

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.

# **Identification and characterisation of rumen bacteria with prominent roles in the ruminal metabolism of forages**

A thesis presented in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy (Microbiology and Genetics) at Massey University,  
Palmerston North, New Zealand

**Sam C. Mahoney-Kurpe**

**2022**

# Abstract

This thesis documents the characterisation of two groups of rumen bacteria that are both prominent in forage-fed ruminants, with the aim to better understand their roles in ruminal metabolism. The first group, referred to as the R-7 group, has in recent years been shown to be one of the most abundant rumen bacterial groups, though the few isolated representative strains available were uncharacterised. Two strains of the group included in the Hungate1000 culture collection, R-7 and WTE2008, were selected for characterisation. To facilitate phylogenetic analyses of this group, the complete genomes of an additional three previously isolated R-7 group strains were sequenced. Genomic, phylogenetic and phenotypic characterisation of R-7 and WTE2008 demonstrated that despite their 16S rRNA gene sequences sharing 98.6-99.0% nucleotide identity, their genome-wide average nucleotide identity of 84% assigned them as separate species of a novel genus and family of the proposed order '*Christensenellales*' using the Genome Taxonomy Database. Phenotypic characterisation showed that the strains were identical in morphology, and both possessed the ability to degrade plant cell wall polysaccharides xylan and pectin, but not cellulose. Acetate, ethanol, hydrogen and lactate were produced by both strains, though R-7 produced greater amounts of hydrogen than WTE2008, which instead produced more lactate. Based on these analyses, it is proposed that R-7 and WTE2008 belong to separate species (*Aristaeella* gen. nov. *hokkaidonensis* sp. nov. and *Aristaeella lactis* sp. nov., respectively) of a newly proposed family (*Aristaeellaceae* fam. nov.).

The second bacterial group of interest, due to their dominant role in ruminal propionate production, was the *Prevotella* 1 group, following analyses of metatranscriptome datasets of rumen microbial communities of lucerne-fed sheep for dominant community members that express propionate pathway genes from succinate. Screening of 14 strains spanning the diversity of *Prevotella* 1 found that all except one *P. brevis* strain produced propionate in a cobalamin (vitamin B<sub>12</sub>)-dependent manner. To better understand the pathway and regulation of propionate production from succinate, a comparative multi-omics approach was used to test the hypothesis that propionate production is regulated by a cobalamin-binding

riboswitch. Scanning of a completed genome assembly of *Prevotella ruminicola* KHP1 identified four 'cobalamin' family riboswitches. However, the riboswitches were not in close proximity to genes putatively involved in converting succinate to propionate, nor were these genes arranged in a single operon. Comparative genomics of the 14 screened strains found that all strains possessed all homologues of candidate propionate pathway genes identified in the KHP1 genome. However, the 13 propionate-producing strains possessed a putative transporter and three subunits encoding a putative methylmalonyl-CoA decarboxylase upstream but antisense to two genes encoding methylmalonyl-CoA mutase subunits, whereas the non-producing strain did not. Comparative transcriptomics and proteomics of KHP1 cultures in the presence and absence of cobalamin demonstrated that some gene candidates were upregulated by cobalamin at the transcriptome level, including co-located genes annotated as phosphate butyryltransferase and butyrate kinase, despite the strain not producing butyrate, suggesting that propionate production may occur *via* propionyl phosphate. However, only both subunits of methylmalonyl-CoA mutase showed greater transcript and protein abundances in the presence of cobalamin. These results show that while some propionate pathway candidate genes were differentially expressed between cobalamin treatments, they did not appear to be under direct control of a cobalamin-binding riboswitch. This study has contributed to our understanding of the roles of both *Aristaeellaceae* fam. nov. and *Prevotella* 1 in ruminal metabolism.

## List of publications

- Mahoney-Kurpe, S. C., Palevich, N., Noel, S. J., Kumar, S., Gagic, D., Biggs, P. J., Janssen, P. H., Attwood, G. T., Moon, C. D. (2021) Complete genome sequences of three *Clostridiales* R-7 group strains isolated from the bovine rumen in New Zealand. *Microbiology Resource Announcements*, 10(26), e00310-21.
- Mahoney-Kurpe, S. C., Palevich, N., Noel, S. J., Gagic, D., Biggs, P. J., Soni, P., Reid, P. M., Koike, S., Kobayashi, Y., Janssen, P. H., Attwood, G. T., Moon, C. D. Description of *Aristaeella hokkaidonensis* gen. nov. sp. nov. and *Aristaeella lactis* sp. nov., two rumen bacterial species of a novel proposed family, *Aristaeellaceae* fam. nov. Under review for publication in the International Journal of Systematic and Evolutionary Microbiology (IJSEM).
- Mahoney-Kurpe, S. C., Palevich, N., Gagic, D., Biggs, P. J., Altshuler, I., Pope, P. B., Attwood, G. T., Moon, C. D. Transcriptomic and proteomic changes associated with cobalamin-dependent propionate production by the rumen bacterium *Prevotella ruminicola*. In preparation for submission to mSystems.

# Acknowledgements

First and foremost, I would like to express my gratitude to all of my supervisors, Dr Christina Moon, Dr Graeme Attwood, Dr Dragana Gagic, and Prof. Patrick Biggs, for the support and guidance you have given me throughout this project. I would also like to thank Dr Nik Palevich for essentially being an unofficial 5<sup>th</sup> supervisor, and also providing valuable input and support.

From AgResearch, I would also like to thank all of the Rumen Microbiology team for always being very helpful with the many queries I had in the lab. Thanks to Peter Reid for carrying out the gas chromatography work. Thanks to Dr Ambarish Biswas and Paul Maclean for bioinformatic support, and to Alan McCulloch for assistance with setting up software on the AgResearch server. Thanks also to Dr Peter Janssen for valuable input regarding the R-7 work, especially for the excellent suggestions for the etymology of the new proposed species.

From Massey, I would like to thank the Massey Microscopy and Imaging Centre, particularly to Niki Minards, Dr Pani Vijayan and Raoul Solomon for their assistance with the TEM work. Thanks to Ann Truter for postgraduate administrative support, and to Xiao Xiao Lin and Lorraine Berry for Massey Genome Service sequencing support.

Outside of AgResearch and Massey, thanks to collaborators Dr Satoshi Koike and Prof. Yasuo Kobayashi (Hokkaido University, Japan), and Dr Ianina Altshuler and Assoc. Prof. Phil Pope (Norwegian University of Life Sciences, Norway). Thanks also to Simon Zhang (Custom Science) for assistance with organising sequencing. I am grateful to the New Zealand Ministry of Business, Innovation and Employment (MBIE) for project and stipend funding through the Strategic Science Investment Fund Microbiomes programme.

# List of figures

Figure 1.1. Rumen microbial fermentation .....	4
Figure 1.2. Heterogeneity in structures of cell wall polysaccharides in plants.....	14
Figure 1.3. Contributions of different prokaryotic genera to polysaccharide degradation and fermentation in the rumen .....	18
Figure 1.4. Known pathways for propionate formation in gut bacteria .....	19
Figure 1.5. Schematic representation of protein degradation and fate of end products in the rumen .....	23
Figure 1.6. Plant-directed strategies of promoting ruminal fibre degradation .....	26
Figure 3.1. Maximum-likelihood phylogeny based on 16S rRNA gene sequences from members of the order ' <i>Christensenellales</i> '. .....	63
Figure 3.2. Matrix of pairwise average nucleotide identity (ANI) values between the genomes of members of ' <i>Christensenellales</i> '. .....	65
Figure 3.3. Cell morphology of R-7 <sup>T</sup> and WTE2008 <sup>T</sup> .....	66
Figure 3.4. Optical density and production of fermentation end products of R-7 <sup>T</sup> and WTE2008 <sup>T</sup> over 24 hours of growth.....	70
Figure 4.1. Alteration of the 'Top Percent' parameter enhances the specificity of taxonomic read assignment in MEGAN.....	81
Figure 4.2. Taxonomic distributions of reads mapping to functional categories involved in the pathway of propionate production from succinate.....	85
Figure 4.3. Screening of <i>Prevotella</i> 1 strains for cobalamin-dependent propionate production .....	93
Figure 4.4. Widespread dispersal of cobalamin riboswitches and candidate propionate pathway genes on the KHP1 genome .....	95
Figure 4.5. Conserved arrangement of methylmalonyl-CoA mutase, putative transporter and methylmalonyl-CoA decarboxylase genes across all propionate-producing Hungate1000 strains but absent in the non-producing P6B11 strain .....	97
Figure 4.6. Optical density and fermentation end product formation of KHP1 cultures throughout growth .....	98
Figure 4.7. Cobalamin-induced shifts in the KHP1 transcriptome and proteome .....	100
Figure 4.8. Gene and protein expression of genes in close proximity to predicted cobalamin family riboswitches .....	103
Figure 4.9. Impact of cobalamin on the expression of candidate genes putatively involved in converting succinate to propionate.....	105

## List of tables

<b>Table 2.1. Substrates</b> .....	35
<b>Table 2.2. Functional categories corresponding to enzymes involved in the conversion of succinate to propionate included in metatranscriptome analyses</b> .....	51
<b>Table 2.3. Bioinformatic tools used in this thesis</b> .....	53
<b>Table 3.1. Genome details of R-7 group strains</b> .....	58
<b>Table 3.2. Phenotypic characteristics of R-7<sup>T</sup> and WTE2008<sup>T</sup> compared to other related type strains</b> .....	68
<b>Table 7.1. Genome characteristics of R-7 group strains</b> .....	143
<b>Table 7.2. Complete cellular fatty acid profiles of R-7<sup>T</sup> and WTE2008<sup>T</sup></b> .....	144
<b>Table 7.3. Production of ethanol and total SCFAs by R-7<sup>T</sup> and WTE2008<sup>T</sup> on insoluble substrates</b> .....	145
<b>Table 7.4. Genome statistics of <i>P. ruminicola</i> KHP1</b> .....	146
<b>Table 7.5. Predicted 'cobalamin' family riboswitches in the KHP1 genome</b> .....	147
<b>Table 7.6. Homologues in the KHP1 genome of characterised genes converting succinate to propionate in other bacteria</b> .....	148
<b>Table 7.7. Positioning of candidate succinate pathway genes between succinate and propionate in the KHP1 genome</b> .....	151
<b>Table 7.8. Searching the Hungate1000 genomes of screened strains for homologues of candidate propionate pathway genes identified in the KHP1 genome</b> .....	152
<b>Table 7.9. Significantly differentially expressed KHP1 genes between cobalamin treatments</b> .....	153
<b>Table 7.10. Significantly differentially abundant KHP1 proteins between cobalamin treatments</b> .....	170



## Non-standard abbreviations

ANI	Average nucleotide identity
CAZymes	Carbohydrate-active enzymes
CE	Carbohydrate esterase
COG	Cluster of Orthologous Groups of proteins
DEG	Differentially expressed gene
DFM	Direct-fed microbial
dH <sub>2</sub> O	Distilled water
FDR	False discovery rate
GC	Gas chromatography
GRC	Global Rumen Census
GTDB	The Genome Taxonomy Database
GWAS	Genome-wide association study
IMG	Integrated Microbial Genomes
JGI	Joint Genome Institute
KEGG	Kyoto Encyclopedia of Genes and Genomes
LFQ	Label-free quantification
MAG	Metagenome-assembled genome
NCBI	National Center for Biotechnology Information
OD	Optical density
OTU	Operational Taxonomic Unit
PCA	Principal component analysis
PL	Polysaccharide lyase
RPM	Reads per million
SCFA	Short-chain fatty acid
SNP	Single nucleotide polymorphism
TEM	Transmission electron microscopy

# Table of contents

<b>Abstract</b> .....	<b>i</b>
<b>List of publications</b> .....	<b>iii</b>
<b>Acknowledgements</b> .....	<b>iv</b>
<b>List of figures</b> .....	<b>v</b>
<b>List of tables</b> .....	<b>vi</b>
<b>Table of contents</b> .....	<b>viii</b>
<b>1 Literature review</b> .....	<b>1</b>
<b>1.1 The ruminant digestive tract</b> .....	<b>2</b>
<b>1.2 The rumen microbiome</b> .....	<b>3</b>
1.2.1 Taxonomic composition of the rumen microbiota.....	4
1.2.1.1 Successional development of the mature rumen microbiota .....	5
1.2.1.2 Major microbial groups in the mature rumen .....	6
1.2.1.2.1 Bacteria .....	6
1.2.1.2.2 Archaea .....	7
1.2.1.2.3 Protozoa .....	7
1.2.1.2.4 Fungi.....	8
1.2.1.2.5 Viruses.....	9
1.2.1.3 Determinants of microbiota composition .....	9
1.2.1.4 Ecological characteristics of the rumen microbiota .....	11
1.2.2 Microbial metabolic processes in the rumen .....	12
1.2.2.1 Fibre degradation .....	13
1.2.2.1.1 Structural composition of the plant cell wall .....	13
1.2.2.1.2 Microbial fibre degradation processes.....	14
1.2.2.2 Rumen fermentation .....	17
1.2.2.2.1 SCFA production .....	17
1.2.2.2.2 Methane production .....	21
1.2.2.3 Protein metabolism .....	22
1.2.3 Rumen microbiome manipulation targets.....	24
1.2.3.1 Enhancing fibre degradation .....	24
1.2.3.2 Promoting propionate production .....	26

