, prst
SR2Q5_1003	GT41	Putative PEP-CTERM system TPR-repeat lipoprotein, prst
SR2Q5_306	GT41	Putative PEP-CTERM system TPR-repeat lipoprotein, prst
SR3Q1_1574	GT41	Surface carbohydrate biosynthesis protein
SR3Q1_1186	GT41	Type VI secretion lipoprotein, VC
SR1Q7_208	GT5	Glycogen/starch synthase, ADP-glucose type, glga
SR2Q5_1469	GT5	Glycogen/starch synthase, ADP-glucose type, glga
SR3Q1_1694	GT5	Glycogen/starch synthase, ADP-glucose type, glga
SR1Q5_1079	GT51	Penicillin-binding protein, 1A family
SR1Q5_1167	GT51	Penicillin-binding protein, 1A family
SR1Q7_1592	GT51	Penicillin-binding protein, 1A family
SR1Q7_2234	GT51	Penicillin-binding protein, 1A family
SR1Q7_890	GT51	Penicillin-binding protein, 1A family
SR2Q5_1186	GT51	Penicillin-binding protein, 1A family
SR2Q5_1279	GT51	Penicillin-binding protein, 1A family
SR3Q1_1596	GT51	Penicillin-binding protein, 1A family
SR3Q1_334	GT51	Penicillin-binding protein, 1A family
SR3Q1_364	GT8	CXXX repeat peptide maturase
SR1Q7_213	GT8	Deoxycytidine triphosphate deaminase, dcd
SR1Q7_2218	GT8	Excinuclease ABC subunit B, uvrb

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR1Q7_1952	GT8	Glycosyl transferase family 8
SR1Q7_2219	GT8	Glycosyl transferase family 8
SR1Q7_2220	GT8	Glycosyl transferase family 8
SR3Q1_354	GT8	Glycosyl transferase family 8
SR3Q1_357	GT8	Glycosyl transferase family 8
SR3Q1_365	GT8	Glycosyl transferase family 8
SR1Q7_2216	GT8	Hypothetical protein
SR1Q7_2217	GT8	Hypothetical protein
SR2Q5_1184	GT8	Hypothetical protein
SR3Q1_353	GT8	Hypothetical protein
SR3Q1_355	GT8	Hypothetical protein
SR3Q1_356	GT8	Hypothetical protein
SR3Q1_87	GT8	Hypothetical protein
SR1Q5_73	GT8	Poly-β-1,6 *N*-acetyl-D-glucosamine synthase, pgac
SR1Q5_900	GT81	Glycosyltransferase, TIGR04182 family
SR3Q1_129	GT81	Glycosyltransferase, TIGR04182 family
SR1Q7_1272	GT83	Dolichyl-phosphate-mannose-protein mannosyltransferase
SR2Q5_30	GT83	Glycosyl transferase
SR3Q1_2287	GT83	Glycosyl transferase
SR3Q1_2105	GT83	Oligosaccharyl transferase, archaeosortase A system- associated
SR1Q7_1268	GT9	Lipopolysaccharide heptosyltransferase I, waac
SR1Q7_1563	GT9	Lipopolysaccharide heptosyltransferase I, waac
SR2Q5_34	GT9	Lipopolysaccharide heptosyltransferase I, waac
SR3Q1_2109	GT9	Lipopolysaccharide heptosyltransferase I, waac
SR3Q1_2283	GT9	Lipopolysaccharide heptosyltransferase I, waac
SR3Q1_695	GT9	Lipopolysaccharide heptosyltransferase I, waac
SR1Q5_1659	GT9	Lipopolysaccharide heptosyltransferase II, waaf
SR1Q7_598	GT9	Lipopolysaccharide heptosyltransferase II, waaf
SR2Q5_819	GT9	Lipopolysaccharide heptosyltransferase II, waaf
SR3Q1_1170	GT9	Lipopolysaccharide heptosyltransferase II, waaf
SR1Q5_181	GT9	Putative lipopolysaccharide heptosyltransferase III, rfaq
SR1Q7_2075	GT9	Putative lipopolysaccharide heptosyltransferase III, rfaq
SR2Q5_1350	GT9	Putative lipopolysaccharide heptosyltransferase III, rfaq
SR3Q1_1870	GT9	Putative lipopolysaccharide heptosyltransferase III, rfaq
SR2Q5_321	GT92	Glycosyltransferase, TIGR04182 family

Quinella genome bin genes	CAZyme families match	GAMOLA2 annotation match
SR1Q5_1130	GT92	Hypothetical protein
SR1Q7_2203	GT92	Hypothetical protein
SR1Q7_2204	GT92	Hypothetical protein
SR1Q7_268	GT92	Hypothetical protein
SR2Q5_1067	GT92	Hypothetical protein
SR2Q5_1467	GT92	Hypothetical protein
SR3Q1_1218	GT92	Hypothetical protein
SR3Q1_1219	GT92	Hypothetical protein
SR3Q1_1220	GT92	Hypothetical protein
SR3Q1_1768	GT92	Hypothetical protein
SR3Q1_2335	GT92	Hypothetical protein

Table A6.2 Key enzymes in sugar fermentation and associated energetics found in *Quinella* genome bins.

Name of enzymes	EC numbers	*Quinella* genome bins and genes numbers			
		SR1Q5	SR1Q7	SR2Q5	SR3Q1
β-Glucosidase		759	451	479	185
α-Phosphotrehalase	3.2.1.1	127	1085	1500	1971
1,4-α-glucan branching enzyme		1565	209, 1682	1470, 1651	445, 1695, 2497
Phosphoenolpyruvate-dependent phosphotransferase (ptsI)	2.7.3.9	1813	987	1461	2430
Phosphocarrier protein, Hpr	2.7.11.	332, 1814, 1815	988, 1602	883, 997, 1462	1951, 2429
Phosphoglucomutase	5.4.2.2	59	320	1353, 1357	792
Glucose-1-phosphatase	3.1.3.10	1118	1004	326	553
Glucokinase	2.7.1.1	1337	1991	1371	1543
Glucose-6-phosphate isomerase	5.3.1.9	58	321	1352, 1356	793, 2391, 2467
6-Phosphofructokinase	2.7.1.11	1017, 1087	529, 1648	1089, 1662	66, 486
Fructose-1,6-bisphosphatase I	3.1.3.11	238	1559	-[a]	-
Fructose-bisphosphate aldolase	4.1.2.13	1533	2286	16	1198, 2174
Triosephosphate isomerase[b]	5.3.1.1	314	1842, 1909	284	272

Name of enzymes	EC numbers	*Quinella* genome bins and genes numbers			
		SR1Q5	SR1Q7	SR2Q5	SR3Q1
Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	313	1908	283	273
Phosphoglycerate kinase[b]	2.7.2.3	314	1842, 1909	284	272
Phosphoglycerate mutase	5.4.2.12	316	1845, 1911	287, 748	270
Phosphoglycerate enolase	4.2.1.11	897	1816	309	132
Phosphoenolpyruvate carboxykinase (ATP)	4.1.1.49	1272	705	1220	1538
Pyruvate kinase	2.7.1.40	1341	868	-	423
L-Lactate dehydrogenase	1.1.1.27	964	1633	1519	1033, 2465
Ribose-5-phosphate isomerase	5.3.1.6	135, 379, 1511	530, 1024	461	200, 1960, 1962
Ribulose-5-phosphate 3-epimerase	5.1.3.1	437, 1718	96, 615	861	-
Transketolase	2.2.1.1	2058, 2062	460, 757, 758, 881, 882, 1667,1668	469, 470, 471, 472, 481, 1571, 1572	189, 190, 1448, 1449
Transaldolase	2.2.1.2	-	-	-	2173
Pyruvate-ferredoxin/flavodoxin oxidoreductase	1.2.7.1	1976	152	868	2206
Flavodoxin		18, 1794	124, 1500	616, 1441	157, 1712
2-Oxoglutarate ferredoxin oxidoreductase, α subunit	1.2.7.8 or 1.2.7.3	1125	772	2	542