1.2.4	Elucidation of the uncharacterised rumen microbial majority through next-generation omics'	28
<b>1.3</b>	<b>Project objectives</b>	<b>29</b>
<b>2</b>	<b>Materials and Methods</b>	<b>31</b>
<b>2.1</b>	<b>Materials</b>	<b>32</b>
2.1.1	Biological material	32
2.1.2	Buffers and solutions	32
2.1.3	Media and media additives	33
2.1.3.1	Media additives	33
2.1.3.2	Media	37
<b>2.2</b>	<b>Methods</b>	<b>38</b>
2.2.1	Revival and maintenance of cultures	38
2.2.2	Colony purification	38
2.2.3	Measurement of fermentation end products	39
2.2.4	Cellular fatty acid analysis	40
2.2.5	Microscopy	40
2.2.6	Batch culture experiments	42
2.2.6.1	Substrate utilisation assay	42
2.2.6.2	Time series experiment of fermentation end product formation during growth	43
2.2.6.3	Screening of strains for cobalamin-dependent propionate production	43
2.2.6.4	Comparative omics' experiment of cobalamin-induced propionate production	44
2.2.7	Molecular methods	44
2.2.7.1	16S rRNA gene sequencing	44
2.2.7.2	Agarose gel electrophoresis	45
2.2.7.3	DNA extraction and genome sequencing	45
2.2.7.4	RNA extraction, library preparation and sequencing	46
2.2.7.5	DNA/RNA quantification and quality control	47
2.2.7.6	Protein extraction and mass-spectrometry	47
2.2.8	Bioinformatics	48
2.2.8.1	16S rRNA gene sequence analysis	48
2.2.8.2	Genome assembly and analyses	49
2.2.8.3	Creation of a Hungate1000 protein database	50
2.2.8.4	Metatranscriptome analyses	50
2.2.8.5	Transcriptome analyses	51

2.2.8.6	Proteome analyses .....	52
<b>3</b>	<b>Characterisation of the rumen R-7 bacterial group .....</b>	<b>54</b>
3.1	Introduction.....	55
3.2	Complete genome sequences of three <i>Clostridiales</i> R-7 group strains isolated from the bovine rumen in New Zealand .....	56
3.2.1	Abstract .....	56
3.2.2	Announcement .....	56
3.3	Description of <i>Aristaeella hokkaidonensis</i> gen. nov. sp. nov. and <i>Aristaeella lactis</i> sp. nov., two rumen bacterial species of a novel proposed family, <i>Aristaeellaceae</i> fam. nov.....	60
3.3.1	Abstract .....	60
3.3.2	Introduction .....	61
3.3.3	Habitat and isolation .....	61
3.3.4	16S rRNA gene phylogeny .....	62
3.3.5	Genome features .....	64
3.3.6	Physiology and Chemotaxonomy .....	65
3.3.7	Protologue .....	71
3.3.7.1	Description of <i>Aristaeellaceae</i> fam. nov.....	71
3.3.7.2	Description of <i>Aristaeella</i> gen. nov.....	71
3.3.7.3	Description of <i>Aristaeella hokkaidonensis</i> sp. nov.....	71
3.3.7.4	Description of <i>Aristaeella lactis</i> sp. nov.....	72
3.4	Discussion and Conclusion .....	74
<b>4</b>	<b>Unravelling the regulation of cobalamin-dependent propionate production by the rumen bacterium <i>Prevotella ruminicola</i> .....</b>	<b>77</b>
4.1	Introduction.....	78
4.2	Bioinformatic identification of dominant ruminal propionate producing bacteria .....	79
4.2.1	Dominance of <i>Prevotella</i> and <i>Succinivibrio</i> in the expression of enzymes predicted to convert succinate to propionate in metatranscriptome datasets.....	82
4.3	Transcriptomic and proteomic changes associated with cobalamin-dependent propionate production by the rumen bacterium <i>Prevotella ruminicola</i> .....	86
4.3.1	Abstract .....	86

4.3.2	Introduction .....	87
4.3.3	Materials and Methods .....	89
4.3.3.1	Biological material and growth conditions .....	89
4.3.3.2	Screening of <i>Prevotella</i> 1 strains for cobalamin-dependent propionate production .....	89
4.3.3.3	Comparative transcriptomics/proteomics experiment of cobalamin-induced propionate production .....	89
4.3.3.4	16S rRNA phylogeny .....	90
4.3.3.5	Genome sequencing and analyses .....	90
4.3.3.6	RNA-seq analyses .....	91
4.3.3.7	Proteome analyses .....	91
4.3.4	Results .....	92
4.3.4.1	Most ruminal <i>Prevotella</i> 1 strains produce propionate in the presence of cobalamin .....	92
4.3.4.2	The complete KHP1 genome contains four putative cobalamin riboswitches .....	93
4.3.4.3	Candidate succinate pathway genes are widely dispersed along the KHP1 chromosome and not in close proximity to predicted cobalamin riboswitches.....	94
4.3.4.4	Comparative genomics of <i>Prevotella</i> 1 strains differing in their propionate production .....	96
4.3.4.5	Comparative transcriptomics and proteomics of the impact of cobalamin on KHP1 gene expression.....	97
4.3.4.6	Differential expression of genes and proteins in close proximity to cobalamin riboswitches .....	101
4.3.4.7	Differential expression of some candidate propionate pathway transcripts between cobalamin treatments, but effects on the proteome limited to overexpression of both methylmalonyl-CoA mutase subunits .....	104
4.3.5	Discussion .....	106
<b>4.4</b>	<b>Discussion and conclusion.....</b>	<b>111</b>
<b>5</b>	<b>Conclusion and future directions.....</b>	<b>113</b>
<b>5.1</b>	<b>Conclusion .....</b>	<b>114</b>
<b>5.2</b>	<b>Future directions.....</b>	<b>115</b>
5.2.1	Future work characterising the R-7 group.....	115
5.2.2	Future work characterising propionate production in <i>P. ruminicola</i> .....	117
<b>5.3</b>	<b>Towards the development of synthetic rumen microbial consortia</b>	<b>119</b>
<b>6</b>	<b>Bibliography .....</b>	<b>121</b>
<b>7</b>	<b>Appendix .....</b>	<b>139</b>

<b>7.1</b>	<b>Complete genome sequences of three <i>Clostridiales</i> R-7 group strains isolated from the bovine rumen in New Zealand .....</b>	<b>140</b>
<b>7.2</b>	<b>Genome statistics of resequenced R-7<sup>T</sup> and WTE2008<sup>T</sup> genomes....</b>	<b>143</b>
<b>7.3</b>	<b>Cellular fatty acid profiles .....</b>	<b>144</b>
<b>7.4</b>	<b>Ethanol and SCFA production on insoluble substrates .....</b>	<b>145</b>
<b>7.5</b>	<b>Genome details of the resequenced <i>P. ruminicola</i> KHP1 assembly.</b>	<b>146</b>
<b>7.6</b>	<b>Sequences and locations of cobalamin riboswitches .....</b>	<b>147</b>
<b>7.7</b>	<b>Protein homologues, encoded by the KHP1 genome, of characterised proteins/enzymes involved in converting succinate to propionate in other bacteria .....</b>	<b>148</b>
<b>7.8</b>	<b>Genome location of candidate KHP1 succinate pathway genes between succinate and propionate.....</b>	<b>151</b>
<b>7.9</b>	<b>Homologues of candidate KHP1 propionate pathway genes in Hungate1000 assemblies of screened strains .....</b>	<b>152</b>
<b>7.10</b>	<b>Differentially abundant KHP1 genes.....</b>	<b>153</b>
<b>7.11</b>	<b>Differentially abundant KHP1 proteins.....</b>	<b>170</b>

# 1 Literature review

Ruminants have shared a close relationship with humans since ancient times (Ajmone-Marsan et al., 2010), owing to their ability to produce a variety of products of both nutritional and practical use (Russell & Rychlik, 2001). Ruminants are farmed intensively in New Zealand, with a total of 37 million animals farmed over 9 million hectares of pastoral land (Statistics New Zealand, 2019). The resulting products of this farming (mainly meat, milk and wool) account for approximately 40% of New Zealand's total annual exports (Statistics New Zealand, 2020). As the global human population continues to increase, the demand for ruminant products is expected to increase (Springmann et al., 2016; Tilman & Clark, 2014). However, the environmental impacts of ruminant production mean that more efficient pastoral production systems are needed (Beukes et al., 2010).

## 1.1 The ruminant digestive tract

Ruminants have evolved a multi-compartmented stomach to facilitate the degradation and extraction of energy from especially recalcitrant lignocellulosic (fibrous) feed material. The ruminant stomach consists of four compartments: the reticulum, rumen, omasum, and abomasum. The reticulum is an anterior pouch connected with the much larger rumen. The reticulum and rumen function together and provide conditions in which digesta can be held anaerobically to facilitate fermentation, and are thus generally referred to collectively as the reticulo-rumen, or more simply the 'rumen'. Digesta passes from the rumen through the reticulo-omasal orifice, which functions as a sieve permitting the passage of only small particles of digesta into the omasum, which acts as the absorption site of water and vitamins. The digesta then enters the abomasum, which is analogous to the stomach of monogastric mammals. The abomasum is an acid-secreting stomach that produces pepsin and other digestive enzymes to further degrade digesta. Digesta flows into the small intestine and through the hindgut, which is also analogous in both structure and function to that of monogastric mammals (Hungate, 1966).

At birth, the rumen of the newborn is undeveloped and small compared with the abomasum. Newborn ruminants are initially fed colostrum from their mothers, and upon the feeding of



milk, the act of suckling causes a muscular contraction in the reticulo-omasal groove resulting in the ingested liquid bypassing the rumen and flowing directly into the abomasum. However, with increasing solid feed intake the rumen develops rapidly, eventually becoming the largest compartment of the ruminant stomach typically holding an amount of digesta and saliva weighing approximately one seventh that of the entire animal (Hungate, 1966). This increased rumen size also increases the surface area from which the generated short-chain fatty acids (SCFAs) can be absorbed by the animal as its major source of energy (Bergman, 1990). The temperature of the rumen is maintained at a constant 39-40°C, though this increases slightly after feeding due to increased microbial fermentative processes (Dale et al., 1954). Ruminal pH is also maintained between 5.8 and 6.8 by the bicarbonate present in secreted host saliva, which acts to counter the acidity from the SCFAs produced by ruminal fermentation (Hungate, 1966).

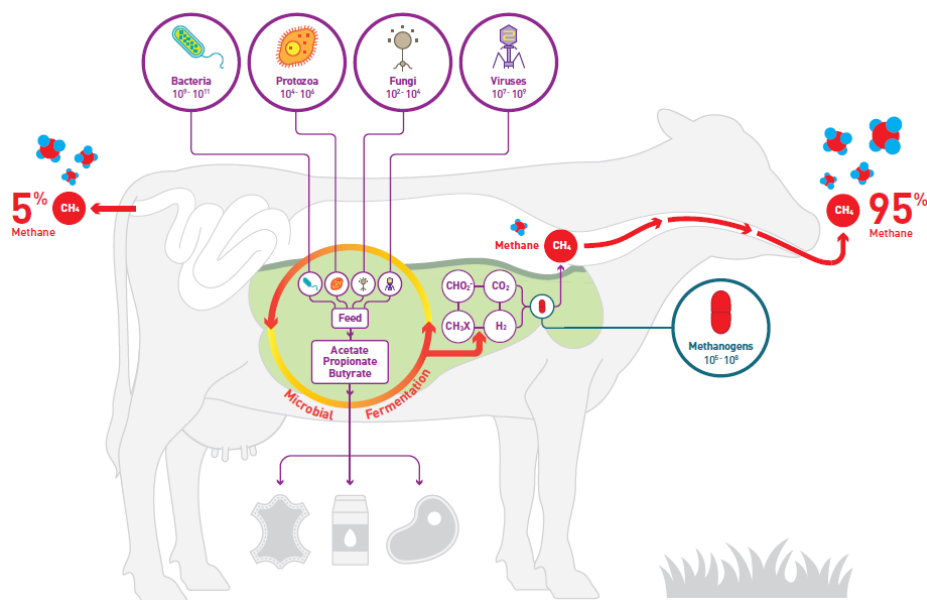
## 1.2 The rumen microbiome

Ruminants rely on the enzymatic actions of a complex microbial community residing in their rumens to degrade lignocellulosic feed material, as they do not possess the necessary enzymes themselves. The process of digestion is initiated by mastication by the host, which mechanically breaks down the ingested feed and creates damaged sites that are colonised by resident microbes (Beauchemin, 1992). The rumen microbial community further instigates the breakdown of lignocellulose through its production of a wide range of enzymes. Physical and enzymatic breakdown of ingested feed provides substrate for microbes to ferment to generate SCFAs, which the host then absorbs across the rumen epithelium (Bergman, 1990). Additionally, microbial protein leaving the rumen accounts for up to 90% of the amino acids reaching the small intestine (Russell & Rychlik, 2001). While the importance of the rumen microbiome has been long recognised (Hungate, 1966), the lack of cultured representatives of some of the major microbial groups of the rumen microbiome had prevented their characterisation using traditional microbiological methods (Krause et al., 2013). However, the increased awareness of these groups brought about by cultivation-independent studies spurred intensive cultivation efforts that have isolated representative strains of many of these uncharacterised but abundant rumen bacterial groups, many of which have been genome

sequenced (Seshadri et al., 2018). In addition to the rapidly increasing number of rumen meta'omic datasets (McCann et al., 2014), these technological advancements hold promise to greatly enhance our understanding of the structure and function of the rumen microbiome, which may ultimately give rise to novel strategies for manipulating the rumen microbiome to improve host health, production efficiency and environmental outcomes. However, before such goals can be achieved, a greater understanding of the rumen microbiome is still necessary.

### 1.2.1 Taxonomic composition of the rumen microbiota

The rumen hosts a densely populated microbial community, consisting of up to  $10^{11}$  microbial cells per gram of rumen contents. This community comprises a myriad of bacteria, archaea, fungi, protozoa and viruses (Figure 1.1). This section describes the successional development of the rumen microbiome from birth, before outlining the roles of major microbial groups in the mature rumen. The factors that collectively shape rumen microbial community composition are then discussed, before summarising the ecological characteristics of the established rumen microbiome.



**Figure 1.1. Rumen microbial fermentation.** Microbial numbers are listed as per gram of ruminal contents. Abbreviations: CO<sub>2</sub>, carbon dioxide; H<sub>2</sub>, hydrogen, CHO<sub>2</sub><sup>-</sup>, formate; CH<sub>3</sub>X, representing methoxy compounds or methylamines; CH<sub>4</sub>, methane. Figure and legend adapted from FAO (2019).

### 1.2.1.1 Successional development of the mature rumen microbiota

Whether the microbiota of mammalian gastrointestinal tracts is completely acquired after birth, or is colonised *in utero* has long been a matter of debate, and it was generally thought that the rumen was sterile prior to birth (Mackie et al., 1999). Some initial reports of *in utero* colonisation using cultivation-independent molecular methods were received with scepticism due to a lack of controls that could confidently rule out the possibility of contamination (Salter et al., 2014; Perez-Muñoz et al., 2017). However, a recent study assessed various components of the gastrointestinal tracts of calf foetuses at 5, 6 and 7 months of gestation, and included appropriate control measures, which confirmed the presence of a pioneer microbiota with distinct communities forming in different compartments of the gastrointestinal tract. These included bacteria of the phyla *Actinobacteria*, *Firmicutes* and *Proteobacteria*, and archaea of the order *Methanobacteriales* present in foetal rumen fluid. The authors further confirmed the viability of bacteria in the rumen fluid by culturing (Guzman et al., 2020). Following birth, the rumen is rapidly further colonized in the first 1-2 days of life by bacteria derived from the mother and surroundings of the newborn (Yáñez-Ruiz et al., 2015). Ingested feed material provides substrates for the subsequent development of the rumen (Hobson & Stewart, 1997). Rumen methanogens also establish in the rumen 1-3 days after birth (Fonty et al., 1987; Skillman et al., 2004). Rumen fungal and protozoal communities establish slightly slower, instead taking approximately ten days to colonize the rumen (Fonty et al., 1988). During the first few days of life a dramatic shift in bacterial community composition occurs, whereby facultatively anaerobic 'first colonizers' of the phyla *Proteobacteria* are replaced by obligately anaerobic *Bacteroidetes*, particularly of the genus *Prevotella* (Dill-McFarland et al., 2017; Jami et al., 2013; Rey et al., 2013). Alpha-diversity (a measurement of the extent of microbial diversity within samples) increases, whereas beta-diversity (a measurement of the compositional similarity of microbial communities between samples) (Whittaker, 1972) instead decreases with age (Dill-McFarland et al., 2017). Rather than reaching a level of complete taxonomic stability, the bacterial community of Holstein dairy cows fed a total mixed ration diet was shown to be dynamic in its composition when sampled over two successive lactations (Jewell et al., 2015).

## 1.2.1.2 Major microbial groups in the mature rumen

### 1.2.1.2.1 Bacteria

Bacteria are the most diverse and abundant microbial group in the rumen. Significant insight into rumen bacterial community composition was gained from the Global Rumen Census (GRC) study, which used cultivation-independent metabarcoding to compare bacterial, archaeal (section 1.2.1.2.2) and protozoal (section 1.2.1.2.3) community composition of 742 ruminants of various species fed different diets from different parts of the world. Results of this study found that the seven most abundant bacterial genera in the rumen comprise approximately 70% of total rumen bacterial community, which were *Prevotella*, *Butyrivibrio* and *Ruminococcus*, as well as unclassified groups of *Lachnospiraceae*, *Ruminococcaceae*, *Bacteroidales* and *Clostridiales*. These groups were universally present across all animals of the study, thereby representing a core rumen bacterial microbiota (Henderson et al., 2015). *Prevotella* is often the most abundant rumen bacterial genus of mature ruminants (Henderson et al., 2015), although currently characterised representatives only comprise a small proportion of this total population (McCann et al., 2014). The bacterial groups that dominate the rumen of a given animal varies considerably depending on a variety of factors, including host genotype, environmental factors, and especially the host diet (see section 1.2.1.3). The rumen bacterial community participates in almost all major rumen microbial functions including fibre degradation, fermentation of the resulting sugars, and protein metabolism (discussed further in Section 1.2.2). Bacterial protein is also the main source of amino acids reaching the small intestine (Russell & Rychlik, 2001). It is also important to note that in addition to abundant groups, other bacterial groups also play important roles while typically having low (<1%) relative abundances, such as the specialist cellulose degrader *Fibrobacter succinogenes* (Stevenson & Weimer, 2007) and some hyper ammonia-producing bacteria such as *Clostridium sticklandii* (Krause & Russell, 1996) and *Peptostreptococcus anaerobius* (Paster et al., 1993).

#### 1.2.1.2.2 Archaea

The rumen hosts an ecologically significant archaeal community, the members of which all appear to be methanogenic (Janssen & Kirs, 2008). The most abundant genus is *Methanobrevibacter*, and isolates of several species have been genome sequenced (Kelly et al., 2014; Kelly et al., 2016; Lambie et al., 2015; Leahy et al., 2010). Some archaeal methanogens in the rumen form ectosymbioses with protozoal species (Finlay et al., 1994), which in some cases may be mediated by archaeal adhesins expressed on their surface with affinity for surface proteins of hydrogen-producing protozoa (Ng et al., 2015). An amplicon sequencing-based study assessed rumen methanogen communities in sheep on various diets and geographical origins, and found that more than half of the reads per sample mapped to only five species-level operational taxonomic units (OTUs) detected across all 252 samples analysed (Seedorf et al., 2015). This suggests that the archaeal community in the rumen is dominated by a small number of abundant taxa. The GRC confirmed this finding, and further showed that rumen archaeal communities of different animals in different parts of the world show a high degree of similarity (Henderson et al., 2015).

#### 1.2.1.2.3 Protozoa

The rumen also hosts a protozoal community which can comprise up to 50% of total rumen microbial biomass (Newbold et al., 2015). The protozoa consist mostly of two major groups belonging to the subclass *Trichomastix*, referred to as holotrichs and entodiniomorphids (Williams & Coleman, 1997). Protozoal communities appear to exhibit a greater degree of variability between individual animals compared with prokaryotes (Henderson et al., 2015).

The difficulties involved in isolating protozoa in axenic culture complicates the assignment of particular functions to individual protozoal species. Metagenomic screening has confirmed the presence of glycosyl hydrolases in the genomes of rumen protozoa (Bera-Maillet et al., 2009), consistent with their suggested contribution to fibre degradation (Gijzen et al., 1988). Numerous defaunation (removing rumen protozoa) experiments have demonstrated that

fauna-free animals also often show decreased rates of protein degradation than normally faunated animals, thereby improving nitrogen uptake efficiency of the host and reducing rates of nitrogen excretion (Newbold et al., 2015). Holotrich protozoa also produce hydrogen, and thus establish metabolic interactions with hydrogen-utilising methanogens (Hillman et al., 1989). Rumen protozoa also predate bacteria (Bonhomme, 1990), and as such, many of the impacts associated with rumen protozoa are likely due to their impacts towards the prokaryotic rumen microbiome rather than being a direct consequence of the protozoal community. Insights into the impacts of various rumen protozoal groups was recently shown in a study comparing different groups of sheep inoculated with differing combinations of protozoa and found that the degree of bacterial degradation per protozoal cell was related to protozoal cell size, whereby larger protozoa had a greater bactericidal activity than smaller entodiniomorphid protozoa (Belanche et al., 2015).

#### 1.2.1.2.4 Fungi

While less phylogenetically diverse than rumen prokaryotic communities, the rumen fungal community can account for as much as 20% of the total rumen microbial biomass (Edwards et al., 2017). This community is composed primarily of species belonging to the phylum *Neocallimastigomycota*, an early-diverging lineage most closely related to the *Chytridiomycetes* (Hibbett et al., 2007). The rhizoids of rumen fungi are able to physically penetrate and break apart ingested plant tissue and produce enzymes which cleave ester linkages between lignin and hemicellulose (Akin & Borneman, 1990), thereby allowing other fibrolytic rumen microbes access to otherwise inaccessible cellulose (Ho et al., 1988). Due to these superior fibrolytic abilities, there is considerable interest in using fungal lignocellulolytic enzymes for biotechnological applications such as the generation of cellulosic biofuels (Ranganathan et al., 2017) and biogases (Procházka et al., 2012), or to incorporate such fungal strains into direct-fed microbial (DFM) strategies (Hartinger & Zebeli, 2021; Puniya et al., 2015). However, the difficulties involved in setting up efficient continuous cultures of rumen anaerobic fungi currently impedes their widespread commercialisation (Gruninger et al., 2014).

#### 1.2.1.2.5 Viruses

The rumen hosts a diverse viral community dominated by bacteriophage which infect a wide variety of rumen bacteria, although to date this community is poorly understood. Despite rumen bacteriophages being first isolated as early as 1966 (Adams et al., 1966), only a very select few with potential biotechnological applicability have since been characterised. Due to the absence of universal phylogenetic markers, viromes cannot be characterised using amplicon-based metabarcoding techniques (Sullivan, 2015). However, a shotgun metagenomic survey of virus-like DNA from the rumen of dairy cattle found large taxonomic variability between animals while the functional profile of the virome appears more conserved, suggestive of functional redundancy (Ross et al., 2013). More recently, Gilbert et al. (2017) sequenced the genomes of five lytic phages capable of infecting various characterised rumen bacteria, providing insights into the infection/defence strategies occurring between each bacteriophage and their bacteria they infect. Future work is necessary to determine the extent to which variations in the rumen virome affect wider microbial community composition and metabolic function in the rumen.

#### 1.2.1.3 Determinants of microbiota composition

The composition of microbes residing in the rumen is dynamically shaped by a wide range of factors, resulting in individual animals exhibiting unique microbiomes. Concurrently, a 'core' (Shade & Handelsman, 2012) rumen microbiome has been identified, consisting of a number of species that are typically present across different animals regardless of their breed or diet regimen (Creevey et al., 2014; Henderson et al., 2015; Jami & Mizrahi, 2012). This section outlines the various biotic and abiotic influences which collectively determine community composition of the rumen microbiome.

Diet is the major driver of rumen microbial community composition (Ellison et al., 2014; Henderson et al., 2015; Tapio et al., 2017). For example, animals fed forage-rich diets tended to have higher relative abundances of unclassified *Bacteroidales* and *Ruminococcaceae*, whereas animals on high-grain diets instead have higher relative abundances of unclassified

*Succinovibrionaceae*, and *Prevotella* (Henderson et al., 2015). Similarly, species involved in fibre degradation tend to increase in abundance in animals fed high fibre diets (Carberry et al., 2012; Zhang et al., 2017). In natural settings rumen microbial community composition is also influenced by season, as the *Bacteroidetes* BS11 group was shown to increase in abundance when moose were grazed on a pasture in winter when plants contained higher amounts of lignin and lower protein than when grazed on the same pasture in summer (Solden et al., 2017).

While having less of an influence than diet, the host also actively shapes rumen microbial community composition (Henderson et al., 2015). For example, subtle species-specific differences in both total microbiome composition (Kittelmann et al., 2013) as well as methanogen communities (King et al., 2011) have been identified. Additionally, another study assessed both the microbial community and metabolism of hybrid crosses of deer and elk to that of their parents under controlled conditions, and found genotype-specific effects towards both rumen microbial community composition and metabolism of host animals (Li et al., 2016). Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in regions of chromosomes of the bovine host correlated with abundances of specific rumen bacterial groups, further demonstrating the host genotype effect towards rumen community composition (Abbas et al., 2020; Li et al., 2019; Sasson et al., 2017).

Microbes in the rumen interact extensively with one another, both mutualistically and antagonistically (Wolin et al., 1997). As such, the culmination of these interactions can influence the composition of the rumen microbiome. Such interactions include those between closely related microbial strains such as those that produce antimicrobial bacteriocins (Russell & Mantovani, 2002), to inter-kingdom interactions between distantly related organisms. For example, protozoa are predators of bacteria and symbionts with methanogenic archaea, and thus the composition of the protozoal community exerts an influence on the prokaryotic community composition (Solomon et al., 2022). Similarly, fungi



develop both mutualistic (Cheng et al., 2009) and antagonistic (Bernalier et al., 1993; Swift et al., 2021) relationships with different groups of rumen prokaryotes.