Name of enzymes	EC numbers	*Quinella* genome bins and genes numbers			
		SR1Q5	SR1Q7	SR2Q5	SR3Q1
2-Oxoglutarate ferredoxin oxidoreductase, β subunit	1.2.7.8	1126	773	1	541
2-Oxoglutarate ferredoxin oxidoreductase, γ subunit	1.2.7.8	1127	774	-	540
2-Oxoglutarate ferredoxin oxidoreductase, δ subunit/ 4Fe-4S ferredoxin		1124	771	3	543
NAD(P)H:flavin oxidoreductase/NADPH-flavin oxidoreductase		939, 1068	10, 1168, 1359	388, 774	307, 315, 769, 1313, 2281, 2513
Succinate CoA-transferase	2.8.3.-	482	972	1058	778
Formate dehydrogenase major subunit, fdoG	1.2.1.2	-	-	-	1644
Formate dehydrogenase iron-sulfur subunit, fdoH		-	-	-	1646
Formate dehydrogenase gamma subunit, fdoI		-	-	-	1647
Lactaldehyde reductase		-	413	-	-
Malate dehydrogenase		368	488	1303	2327
NiFe hydrogenase large subunit	1.12.2.1	381	532	1553	2344, 2244

Name of enzymes	EC numbers	*Quinella* genome bins and genes numbers			
		SR1Q5	SR1Q7	SR2Q5	SR3Q1
NiFe hydrogenase small subunit		380	531	1552	2245, 2502, 2343
NiFe hydrogenase cytochrome-*b* subunit		382	533	1554	2243, 2345
Methylmalonyl-CoA mutase		1136, 1137, 1910	573, 574, 2128	1082, 1083	526, 527, 2241, 2242
Methylmalonyl-CoA epimerase		485	969	1056	416, 775, 2283
Methylmalonyl-CoA decarboxylase, mmdA	4.1.1.41	486	969	1055	774
Methylmalonyl-CoA decarboxylase, mmdB		1999/767[c]	200	371, 1126	146
Methylmalonyl-CoA decarboxylase, mmdC		488	968	1055	773
Methylmalonyl-CoA decarboxylase, mmdD		2000	201	370, 1127	145
Oxaloacetate decarboxylase, oadA		441	1912	1198	582, 2395
Fumarate reductase subunit A		1181	2326	1218	1943
Fumarate reductase subunit B		1179	2328	1219	1944
Fumarate reductase subunit C		1182	2325	1217	1942
ATP synthase F_o, a subunit		1785	891	1188	1281
ATP synthase F_o, b subunit		1787	893	1190	1279

Name of enzymes	EC numbers	*Quinella* genome bins and genes numbers			
		SR1Q5	SR1Q7	SR2Q5	SR3Q1
ATP synthase F_o, c subunit		1786	892	1189	1280
ATP synthase F_1, α subunit	3.6.3.14	1789	896	1193	1277
ATP synthase F_1, β subunit		1791	898	1196	415
ATP synthase F_1, γsubunit		1790	897	1194	417
ATP synthase F_1, δ subunit		1788	894	1191	1278
ATP synthase F_1, ε subunit		1939	899	1197	414
Na^+/H^+ antiporter, nhac		520	867	1606	1074, 1545

[a]–, not detected.

[b]Fused genes present in all *Quinella* genome bins.

[c]Parts of the gene were present on the termini of different contigs and were assembled together to a generate full gene sequence.

Table A6.3 PTS transporter found in *Quinella* genome bins using TransportDB 2.0 database (Elbourne et al., 2017).