#### 1.2.1.4 Ecological characteristics of the rumen microbiota

The term ‘functional redundancy’ refers to the overlapping physiological capabilities of different taxa, which enables compositionally distinct communities to fulfil the same functions (Allison & Martiny, 2008). Similar to other microbial ecological contexts (Burke et al., 2011), evidence of functional redundancy exists in the rumen microbiome (Weimer, 2015). For example, mapping metagenomic reads to a metabolic network found that rumen microbiomes with differing taxonomic compositions were more similar in terms of their metabolic potential (Taxis et al., 2015). A study assessing the metatranscriptomes of four Holstein cows over time during feeding, also found that microbiota compositions of individual animals assessed were highly variable, whereas their activities showed functional redundancy across and within microbial domains during feed degradation (Söllinger et al., 2018). This therefore emphasises the value of complementing assessments of community composition with methods that also provide functional insights into the rumen microbiota such as shotgun metagenomic (Quince et al., 2017), metatranscriptomic (Bashiardes et al., 2016), metaproteomic (Kunath et al., 2019) and metabolomic (Sangubotla & Kim, 2021) approaches, which are fortunately becoming more prevalent as the associated costs continue to decrease.

Aside from the core microbiome that exists across ruminants irrespective of their diet, host genetics or lifestyle (Henderson et al., 2015), a large number of taxa also tend to differ between different microbiomes. This individuality in community composition between different hosts has been shown for rumen protozoal communities (Eadie, 1962; Henderson et al., 2015), the fibrolytic bacterial community (Weimer et al., 1999), as well as the total rumen bacterial community (Jami & Mizrahi, 2012). Some studies have also shown individuality between animals in their methanogen communities (King et al., 2011; Zhou et al., 2012), although the GRC instead found that archaeal communities showed a high degree of similarity between animals (Henderson et al., 2015). The abovementioned GWAS that

identified SNPs at particular loci in host genomes associated with differential abundances of groups and single OTUs of rumen bacteria (Abbas et al., 2020; Li et al., 2019; Sasson et al., 2017) suggests that the host genotype at least partially accounts for the observed host specificity in rumen community composition, though the mechanisms responsible remain elusive.

Despite showing subtle but significant temporal fluctuations in community composition (Jewell et al., 2015), the mature rumen microbiome also exhibits resilience to change (Weimer, 2015). This was clearly illustrated in a study whereby a one-time near-total (>95%) exchange of ruminal contents between dairy cows differing in ruminal bacterial community composition resulted in ruminal pH and SCFA concentrations returning to pre-exchange values, and bacterial community composition to be more similar to pre-exchange communities after 61 days (Weimer et al., 2010). Similarly, while near total exchange of rumen contents between high and low milk production efficiency (MPE) cows resulted in a reversal of MPE status that corresponded with a concomitant shift in bacterial (Weimer et al., 2017) but not fungal community composition (Cox et al., 2021), this reversal in MPE status lasted only ~10 days (Cox et al., 2021; Weimer et al., 2017). As a result of the apparent stability of the mature rumen microbiome, strategies for manipulating the rumen microbiome through inoculation of introduced strains may be more successful if inoculation occurs at an early stage of host development before such stabilisation occurs (Yáñez-Ruiz et al., 2015), providing an opportunity to influence both microbiome establishment and gastrointestinal tract development.

### 1.2.2 Microbial metabolic processes in the rumen

The rumen microbiome catalyses a number of important metabolic processes that have crucial implications for the nutrition, development and physiology of their hosts. Microbial degradation of ingested lignocellulose and fermentation of the resulting sugars results in the generation of SCFAs, which the host then utilises as its major energy source. Additionally, microbes are responsible for both degrading protein in the rumen, as well as acting as a major

protein source for their hosts in the lower digestive tract. The following section describes each of these processes in more detail.

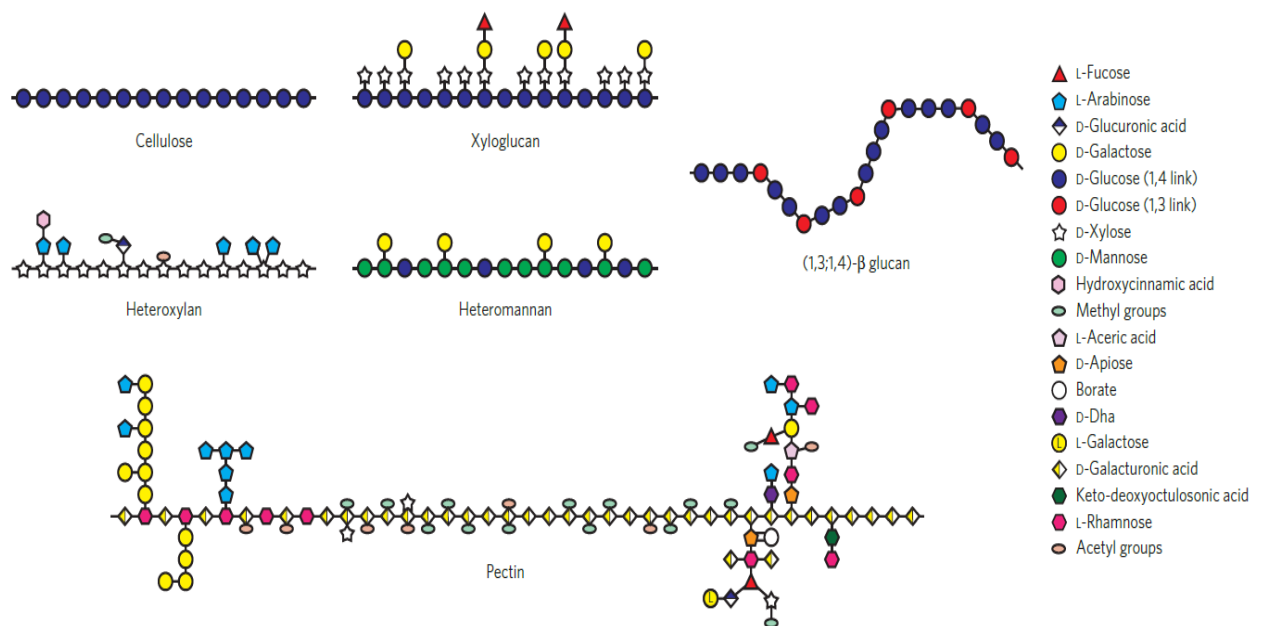
### 1.2.2.1 Fibre degradation

#### 1.2.2.1.1 Structural composition of the plant cell wall

The evolutionary success of ruminants can be largely attributed to their ability to digest the recalcitrant structural lignocellulolytic polysaccharides (fibre) found in plant cell walls (Hungate, 1966). Plant 'fibre' refers broadly to the indigestible portion of plant cell walls, encompassing a wide variety of plant polysaccharides (Jung et al., 2012). Plants contain both primary and secondary cell walls. Primary cell walls surround growing cells, and are broadly classified as belonging to one of two types: Type-I primary cell walls are primarily composed of cellulose surrounded by hemicelluloses rich in xyloglucans, pectin, structural proteins and phenolics. Type-II primary cell walls are found exclusively in grasses, and instead contain cellulose embedded in glucuronoarabinoxylans, mixed linkage glucans and phenolics. Secondary cell walls are much thicker and act to further fortify mature tissues, and are composed mainly of cellulose, the hemicelluloses xylan and glucomannan, and lignin (Zhong & Ye, 2015). Plant cell walls also contain various proteins responsible for a number of structural and developmental roles (Cassab, 1998). The overall structure of the plant cell wall can differ considerably between different plant species, as well as between different tissues of individual plants (Burton et al., 2010).

Primary and secondary plant cell walls are composed largely of cellulose, hemicelluloses, pectin, and lignin. Cellulose is considered the core component of the plant cell wall, and is composed of long polymeric chains of  $\beta$ -1,4-linked glucose units that form microfibrils held together by hydrogen bonds (Figure 1.2) (Somerville, 2006). The second most abundant component of the plant cell wall is hemicellulose, which refers to a number of heteropolymers comprised of various pentose and hexose sugars and sugar acids, that may be branched, and provide the plant cell wall with additional support through crosslinking cellulose and pectin (Scheller & Ulvskov, 2010). Pectin is structurally the most complex of the plant

polysaccharides, and is composed of predominantly a galacturonic acid residue backbone with various side-chains (Figure 1.2) (Voragen et al., 2009). Lignin is composed mainly of three cross-linked phenolic components- *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, and acts as a molecular ‘glue’ providing further support to the plant cell wall and preventing pathogen invasion (Campbell & Sederoff, 1996). Lignin cannot be degraded by rumen microbes, and thus its presence in plant cell walls significantly limits the degradation of ingested fibre (Vanholme et al., 2010).



**Figure 1.2. Heterogeneity in structures of cell wall polysaccharides in plants.** The backbone structures of these polysaccharides are based on (1,4)- $\beta$ -linked monosaccharides in the case of cellulose, xyloglucan, (1,3;1,4)- $\beta$ -glucan, heteroxylan and heteromannan, whereas the backbone of the galacturonans that are major constituents of pectic polysaccharides are based on (1,4)- $\alpha$ -linked chains of galacturonosyl residues. Extensive chain aggregation of the type found in cellulose is sterically hindered in the noncellulosic wall polysaccharides mainly through the addition of short oligosaccharides, monosaccharides or acetyl groups. However, in the (1,3;1,4)- $\beta$ -glucans, main chain aggregation is prevented through the irregular conformation of the polysaccharide. Figure and legend adapted from Burton et al. (2010) with permission obtained through RightsLink.

#### 1.2.2.1.2 Microbial fibre degradation processes

Fibrolitic microbes employ a variety of different mechanisms to break down ingested fibre through their expression of a wide variety of glycoside hydrolases (GHs) (Naumoff, 2011),

carbohydrate esterases (CEs) (Nakamura et al., 2017) and polysaccharide lyases (PLs) (Linhardt et al., 1986) which catalyse the degradation of bonds within and between carbohydrates, or between carbohydrates and non-carbohydrate molecules (Krause et al., 2003). This process is initiated by physical attachment of fibrolytic microbes to fibre, particularly to the damaged surfaces of the feed, which occurs rapidly following ingestion and chewing (Koike et al., 2003). Some fibrolytic bacteria degrade fibre through the expression of multi-enzyme complexes referred to as cellulosomes (Bayer et al., 1994), which anchor numerous fibrolytic enzyme complexes together to cooperatively deconstruct lignocellulose (Artzi et al., 2017). The backbone of the cellulosome is referred to as a scaffoldin, which contains numerous cohesin modules that bind to dockerins located on various fibrolytic enzymes and carbohydrate binding modules. However, the specialist cellulose degrader *Fibrobacter succinogenes* secretes endoglucanases resulting in the breakdown of cellulose to form cellodextrins, which are then taken up by the cell and further degraded in the periplasm (Burnet et al., 2015). Rumen eukaryotes are also actively involved in fibre degradation. The rhizoids of rumen fungi are capable of penetrating ingested fibre, thereby permitting the breakdown of otherwise inaccessible cellulose (Orpin, 1977). Fibrolytic enzymes produced by rumen fungi have also been shown to be exceptionally active (Wilson & Wood, 1992) which has spurred interest in these enzymes having biotechnological applicability. Some species of rumen entodiniomorphid protozoa are also capable of engulfing and degrading cellulose particles *via* phagocytosis (Coleman, 1992). Contrasting the recalcitrance of cellulose, hemicelluloses are efficiently degraded by rumen bacteria, especially by members of *Butyrivibrio* and *Prevotella* (Emerson & Weimer, 2017; Morais & Mizrahi, 2019).

Recent advances in next-generation sequencing technologies have unveiled a great magnitude of diversity of lignocellulolytic enzymes present in the rumen. Metagenomic analyses have detected thousands of novel glycoside hydrolases (Brulc et al., 2009; Hess et al., 2011) highlighting the potential of the rumen microbiome as a reservoir of biotechnologically useful enzymes for biomass conversion. A number of studies have also assessed the metatranscriptome of the fibre-adherent rumen community, which provides additional insights into the gene expression of the active microbial community in the sample (Bashiardes et al., 2016). Dai et al (2015) revealed that glycoside hydrolases mostly belonging

to four families (GH5, GH9, GH45 and GH48) are expressed in the rumen of dairy cows fed corn straw, predominantly by bacteria belonging to the genera *Ruminococcus*, *Prevotella* and *Fibrobacter*. Transcripts that encode proteins of a number of GH families involved in hemicellulose degradation were also detected, which were assigned mostly to *Ruminococcus* and *Prevotella* (Dai et al., 2015). Shinkai et al. (2016) assessed the fibre-adherent metatranscriptome of cows fed Timothy hay and showed a number of the same GH families described in Dai et al. (2015), the majority of which were expressed by *Fibrobacter succinogenes*. Comtet-Marre et al. (2017) simultaneously assessed both the prokaryotic and eukaryotic fibrolytic communities of dairy cows fed a total mixed ration diet, and found that the eukaryotic component of the fibrolytic community was vastly underrepresented by metagenomic analyses compared with metatranscriptomic data, suggestive of a high metabolic activity relative to its biomass.

In addition to the abovementioned insights from metagenomic and metatranscriptomic studies of the complex communities present in rumen samples, genomic, transcriptomic and proteomic analyses of pure cultures and/or co-cultures of rumen bacteria are greatly improving our understanding of the contributions of specific rumen bacteria to fibre degradation. For example, a comparison of the genomes of three *Ruminococcus albus* and *Ruminococcus flavefaciens* isolates showed that strains of each species shared vastly different genes encoding cellulosome components (Dassa et al., 2014), illustrating the genetic basis underlying species-specific differences in mechanisms of adhesion to and hydrolysis of ingested plant material (Krause et al., 2003). Comparative transcriptomics of *Ruminococcus albus* 7 grown on cellulose versus cellobiose found that cellulose-degrading cultures had increased expression of transcripts encoding carbohydrate binding module 37 (CBM37), and enzymes involved in tryptophan biosynthesis (Christopherson et al., 2014). Insights into the genetic basis underlying the ability of *Prevotella bryantii* to degrade xylan were gained using comparative transcriptomics to identify genes upregulated when grown on xylan versus on a mixture of its composing sugars. Some of the most highly upregulated genes encoded hypothetical proteins, which were subsequently characterised and confirmed as novel xylanases with GH5 activity (Dodd et al., 2010).