Quinella genome bin genes	Sub family	Substrate
SR1Q7_1675	EnzymeIIC	Ascorbate
SR1Q7_1676	EnzymeIIC	Ascorbate
SR1Q7_455	EnzymeIIC	Ascorbate
SR2Q5_487	EnzymeIIC	Ascorbate
SR1Q7_455	EnzymeIIC	Ascorbate
SR1Q5_1523	EnzymeIIA	Fructose
SR1Q7_1384	EnzymeIIAB	Fructose
SR1Q7_1385	EnzymeIIABC	Fructose
SR1Q7_1677	EnzymeIIA	Fructose
SR1Q7_454	EnzymeIIA	Fructose
SR1Q7_1754	EnzymeIIAB	Fructose
SR1Q7_1757	EnzymeIIAB	Fructose
SR2Q5_1321	EnzymeIIABC	Fructose
SR2Q5_1509	EnzymeIIABC	Fructose
SR2Q5_1757	EnzymeIIABC	Fructose
SR2Q5_181	EnzymeIIAB	Fructose
SR2Q5_488	EnzymeIIA	Fructose
SR3Q1_2251	EnzymeIIB	Fructose
SR3Q1_2252	EnzymeIIC	Fructose
SR3Q1_2253	EnzymeIIA	Fructose
SR3Q1_783	EnzymeIIABC	Fructose
SR3Q1_786	EnzymeIIAB	Fructose
SR3Q1_1576	EnzymeIIA	Fructose
SR1Q7_456	EnzymeIIB	Galactitol
SR2Q5_485	EnzymeIIB	Galactitol
SR2Q5_486	EnzymeIIB	Galactitol
SR1Q5_1287	EnzymeIIA	Glucitol/sorbitol
SR1Q5_434	EnzymeIIC	Glucitol/sorbitol
SR1Q5_435	EnzymeIIB	Glucitol/sorbitol
SR1Q7_802	EnzymeIIA	Glucitol/sorbitol
SR1Q7_432	EnzymeIIA	Glucitol/sorbitol
SR1Q7_617	EnzymeIIA	Glucitol/sorbitol
SR2Q5_1289	EnzymeIIA	Glucitol/sorbitol
SR2Q5_605	EnzymeIIA	Glucitol/sorbitol
SR3Q1_1106	EnzymeIIA	Glucitol/sorbitol
SR3Q1_589	EnzymeIIA	Glucitol/sorbitol
SR3Q1_341	EnzymeIIA	Glucitol/sorbitol
SR1Q5_2046	EnzymeIIABC	Glucose
SR1Q7_793	EnzymeIIABC	Glucose

Quinella genome bin genes	Sub family	Substrate
SR2Q5_586	EnzymeIIABC	Glucose
SR3Q1_1024	EnzymeIIABC	Glucose
SR3Q1_279	EnzymeIIABC	Glucose
SR3Q1_619	EnzymeIIABC	Glucose
SR1Q5_550	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR1Q5_873	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR1Q5_1782	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR1Q7_565	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR1Q7_683	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR1Q7_1999	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR2Q5_1011	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR2Q5_1128	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR3Q1_143	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR3Q1_1581	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR3Q1_1499	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR3Q1_942	EnzymeIIABC	Glucose/maltose/*N*-acetylglucosamine
SR1Q5_325	EnzymeIIABC	Mannitol
SR1Q7_1379	EnzymeIIABC	Mannitol
SR2Q5_900	EnzymeIIABC	Mannitol
SR1Q5_508	EnzymeIID	Mannose/fructose
SR1Q5_509	EnzymeIIC	Mannose/fructose
SR1Q5_510	EnzymeIIB	Mannose/fructose
SR1Q5_511	EnzymeIIA	Mannose/fructose
SR1Q7_1652	EnzymeIIA	Mannose/fructose
SR1Q7_1653	EnzymeIIB	Mannose/fructose
SR1Q7_1654	EnzymeIIC	Mannose/fructose
SR1Q7_1655	EnzymeIID	Mannose/fructose
SR1Q7_1899	EnzymeIIA	Mannose/fructose
SR1Q7_1900	EnzymeIIB	Mannose/fructose
SR1Q7_1901	EnzymeIIC	Mannose/fructose
SR2Q5_41	EnzymeIID	Mannose/fructose
SR2Q5_42	EnzymeIIC	Mannose/fructose
SR2Q5_43	EnzymeIIB	Mannose/fructose
SR2Q5_44	EnzymeIIA	Mannose/fructose