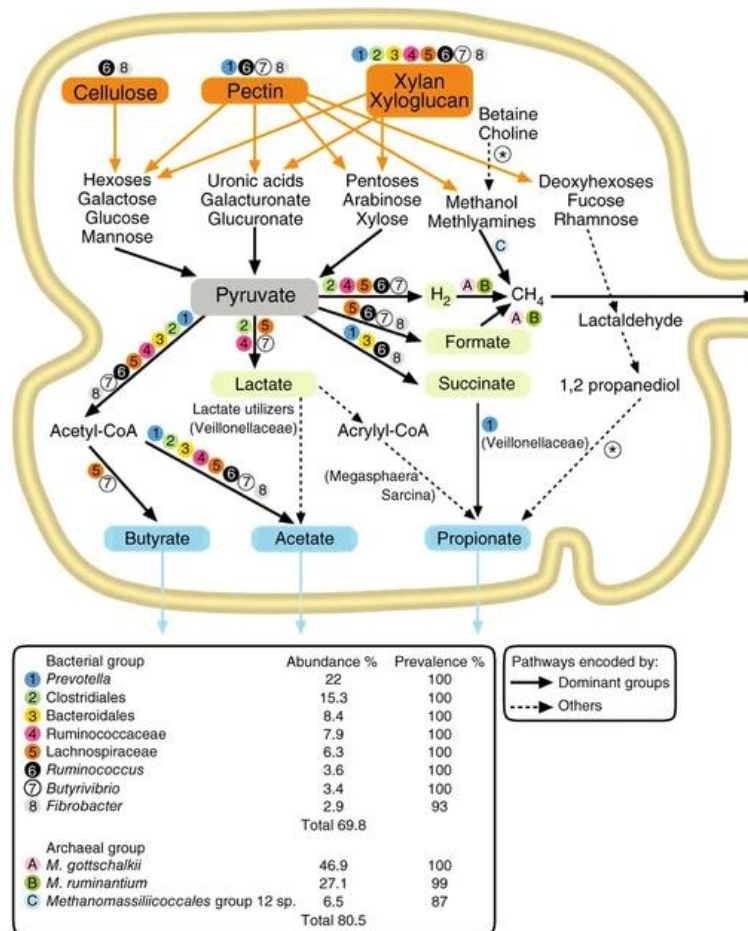
### 1.2.2.2 Rumen fermentation

Following the mechanical and enzymatic breakdown of ingested lignocellulose, the resulting monomeric sugars (hexoses and pentoses) are fermented to form short-chain fatty acids (SCFAs). SCFAs are short carboxylic acid molecules containing between two and six carbons that are absorbed across the ruminal epithelium, providing their hosts with approximately 70% of their energy requirements (Dijkstra, 1994). Concentrations of SCFAs in blood are low, indicating their rapid utilisation by the host upon their absorption from the rumen (Barcroft et al., 1944).

Ruminal fermentation also generates hydrogen as a by-product, the partial pressure of which varies with the relative concentration of different SCFAs produced. This, in turn, influences the amount of methane produced and emitted by the animal (Janssen, 2010). This section discusses the physiological roles of different SCFAs and the factors that influence their production, as well as the importance of methane, an environmentally important by-product of rumen fermentation, due to its much-increased potency as a greenhouse gas when compared to CO<sub>2</sub>.

#### 1.2.2.2.1 SCFA production

The three major SCFAs produced in the rumen are acetate, propionate and butyrate. The first step in their production is the oxidation of monomeric hexoses and pentoses to pyruvate, which represents the central intermediary branch point from which the metabolic pathways forming different major SCFAs diverge (Russell & Wallace, 1997). Acetate is the most abundant SCFA in the rumen, typically accounting for approximately 60% of the total SCFAs produced (Bergman, 1990). Acetate is produced by all major bacterial groups in the rumen (Figure 1.3) using five identified potential pathways *via* acetyl-CoA, or acetyl phosphate (Zhang et al., 2021). Some holotrich protozoa are also capable of converting pyruvate to acetate and hydrogen using organelles called hydrogenosomes, which function in a similar manner to mitochondria (Yarlett et al., 1983). Ruminal infusions of acetate have been associated with increases in milk fat yield (Maxin et al., 2011; Urrutia & Harvatine, 2017).

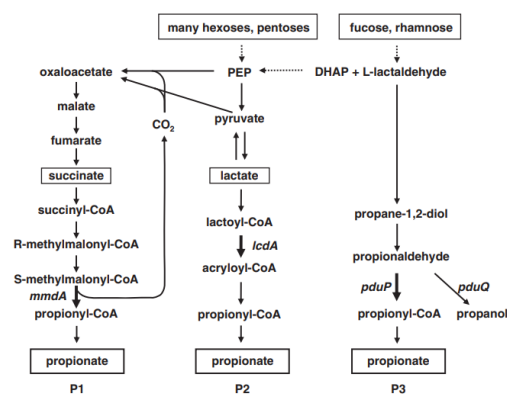


**Figure 1.3. Contributions of different prokaryotic genera to polysaccharide degradation and fermentation in the rumen.** Simplified illustration showing the degradation and metabolism of plant structural carbohydrates by the dominant bacterial and archaeal groups identified in the Global Rumen Census project using information from metabolic studies and analysis of the reference genomes. The abundance and prevalence data shown in the table are taken from the Global Rumen Census project. Abundance represents the mean relative abundance (%) for that genus-level group in samples that contain that group, while prevalence represents the prevalence of that genus-level group in all samples (n = 684). \* The conversion of choline to trimethylamine, and propanediol to propionate generate toxic intermediates that are contained within bacterial microcompartments (BMC). Figure and legend adapted from Seshadri et al. (2018).

Propionate is the second most abundant VFA in the rumen, and holds particular importance due to it being the major precursor of glucose in ruminants *via* gluconeogenesis (Aschenbach et al., 2010). Propionate is formed from sugars *via* three different pathways- the succinate, acrylate and propanediol pathways (Reichardt et al., 2014) (Figure 1.4). The succinate pathway involves either the carboxylation of pyruvate or phosphoenolpyruvate to form oxaloacetate before being reduced to malate by malate dehydrogenase, or alternatively through the reductive carboxylation of pyruvate directly to malate (Paynter & Elsdon, 1970). Malate is then converted to fumarate by fumarate hydratase, which is then reduced by



fumarate reductase to form succinate. Succinyl-CoA is formed by CoA transferase activity, with this reaction coupled with the formation of acetate/butyrate/propionate from their respective CoA thioesters. Succinyl-CoA is converted to *R*-methylmalonyl-CoA by methylmalonyl-CoA mutase, which uses cobalamin (vitamin B<sub>12</sub>) as a cofactor. *R*-methylmalonyl-CoA is converted to *S*-methylmalonyl-CoA by methylmalonyl-CoA epimerase (McCarthy et al., 2001) and then propionyl-CoA is formed by a sodium ion-transporting methylmalonyl-CoA decarboxylase (Hilpert & Dimroth, 1983). Propionyl-CoA is then finally reduced to propionate either by the above CoA-transferase reaction coupled with succinyl-CoA formation, or *via* a two-step phosphate propionyltransferase/propionate kinase reaction in which propionyl phosphate is formed as an intermediate (Louis & Flint, 2017). The acrylate pathway of propionate production involves the reduction of lactate *via* lactoyl-CoA and acryloyl-CoA. The propanediol pathway forms propionate firstly *via* the production of 1-2-propanediol from deoxy-sugars such as fucose and rhamnose, or from dihydroxyacetonephosphate or lactate (Saxena et al., 2010), followed by its further conversion to propionate by other bacteria (Bobik et al., 1999; Reichardt et al., 2014). Given that the production of propionate acts as an alternative hydrogen sink to methane production, propionate levels are often inversely related to the methane yield of the animal (Janssen, 2010; van Nevel et al., 1974). Elevated proportions of propionate are associated with higher milk yields (Miettinen & Huhtanen, 1996), and also correlate with host feed efficiency (Shabat et al., 2016).



**Figure 1.4. Known pathways for propionate formation in gut bacteria.** (P1), Succinate pathway; (P2), acrylate pathway; (P3), propanediol pathway. Substrates utilised are shown in boxes. Genes targeted as molecular markers for the specific pathways are indicated. DHAP, dihydroxyacetonephosphate; PEP, phosphoenolpyruvate. Figure and legend adapted from Reichardt et al (2014) with permission obtained through RightsLink.

Butyrate typically accounts for 10% of total SCFAs produced (Bergman, 1990) and is a major energy source for rumen epithelial cells. Butyrate is produced *via* butyryl-CoA from acetyl CoA, followed by the conversion of butyryl-CoA to butyrate by either phosphate butyryl transferase and butyrate kinase (Miller & Jenesel, 1979), or *via* a butyrate:acetate CoA-transferase, in which its production is coupled with acetyl-CoA production (Trachsel et al., 2016). Butyrate production in the rumen is mainly attributed to unclassified *Lachnospiraceae* and the genus *Butyrivibrio* (Seshadri et al., 2018) although some protozoa (Yarlett et al., 1985) are also capable of producing butyrate from pyruvate. Increased butyrate levels have been shown to negatively affect milk and lactose yields, and also increase milk fat content (Miettinen & Huhtanen, 1996). Butyrate production has also been shown to correlate with feed efficiency (Guan et al., 2008).

Further to the abovementioned characterised pathways for the production of SCFAs, it should also be noted that genome analyses of 48 strains of rumen bacteria carried out in Hackmann et al. (2017) found that 44% lacked genes for characterised pathways to explain their production of various fermentation end products, suggestive of the existence of numerous novel uncharacterised metabolic pathways. For example, rather than succinate formation occurring *via* a currently characterised pathway, the genomes of *Prevotella* and *Selenomonas* strains instead appear to encode an atypical pathway involving an Rnf complex (Kuhns et al., 2020) using an electron bifurcation mechanism (Müller et al., 2018) that oxidises reduced ferredoxin and reduces NAD, coupled with ion transport across the cytoplasm used to generate ATP (Buckel & Thauer, 2018). This finding emphasises the importance of culture-based studies to better characterise these atypical pathways of fermentation end product formation in rumen bacteria (see Chapter 4).

In addition to the three major SCFAs, a number of longer branched chain fatty acids are also produced through the fermentation of amino acids (Hungate, 1966) collectively referred to as isoacids (Andries et al., 1987). Despite being present in comparatively low concentrations, isoacids are specific nutrients for some cellulolytic bacteria (Bryant, 1973) and can also influence microbial fermentation (Andries et al., 1987).

The relative proportions of major SCFAs produced in the rumen is largely determined by diet (Bergman, 1990). For example, in animals fed forage-based diets the ratio of acetate:propionate:butyrate is typically 7:2:1, whereas in animals fed a high grain diet propionate becomes increasingly abundant and the ratio approaches 5:4:1 (Siciliano-Jones & Murphy, 1989). SCFA profiles can also be altered through the use of a range of dietary additives, as discussed in more detail in Section 1.2.3.2.

#### 1.2.2.2.2 Methane production

Hydrogen is produced as an end product of ruminal fermentation, and upon accumulation, it can inhibit fermentative processes (Janssen, 2010). Rumen methanogens use hydrogen as an energy source to produce methane, which is expelled from the rumen *via* eructation. Methane is an especially potent greenhouse gas, with a global warming potential approximately 34 times greater than carbon dioxide (Parry et al., 2007). It is estimated that approximately 12% of global methane emissions are attributable to livestock (Reisinger et al., 2021). Methanogenesis involves the uptake of hydrogen and the stepwise reduction of carbon dioxide, and most of the hydrogen produced as a by-product of fermentation ends up as methane (Russell & Wallace, 1997). Methane can be produced *via* three different pathways: the hydrogenotrophic, methylotrophic and acetoclastic pathways, which use CO<sub>2</sub>, methyl groups and acetate as substrates, respectively. Methane is primarily produced by the hydrogenotrophic pathway in the rumen, although some is also produced *via* the methylotrophic pathway, and an even lesser extent *via* the acetoclastic pathway (Morgavi et al., 2010). Methanogenesis keeps the partial pressure of hydrogen low in the rumen such that hydrogen formation is favoured. When hydrogen accumulates in the rumen as a result of methanogen inhibition, propionate formation is favoured (Janssen, 2010).