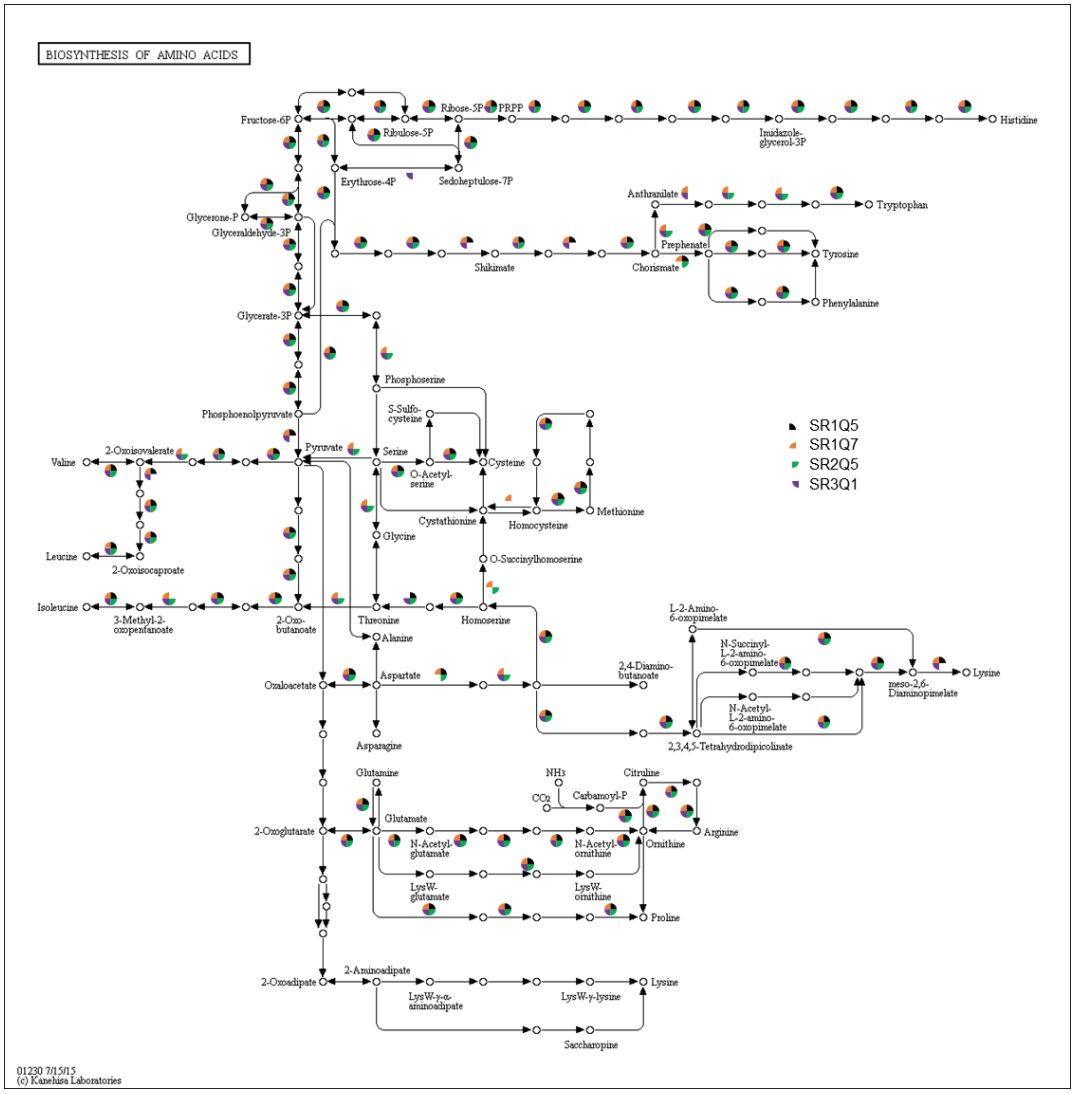


Figure A6.1 Genes implicated in amino acid synthesis in *Quinella*. The four quadrants indicate by colour if that gene was found in each genome bin. No colour means it was absent. The image of amino acid biosynthesis was reproduced with permission (Kanehisa et al., 2017).