Given its high importance as a greenhouse gas, a large body of research has sought to better understand ruminal methanogenesis and factors that influence it (Beauchemin et al., 2020). One such area has been investigating the mechanisms underlying differences between low

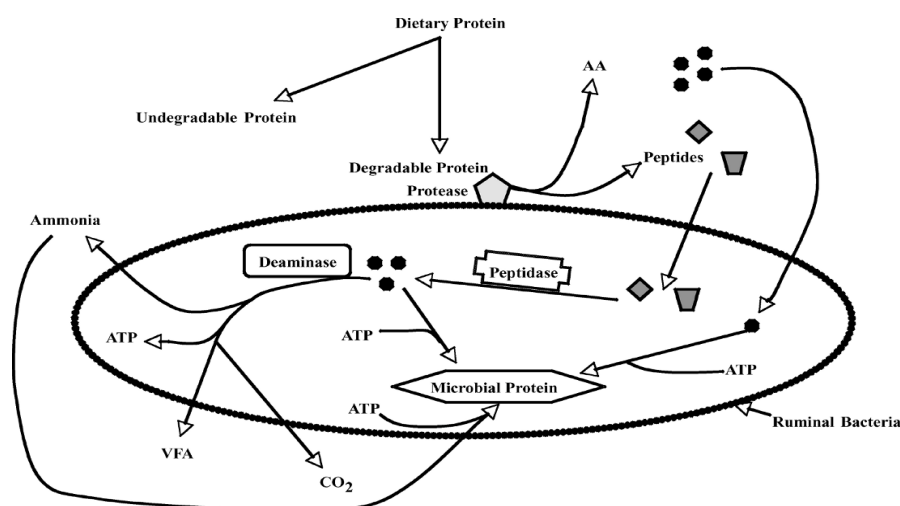
and high methane-emitting animals. Sheep that have lower yields of methane were found to be associated with two different rumen community types, distinct from the bacterial community found in the rumen of high methane-emitting sheep (Kittelmann et al., 2014). Significant insight into ruminal methanogenesis was gained from Shi et al. (2014), which compared deeply sequenced metagenome and metatranscriptome datasets from rumen contents of animals differing in their methane yield phenotypes (4 high, 4 low, and 2 intermediate methane-emitters). Comparison of low and high methane-emitting animals showed no differences in rumen community composition, nor in the abundances of methanogenesis pathway genes in the metagenome. However, metatranscriptome analyses found that genes encoding the methanogenesis pathway were significantly upregulated in high methane-emitting sheep, thereby demonstrating that their difference in methane yield phenotype are due to increased metabolic activity of the rumen methanogen community rather than any change in community composition or abundance (Shi et al., 2014). Further analyses of the metagenomic and metatranscriptomic datasets generated in this study showed that abundances of the lactate-producing and utilizing bacterium *Sharpea azabuensis* and genes involved in the acrylate pathway of propionate production were enriched in low versus high methane-emitting animals (Kamke et al., 2016). It has also been shown that methane yield is positively correlated with rumen volume (Goopy et al., 2014), leading to the hypothesis that smaller rumens with a higher turnover rate of digesta result in less hydrogen and therefore less methane produced (Kamke et al., 2016). The composition of the rumen eukaryotic community appears unrelated to methane yield (Kittelmann et al., 2014).

### 1.2.2.3 Protein metabolism

Dietary protein is an important component of the ruminant diet as it ultimately provides amino acids to the host required for the synthesis of new proteins, *via* the production of rumen microbial protein. Forage diets contain high levels of protein compared to concentrate-based diets, and the excretion of excess nitrogen from forage-fed ruminants in the form of urea in urine has significant impacts on the environment. Analogous to the digestion of carbohydrates, the first step in microbial protein degradation is the physical attachment of proteolytic microbes, followed by the action of cell-bound microbial proteases

(Bach et al., 2005). The synergistic actions of a variety of different proteases is required for complete protein degradation. Once degraded by extracellular proteases, the resulting peptides and amino acids are assimilated by microbes (Figure 1.5). Peptides are further degraded into amino acids by peptidases which are deaminated and incorporated into microbial protein or alternatively fermented into SCFAs. Microbial protein constitutes the main protein supply to the host, accounting for approximately 90% of the total amino acids reaching the small intestine (Russell & Rychlik, 2001). In the rumen, ammonia is both the major product of catabolism as well as the main substrate used by the rumen microbiota to synthesise new protein (Wallace et al., 1997).

The rate of ruminal protein degradation is dependent on a range of factors, including conditions and microbial community composition in the rumen as well as the compositional nature of the protein being degraded. Some peptide bonds between different amino acids are more resistant to ruminal degradation than others, and as such different proteins can have vastly different rates of degradation (Yang & Russell, 1992). Rates of protein degradation are inversely associated with passage rate (Ørskov & McDonald, 1979), and is also reduced at lower rumen pH levels (Calsamiglia et al., 2002). Moreover, plant components such as condensed tannins can bind to protein in the rumen, reducing protein solubility and degradation by microbiota (McMahon et al., 2000).



**Figure 1.5. Schematic representation of protein degradation and fate of end products in the rumen.** Figure adopted from Bach et al (2005) with permission obtained through RightsLink.

### 1.2.3 Rumen microbiome manipulation targets

Given the critical role that the rumen microbiota has in the development and physiology of its host, manipulating the rumen microbiome has long been an attractive prospect to improve production efficiency and environmental sustainability (Chalupa, 1977). Improving the digestibility of ingested forage could provide ruminants with greater nutrition and energy yields, allowing more SCFAs to be produced per unit of ingested feed. Additionally, manipulating fermentation profiles to favour higher concentrations of propionate compared to acetate and butyrate could increase both overall yield and nutrient content of milk while simultaneously reducing methane emissions. This section discusses strategies aimed at optimising these processes in further detail.

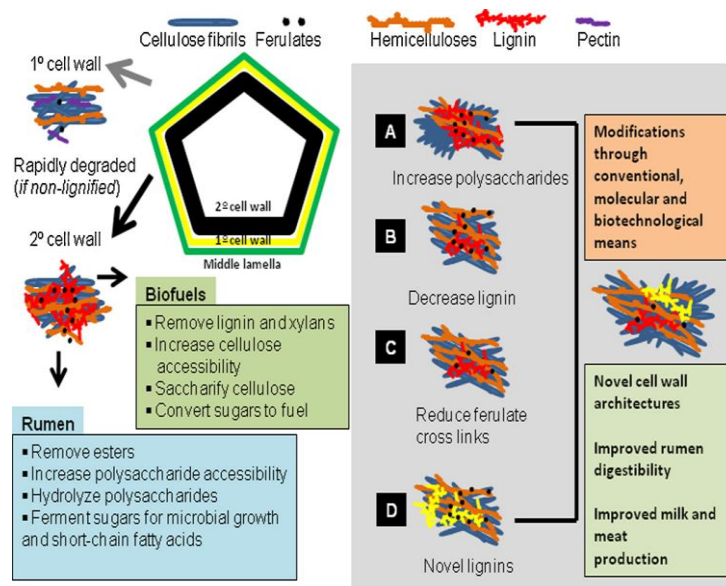
#### 1.2.3.1 Enhancing fibre degradation

The finding that fibre recovered from faeces is fermentable suggests that fibre degradation in the rumen is not complete, and indicates that strategies could potentially be developed to improve its efficiency (Krause et al., 2003). This has led to numerous efforts to optimise ruminal fibre degradation, such as inoculating the rumen with strains exhibiting enhanced fibrolytic activity or by chemically, enzymatically, mechanically, or genetically modifying plant feed to enhance its degradability.

A wide range of studies have assessed the effectiveness of microbial inoculants, both genetically modified and natural strains, that elicit enhanced fibre digesting capabilities or that are more resistant to certain environmental conditions, thus making them more competitive (Krause et al., 2003). Unfortunately, in many of these cases these introduced strains seem unable to successfully persist in the rumen microbiome (Weimer, 2015). One particularly extreme example of this was seen in a study where 6 L of culture of the fibrolytic bacterium *Clostridium longisporum* was inoculated into emptied bovine rumens, along with 20 L of buffer. Despite being inoculated in such an enormous quantity, after 48 hours the strain was undetectable (Varel et al., 1995). Similarly, repeated ruminal dosing with fibre

degrading *Ruminococcus* spp. led to a rapid decline in the concentrations of the inoculated strains post-inoculation, and no observable improvement in fibre degradation (Krause et al., 2001). Furthermore, in another study, stimulating a ten-fold increase in the concentration of cellulolytic bacteria by feeding a high-cellulose diet did not coincide with any significant increase in the rate of fibre degradation (Dehority & Tirabasso, 1998). Given their prominent role in fibre degradation, there is also considerable interest in using rumen fungi as inoculants (Hartinger & Zebeli, 2021).

Given that limitations to polysaccharide degradation in the rumen are typically regarded as being the result of the nature of the substrate or its processing rather than a lack of microbial fibrolytic activity *per se* (Weimer, 1996), it has been argued (Weimer, 1998) that a better strategy of enhancing the efficiency of fibre digestion may be to modify crops to render their cell walls more readily digestible. Lignin acts as a shield to the digestion of lignocellulose (Moore & Jung, 2001), and as a consequence much of the cellulose present in ingested feed is inaccessible to the fibrolytic microbial community. As such, reducing and/or altering the structure of lignin in forages has been widely studied as a possible means of enhancing ruminal fibre degradation. This has been achieved in a range of forages by genetic manipulation or suppression of genes involved in lignin biosynthesis. For example, antisense gene suppression of an enzyme involved in lignin biosynthesis in the tropical forage legume *Stylosanthes humilis* led to a 10% increase in digestibility *in vitro* (Rae et al., 2001). Lignin-deficient corn plants containing brown midrib-3 mutations also improved digestion of beef cattle, although this did not translate into improved animal performance (Tjardes et al., 2000). It should also be noted that given the important role lignin plays in providing rigidity to the plant as well as preventing invasion by fungal pathogens (Miedes et al., 2014), strong reductions in lignin content can also be deleterious to plant health (Jung et al., 2012).



**Figure 1.6. Plant-directed strategies of promoting ruminal fibre degradation.** Rumen digestion and biofuel conversion processes are summarized, and four possible modifications to cell wall structure that may improve digestibility and/or conversion processes are illustrated. Figure adopted from Jung et al. (2012) with permission obtained through RightsLink.

### 1.2.3.2 Promoting propionate production

Given that glucose is rapidly fermented to SCFAs in the rumen, ruminants synthesise the glucose they require for their metabolism through gluconeogenesis. Propionate is the major glucogenic precursor in ruminants, contributing 60-74% of the total carbon used for gluconeogenesis (Aschenbach et al., 2010). The ratio of propionate to the other two major SCFAs (acetate and butyrate) therefore plays an important role in the nutrient composition and the overall rate of milk production (Miettinen & Huhtanen, 1996). Additionally, as propionate is an alternate electron sink to methane production, enhancing propionate production typically coincides with a reduction in methane production (Janssen, 2010). As such, strategies to enhance the proportion of ruminal propionate produced are especially desirable.

Inoculation of the rumen with propionate producing bacteria has been assessed in a number of studies, with varying levels of success. Ruminal inoculation with *Propionibacterium* P169 promoted propionate production in beef steers (Lehloenya et al., 2008), as well as dairy cows (Stein et al., 2006). However, another study found that in beef heifers on a high-forage diet,



daily administration of three different *Propionibacterium* strains had no effect on molar proportions of SCFAs (Vyas et al., 2014).

Another strategy that has been used to increase the ratio of propionate relative to other SCFAs is through the use of selective antimicrobial compounds. The most extensively studied of these is monensin, an antimicrobial ionophore naturally produced by the bacterium *Streptomyces cinnamonensis*, which acts against a range of Gram positive microorganisms (Duffield et al., 2012). Monensin feeding in cattle was shown to increase the molar proportion of propionate by approximately 10%, while simultaneously decreasing molar proportions of acetate and butyrate continuously over a 148-day experiment (Richardson et al., 1976). Similarly, daily monensin feeding of steers increased propionate production of steers fed both forage and concentrate-based diets, although increases in glucose kinetics were minor in comparison to increases in propionate production (Van Maanen et al., 1978). Despite evidence suggesting that monensin has a low potential to facilitate the development of monensin-resistant strains (Russell & Houlihan, 2003) its use has been banned in the European Union since 2006 (Hao et al., 2014), although is still commercially available in the United States, Australia and New Zealand, where it is often used to prevent bloat (Marques & Cooke, 2021).

A range of dietary additives have been shown to provide many of the same benefits conferred by monensin feeding. Plant-derived compounds such as garlic oil, cinnamaldehyde, eugenol, capsaicin and anethole have been shown to mirror the effects of monensin by enhancing propionate production, while simultaneously inhibiting methane production and rumen proteolysis. However, these effects are generally highly dependent on diet and ruminal pH (Calsamiglia et al., 2007). As the inclusion of concentrates in the diet of animals increases, so too does the ratio of propionate produced (Hobson et al., 1997). However, in agricultural systems in New Zealand, animals are typically grazed on pastures and therefore predominantly consume forage outdoors, year-round (Waghorn & Clark, 2004).

### 1.2.4 Elucidation of the uncharacterised rumen microbial majority through next-generation omics'

Rumen microbiological research was greatly facilitated by the development of anaerobic cultivation techniques by Robert E. Hungate, considered to be the father of rumen microbiology (Huws et al., 2018). The pioneering techniques he developed in the 1940s allowed the isolation of pure cultures of rumen microorganisms *in vitro*, enabling their detailed characterisation (Hungate, 1944). For the remainder of the 20<sup>th</sup> century, an extensive body of work greatly improved our understanding of the physiology of some early cultured members of the rumen microbiota, including strains involved in key rumen metabolic processes such as fibre degradation and ammonia production (Hobson & Stewart, 1997). However, following the development and continual increase in throughput of cultivation-independent methods of community analysis, it became increasingly clear that traditionally isolated and well-characterised rumen isolates were representatives of only a small fraction of the total rumen microbial community (Creevey et al., 2014; Edwards et al., 2004; Krause et al., 2013; Ramšak et al., 2000). This finding was later emphasised by the Global Rumen Census study, as 4 of the 7 most abundant rumen bacterial groups that comprised the core rumen microbiota were unclassified beyond the family level (Henderson et al., 2015). This increasing awareness of the lack of cultured representatives of some of the dominant groups of rumen bacteria spurred intensive cultivation efforts, which resulted in the isolation of many representatives of abundant but unstudied rumen bacterial groups (Kenters et al., 2011; Koike et al., 2010; Noel, 2013; Nyonyo et al., 2013).

The increasing number of meta'omic datasets obtained from rumen samples, accelerated by the application of next generation sequencing technologies, has revolutionised our understanding of the structure and functions of the rumen microbiota (McCann et al., 2014). However, genome sequence information from the increasing number of rumen strains available in culture is crucial for improving the accuracy of functional/taxonomic annotation, and thus our interpretation, of rumen meta'omic datasets (Creevey et al., 2014). The availability of rumen microbial genome sequences was greatly facilitated by the initiation of the Hungate1000 project in 2012, which was a collaborative effort involving rumen

microbiology labs around the world led by the Rumen Microbiology team at AgResearch Grasslands (Palmerston North, New Zealand) to create a catalogue of genomes selected from available cultures of rumen microorganisms. This collection currently includes a total of 501 rumen bacteria and archaea, including members of some currently uncharacterised but abundant rumen bacterial groups (Seshadri et al., 2018). The availability of these genome sequences allows for meta'omic datasets to be mined for specific functions, providing insights into the particular taxonomic groups involved in metabolic processes of interest (Greening et al., 2019). While the rapidly increasing number of metagenome-assembled genomes (MAGs) (Sangwan et al., 2016) being reconstructed from rumen samples can also be used to create such databases (Stewart et al., 2019; Xie et al., 2021), culture collections such as the Hungate1000 possess the powerful advantage that organisms of interest, which may be highlighted by 'omics studies, are available for further *in vitro* characterisation and application.

### 1.3 Project objectives

The extensive rumen microbial metagenomic and metatranscriptomic datasets, coupled with the rapidly increasing number of rumen microbial genomes now available (Seshadri et al., 2018) provides the opportunity to better understand the contribution of specific microbial groups of the rumen microbiota to ruminal metabolism (Li et al., 2017). In this thesis, we set out to leverage these new data and culture resources to further characterise rumen bacterial groups and metabolic processes that have prominent, but poorly understood, roles in the ruminal metabolism of ruminants fed forage-based diets, characteristic of New Zealand agricultural systems. The first group of bacteria, characterised in Chapter 3, belonging to an unclassified group of the order *Clostridiales* referred to the R-7 group, was selected due to being one of the most abundant groups of rumen bacteria, particularly in forage-fed animals (Henderson et al., 2015). We wanted to understand the diversity and functional contributions of the R-7 group to ruminal metabolism *via* genome analyses and extensive *in vitro* characterisation of representatives available in the Hungate1000 collection.

In addition, to better understand the organisms and the pathways they utilise for the production of propionate from succinate, in Chapter 4 we sought to use ruminal metatranscriptome datasets from animals fed a forage diet to identify taxa with dominant roles in this process (Section 4.2). Based on these results, a group of strains from the Hungate1000 collection were selected to characterise their propionate production, and investigate its regulation using a comparative transcriptome and proteome approach (Section 4.3).

## 2 Materials and Methods

## 2.1 Materials

### 2.1.1 Biological material

All strains included in this study were sourced from the Hungate1000 Culture Collection (Palmerston North, New Zealand) (Seshadri et al., 2018), aside from the R-7 group strains FE2010, FE2011 and XBB3002, which were revived from glycerol stocks described in Noel (2013).

### 2.1.2 Buffers and solutions

All autoclaved solutions were autoclaved by heating to 121°C for 20 min and were stored at room temperature, unless otherwise indicated. Where necessary, pH adjustments were made by addition of concentrated HCl or NaOH.

#### **Tris-acetate-EDTA (TAE) buffer**

A 50× solution containing glacial acetic acid (950 mM, VWR International Ltd, Poole, UK), 50 mM EDTA (pH 8.0) and 2 M Trizma base (Thermo Fisher Scientific Inc., Waltham, MA, USA) was prepared in dH<sub>2</sub>O and autoclaved. A working solution (1×) was made by diluting the stock solution 50-fold in dH<sub>2</sub>O.

#### **Tris-EDTA (TE) buffer**

TE buffer was made to final concentrations of 10 mM Trizma base and 1 mM EDTA (pH 8.0), and autoclaved.

#### **Saline-EDTA solution**

Saline-EDTA solution was made to final concentrations of 10 mM EDTA and 150 mM NaCl in dH<sub>2</sub>O, the pH adjusted to 8.0, and autoclaved.

### **Lysis buffer**

A mixture of 1 mg/mL lysozyme (Thermo Fisher Scientific) and 20 µg/mL RNase A (Invitrogen) was prepared in saline-EDTA solution.

### **Diethylpyrocarbonate (DEPC)- treated H<sub>2</sub>O**

DEPC (0.1% (v/v), Sigma-Aldrich, St. Louis, MO, USA) was added to dH<sub>2</sub>O, then autoclaved.

### **RNA extraction buffer**

RNA extraction buffer was prepared to contain final concentrations of 200 mM NaCl and 20 mM EDTA, prepared in RNase-free H<sub>2</sub>O (Invitrogen).

## **2.1.3 Media and media additives**

### **2.1.3.1 Media additives**

#### **Anaerobic glycerol solution**

Anaerobic glycerol solution (40%; v/v) was prepared to 500 mL with the following: 130 mL H<sub>2</sub>O, 85 mL salt solution A, 85 mL salt solution 2B, 200 mL glycerol, 2 drops of resazurin (0.1% w/v), 2.5 g NaHCO<sub>3</sub> and 0.125 g L-cysteine-HCl. The solution was bubbled with N<sub>2</sub> in serum bottles, sealed with a rubber bung, crimp-sealed and autoclaved at 121°C for 20 min. The autoclaved solution was stored at room temperature in the dark until use.

#### **Rumen fluid collection for preparation of media**

Whole rumen contents were collected by trained AgResearch staff under a general animal ethics approval (AgResearch Animal Ethics Committee AE12174) allowing the collection of rumen contents from hay-fed fistulated cattle fasted for approximately 16 hours prior to sampling. Removed rumen contents were filtered through cheesecloth, and the resulting rumen liquor was centrifuged at 10,000 × *g* for 20 min, and the supernatant stored at -20°C. Before addition in media, frozen rumen fluid was thawed and centrifuged again at 10,000 ×

g, and the resulting supernatant constituted 'centrifuged rumen fluid' used in media preparation.

### **Salt Solution A**

Salt solution A was prepared by dissolving 6 g NaCl, 1.5 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 g KH<sub>2</sub>PO<sub>4</sub>, 0.79 g CaCl<sub>2</sub>·2H<sub>2</sub>O, and 1.2 g of MgSO<sub>4</sub>·7H<sub>2</sub>O per litre of dH<sub>2</sub>O.

### **Salt Solution 2B**

Salt solution 2B was prepared by dissolving 6 g K<sub>2</sub>HPO<sub>4</sub> per litre of dH<sub>2</sub>O.

### **Selenite-Tungstate solution**

A solution containing 12.5 mM NaOH, 0.01 mM Na<sub>2</sub>SeO<sub>3</sub>·5H<sub>2</sub>O and 0.01 mM Na<sub>2</sub>WO<sub>4</sub>·2H<sub>2</sub>O per litre of dH<sub>2</sub>O was prepared, and autoclaved at 121°C for 20 min.

### **Substrate preparations**

Substrates were prepared by dissolving or suspending each substrate in dH<sub>2</sub>O to 10% (w/v) in serum bottles. Solutions were then bubbled with N<sub>2</sub> for 15 min, sealed, autoclaved and stored in the dark. All substrate solutions were stored at room temperature except for esculin, starch and xylan, which were instead stored at 39°C due to their tendency to solidify at room temperature.

### **Hemin solution**

A 0.5 mg/mL stock solution was made by dissolving 100 mg of hemin (Thermo Fisher Scientific) in 2 mL of 5 M NaOH, and then adding 198 mL of dH<sub>2</sub>O. The solution was autoclaved and stored at 4°C.



**Table 2.1. Substrates.**

<b>Substrate</b>	<b>Manufacturer</b>
amygdalin	Sigma-Aldrich (St.Louis, MO, USA)
L-arabinose	Sigma-Aldrich (St.Louis, MO, USA)
D-cellobiose	Sigma-Aldrich (St.Louis, MO, USA)
crystalline cellulose	Sigma-Aldrich (St.Louis, MO, USA)
esculin	Sigma-Aldrich (St.Louis, MO, USA)
fructose	Sigma-Aldrich (St.Louis, MO, USA)
D-galactose	Sigma-Aldrich (St.Louis, MO, USA)
D-glucose	Thermo Fisher Scientific (Waltham, MA, USA)
glycerol	Thermo Fisher Scientific (Waltham, MA, USA)
glycogen	Sigma-Aldrich (St.Louis, MO, USA)
inulin	Sigma-Aldrich (St.Louis, MO, USA)
lactose	Thermo Fisher Scientific (Waltham, MA, USA)
maltose	Sigma-Aldrich (St.Louis, MO, USA)
D-mannitol	Sigma-Aldrich (St.Louis, MO, USA)
D-mannose	Sigma-Aldrich (St.Louis, MO, USA)
melezitose	Sigma-Aldrich (St.Louis, MO, USA)
D-melibiose	Sigma-Aldrich (St.Louis, MO, USA)
myo-inositol	Sigma-Aldrich (St.Louis, MO, USA)
pectin	Sigma-Aldrich (St.Louis, MO, USA)
D-raffinose	Sigma-Aldrich (St.Louis, MO, USA)
L-rhamnose	Sigma-Aldrich (St.Louis, MO, USA)
D-ribose	Thermo Fisher Scientific (Waltham, MA, USA)
rutin	Sigma-Aldrich (St.Louis, MO, USA)
salicin	Sigma-Aldrich (St.Louis, MO, USA)
sorbitol	Sigma-Aldrich (St.Louis, MO, USA)
starch	BDH Chemicals Ltd. (Poole, England)
sucrose	Scharlau (Barcelona, Spain)
D-trehalose	Sigma-Aldrich (St.Louis, MO, USA)
xylan	Sigma-Aldrich (St.Louis, MO, USA)
D-xylitol	ICN Biomedicals (Aurora, OH, USA)
D-xylose	Sigma-Aldrich (St.Louis, MO, USA)

**Trace element solution (SL-10)**

Trace element solution (SL-10) (Widdel & Pfennig, 1981) used in BY medium, was made to 1 L with the following: 10 mL of 25% HCl, 1.5 g FeCl<sub>2</sub>·4H<sub>2</sub>O, 190 mg CoCl<sub>2</sub>·4H<sub>2</sub>O, 100 mg MnCl<sub>2</sub>·4H<sub>2</sub>O, 70 mg ZnCl<sub>2</sub>, 6 mg H<sub>3</sub>BO<sub>3</sub>, 36 mg Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O, 24 mg NiCl<sub>2</sub>·6H<sub>2</sub>O, and 2 mg CuCl<sub>2</sub>·2H<sub>2</sub>O. The solution was autoclaved and stored at room temperature.

### **Pfennig's mineral solution**

The following components were dissolved in 1 L of H<sub>2</sub>O, as described in Schaefer et al. (1980): 500 mg Na<sub>4</sub>EDTA, 200 mg of FeSO<sub>4</sub>·7H<sub>2</sub>O, 200 mg of MnCl<sub>2</sub>·4H<sub>2</sub>O, 10 mg of ZnSO<sub>4</sub>·7H<sub>2</sub>O, 30 mg of H<sub>3</sub>BO<sub>3</sub>, 20 mg of CoCl<sub>2</sub>·6H<sub>2</sub>O, 1 mg of CuCl<sub>2</sub>·2H<sub>2</sub>O, 2 mg of NiCl<sub>2</sub>·6H<sub>2</sub>O, and 3 mg NaMoO<sub>4</sub>·2H<sub>2</sub>O. The solution was autoclaved and stored at room temperature.

### **Vitamin 10 concentrate solution**

Vitamin 10 concentrate solution was made to give the following final concentrations in the growth medium, as described in (Janssen et al., 1997): 40 µg/L 4-aminobenzoate, 10 µg/L D-(+)-biotin, 100 µg/L each of nicotinic acid and thiamine chloride hydrochloride, 50 µg/L each of hemicalcium-D-pantothenate and cyanocobalamin, 30 µg/L each of D, L-6, 8- thioctic acid and riboflavin, and 10 µg/L folic acid. Once components had dissolved, O<sub>2</sub>-free N<sub>2</sub> was bubbled through the solution for a further 20 min before the solution was filter sterilised (by passing through a 0.22 µm pore size filter) into sterile N<sub>2</sub>-filled serum bottles. Working stocks were prepared by diluting 10-fold in N<sub>2</sub>-gassed H<sub>2</sub>O, and working stocks were diluted a further 100-fold when added to culture media. The bottles were wrapped in aluminium foil to protect against light and frozen at -20°C, or kept at 4°C as working stocks.