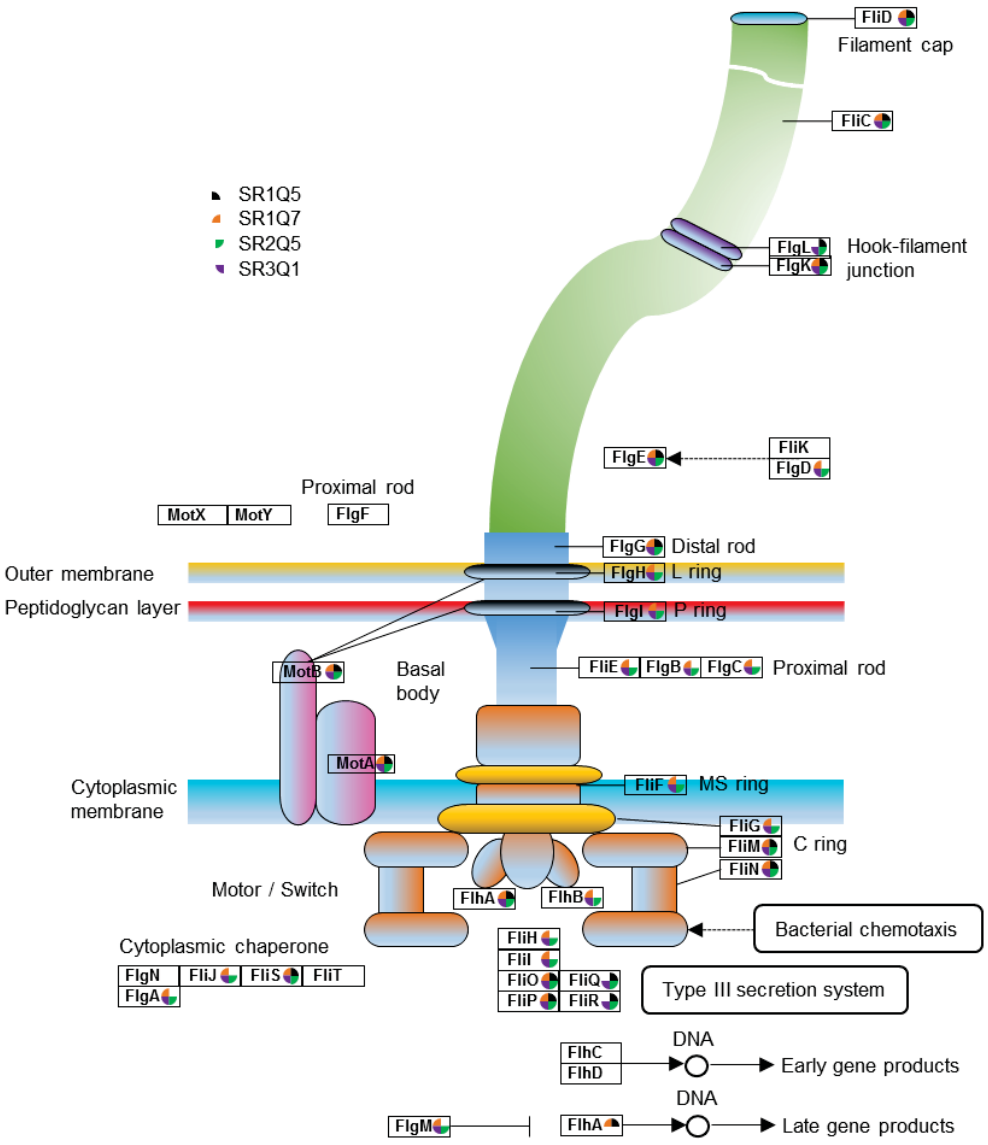


Figure A6.2 Genes associated with flagellar assembly in *Quinella*. The four quadrants indicate by colour if that gene was found in each genome bin. No colour means it was absent. The colours of the protein subunits are for differentiation only.