### **Vitamin solutions (for Strobel medium)**

Vitamin concentrate solutions for use in Strobel medium were made to give the following final concentrations in the growth medium, as described in Cotta and Russell (1982): 2 mg/L each of pyridoxamine dihydrochloride, riboflavin, thiamine hydrochloride, nicotinamide and calcium pantothenate, 1 mg/L of Lipoic acid, 0.1 mg/L para-aminobenzoic acid, and 50 µg/L each of folic acid, biotin and vitamin B<sub>12</sub> (cyanocobalamin). Concentrate solutions were prepared as 100× stocks, with and without vitamin B<sub>12</sub>. Solutions were gassed with N<sub>2</sub> and filter-sterilised through a 0.22 µm filter into sterilised N<sub>2</sub>-filled bottles, covered in aluminium foil and stored at -20°C, or kept at 4°C as working stocks.

### 2.1.3.2 Media

In all cases, media were prepared in dH<sub>2</sub>O and were made anaerobic by boiling in a microwave followed by bubbling and dispensing under 100% CO<sub>2</sub>, and dispensed into CO<sub>2</sub>-filled Hungate tubes or serum bottles sealed with a butyl rubber bung held in place with a plastic screw-top (Hungate tubes) or with a metal crimp (serum bottles). L-cysteine-HCl was added to bubbling media once it had cooled. Media were sterilised by autoclaving at 121°C and stored at room temperature in the dark until use.

#### **Basal medium plus yeast extract (BY)**

BY media was prepared as described in Joblin (2005): The added components (per litre) are as follows: 360 mL H<sub>2</sub>O, 300 mL centrifuged rumen fluid, 170 mL each of salt solutions A and 2B, 1 g of yeast extract (Merck, Darmstadt, Germany), 10 drops of 0.1% (w/v) resazurin, 5 g NaHCO<sub>3</sub>, 1 mL each of trace element solution SL-10 and selenite-tungstate solution, and 500 mg of L-cysteine-HCl.

#### **Defined Strobel medium**

Media used for all *Prevotella* experiments was made based on that described in Strobel (1992), and contained the following per litre: 292 mg of K<sub>2</sub>HPO<sub>4</sub>, 292 mg of KH<sub>2</sub>PO<sub>4</sub>, 480 mg of Na<sub>2</sub>SO<sub>4</sub>, 480 mg of NaCl, 100 mg of MgSO<sub>4</sub>·7H<sub>2</sub>O, 64 mg of CaCl<sub>2</sub>·H<sub>2</sub>O, 530 mg of NH<sub>4</sub>Cl, 600 mg of L-cysteine-HCl, 4 g of Na<sub>2</sub>CO<sub>3</sub>, 1 g of Bacto-peptone; 1 mg of hemin, 1 mM (each) isobutyrate, 2-methylbutyrate, valerate, and isovalerate, 10 mL of Pfennig's micromineral solution. All components were dissolved by stirring prior to boiling aside from the volatile fatty acids and L-cysteine-HCl, which were added while media was bubbled with CO<sub>2</sub> once it had cooled. Vitamin solution was added as a 100× working stock prior to inoculation.

#### **M2GSC medium**

M2GSC medium was prepared as described in Miyazaki et al. (1997), to contain the following per litre: 0.45 g each of K<sub>2</sub>HPO<sub>4</sub> and KH<sub>2</sub>PO<sub>4</sub>, 0.9 g each of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and NaCl, 90 mg each of

MgSO<sub>4</sub>·7H<sub>2</sub>O and CaCl<sub>2</sub>, 10 drops of 0.1% (w/v) resazurin, 10 g of Bacto-peptone, 2.5 g of yeast extract, 2 g each of cellobiose, glucose and soluble starch, 30% (v/v) centrifuged rumen fluid, 4 g of NaHCO<sub>3</sub>, and 1 g of L-cysteine-HCl. Vitamin 10 concentrate solution was added to tubes as a 100× working stock prior to inoculation.

## 2.2 Methods

### 2.2.1 Revival and maintenance of cultures

Culturing of strains in liquid media was carried out anaerobically at gassing stations with lines of 100% CO<sub>2</sub> and N<sub>2</sub> gases (BOC, Palmerston North, New Zealand) aseptically using the Hungate technique (Hungate, 1966), using a sterile CO<sub>2</sub>- flushed needle and syringe to add or remove liquids from Hungate tubes or serum bottles of liquid media, sterilising the butyl stopper with flamed 96% ethanol prior to each puncture. Work with agar plates to confirm culture purity was carried out in an anaerobic glove box chamber (Coy Laboratory Products, MI, USA). Strains used in this study were revived from glycerol stocks by thawing on ice and inoculating approximately 0.3 mL per 9.5 mL of media plus suitable substrate. Isolates were thereafter maintained on a suitable substrate, and sub-cultured every few days as required. For substrate utilisation assays, strains were revived and sub-cultured a maximum of two times prior to their use in experiments to avoid effects due to *in vitro* culture-biased evolution (Papadopoulos et al., 1999).

### 2.2.2 Colony purification

As a means of confirming culture purity, cultures were streaked on agar media to isolate single colonies. Agar plates of BY media containing 1.5% (w/v) agar were prepared and stored in an anaerobic glove box chamber. Cells from liquid cultures were streaked onto plates using a sterile plastic spreader, and plates were stored in a 2.5 L Oxoid AnaeroJar (Thermo Fisher) containing a AnaeroPack sachet (Mitsubishi Gas Chemical) and incubated at 39°C for 3-4 days. Once colonies had grown, cells were taken from single colonies using a sterile plastic spreader

and streaked on another agar plate and re-incubated under the same conditions. Cells from the resulting isolated colonies were re-inoculated into appropriate liquid media using a sterile spreader, incubated, and the identity of resulting cultures confirmed by 16S rRNA sequencing (Section 2.2.7.1).

### 2.2.3 Measurement of fermentation end products

An aliquot of culture to be analysed was transferred to an Eppendorf tube and centrifuged at  $21,000 \times g$  for 10 min, and the supernatant transferred to a new tube. A 10 $\times$  solution of 20% (v/v) orthophosphoric acid containing the internal standard 2-ethylbutyric acid was added to the culture supernatant to working concentration (1 $\times$ ), and frozen overnight. Samples were thawed and centrifuged at  $21,000 \times g$  for 10 min to remove any further precipitated proteins, and the resulting supernatant transferred to a GC vial and crimp sealed. Analysis of aqueous SCFAs (acetate, propionate, isobutyrate, butyrate, isovalerate, valerate, caproate) and alcohols (methanol, ethanol, 1-propanol) was performed by Peter Reid (AgResearch Grasslands) using a Shimadzu GC-2010 Plus gas chromatograph (GC) and AOC 20i auto injector (Shimadzu Corporation, Kyoto, Japan), a Phenomenex Zebron ZB-FFAP Capillary GC Column; 30 m length x 0.53 mm I.D x 1.00  $\mu$ m film thickness (Phenomenex, Torrance, CA, USA), and helium carrier gas (BOC, Palmerston North, NZ). The GC conditions were as follows; 1  $\mu$ L of sample injected direct (on column), injection temperature set to 90 $^{\circ}$ C, flame ionization detector temperature set to 240 $^{\circ}$ C, GC column temperature set to 60 $^{\circ}$ C for 3.5 min then to 120 $^{\circ}$ C (30 $^{\circ}$ C/min), then to 185 $^{\circ}$ C (10 $^{\circ}$ C/min), then to 200 $^{\circ}$ C (15 $^{\circ}$ C/min), then hold for 3 min.

To measure lactate, succinate and formate concentrations, the samples used for aqueous SCFA analysis were derivatised using a downscaled method based on that described in Richardson et al. (1989). Briefly, to 200  $\mu$ L of sample, 100  $\mu$ L of 33% HCl (Sigma Aldrich, St. Louis, MO, USA), 5  $\mu$ L of 1% resazurin dye (Thermo Fisher) and 800  $\mu$ L diethyl-ether (Thermo Fisher) was added to the samples, which were then mixed vigorously, and the top ether layer collected into a new 2 mL Eppendorf tube. The diethyl-ether extraction was repeated to the sample, and the ether layer collected into the same 2 mL tube. To a GC vial, 100  $\mu$ L of

derivatisation agent; N-methyl-N-t-butyltrimethylsilyl trifluoroacetamide (Sigma Aldrich) was added and 800  $\mu\text{L}$  of ether extraction was added and the vial crimped closed immediately. Samples were heated at 80°C for 20 min and left in a fume hood for 48 h to allow lactic acid to fully derivatise.

Hydrogen concentrations in the culture headspaces were determined using an Aerograph 660 gas chromatograph (Varian Associates, Palo Alto, CA, USA) fitted with a Porapak Q80/100 mesh column (Waters Corporation, Milford, MA, USA) and a thermal conductivity detector. A standard gas mixture containing 5% hydrogen, 30% methane gas with nitrogen as the carrier gas (BOC limited, Palmerston North, New Zealand) was used as a reference.

#### 2.2.4 Cellular fatty acid analysis

Cells from 50 mL cultures of strains grown in BY medium were harvested by centrifugation, and the resulting pellets sent to Hokkaido University (Hokkaido, Japan) for cellular fatty acid analysis (Welch, 1991). Profiles of each strain were determined using gas chromatography with flame-ionisation detection, using the MIDI Sherlock Microbial Identification System (MIS), and the Anaerobic Bacteria Library (MOORE6) for peak identification.

#### 2.2.5 Microscopy

Wet mounts of cells were prepared by aseptically removing approximately 0.1 mL of culture using a sterile needle and 1 mL syringe, placing a drop on a microscope slide and covering with a cover-slip, and applying firm pressure to remove excess culture. Gram staining of cells was carried out by aseptically placing a drop of culture on a microscope slide, allowing to air dry before heat-fixing cells by quickly passing the slide through a Bunsen burner flame and then staining fixed cells using the Gram staining procedure (Coico, 2006). Samples were stained with 10% (w/v) crystal violet for 1 min, followed by fixation of the stain with 0.3%

(w/v) iodine and 0.7% (w/v) potassium iodine for 1 min. The cells were then de-colourised with 50% (v/v) acetone, and counter-stained with 2.5% (w/v) safranin for 30 sec. Slides were gently rinsed with tap water between each step. Wet-mounts and Gram stained cell preparations were examined under 1000× magnification under oil immersion using a Leica DM2500 microscope. The micrograph images produced were captured digitally using the Leica application suite software (Leica Microsystems, Wetzlar, Germany).

Transmission electron microscopy (TEM) of samples of each characterised isolate were carried out at the Massey Microscopy and Imaging Centre (MMIC; Massey University, Palmerston North). Cells of overnight cultures of each strain were harvested by centrifugation, followed by resuspending the cells in sterile N<sub>2</sub>-gassed H<sub>2</sub>O to wash the cells. The resulting suspensions were processed and imaged at the Massey Microscopy and Imaging Centre (MMIC, Palmerston North, New Zealand). Cells were fixed to a Formvar grid and stained in 2% (w/v) uranyl acetate, and then viewed using a FEI Tecnai transmission electron microscope.

Embedded thin sections of cells were prepared for TEM by harvesting cells of overnight cultures by centrifugation, and washing three times by resuspending in sterile water, then resuspending in modified Karnovsky's fixative (2% paraformaldehyde and 3% glutaraldehyde in 0.1 M sodium phosphate buffer (pH 7.4) (provided by MMIC), for processing and imaging (MMIC). Thin sections were prepared using an EM UC7 ultra-microtome (Leica Microsystems, Wetzlar, Germany), and viewed using a FEI Tecnai G2 Biotwin transmission electron microscope.

## 2.2.6 Batch culture experiments

### 2.2.6.1 Substrate utilisation assay

BY medium to assay the substrate utilisation of all strains tested was prepared from a single batch of pooled rumen fluid, to minimise possible effects due to batch variation. All prepared media were stored at room temperature in the dark until use. For substrate utilisation tests, the inoculum needed to contain minimal substrate to avoid carry-over of as little substrate as possible into the test assay. Therefore, the amount of carbohydrate substrate allowing good growth of the test strains after 24 h ( $OD_{600nm} = 0.5 - 1.0$ ) but with minimal residual carbohydrate was determined to be: 0.2% (w/v) xylose for R-7, and 0.2% (w/v) cellobiose for WTE2008. Pre-warmed serum bottles containing 54 mL of BY medium and 3 mL of substrate (to give 0.2% w/v final concentration) were inoculated with 3 mL of culture and incubated for 24 h. Hungate tubes containing 9 mL of media and 0.5 mL of 10% stock solution of each substrate were incubated for 24 h prior to the addition of inocula and were checked for turbidity prior to use, to ensure that media and substrate solutions were not contaminated. Next, 0.5 mL aliquots of inocula from serum bottle cultures were aliquoted into each Hungate tube, resulting in 0.5% concentration of substrate per tube. Test isolates were assessed in triplicate per substrate, whereas one tube per substrate was assessed for controls. The optical density ( $OD_{600}$ ) of each tube was measured immediately after inoculation and subtracted from subsequent readings, which were taken at 24, 48 and 72 h after inoculation. In all instances, a Hungate tube containing uninoculated BY medium with no added substrate was used to zero the spectrophotometer. A no-substrate control was also included in triplicate per strain to detect any background growth in the media, and where detected was subtracted from test readings. Substrates that were used to grow inocula (cellobiose or xylose) were also included in tests as positive controls to confirm cell viability.

For insoluble substrate utilisation assays, microbial growth was assessed *via* fermentation end product detection, since spectroscopy was not suitable. Batch 10 mL cultures in Hungate tubes containing 0.5% (w/v) of insoluble substrate were placed in a tube rack horizontally on a shaker with gentle shaking to keep substrate suspended. After five days, an aliquot was



taken and fermentation end product concentrations were measured, and measurements of samples taken immediately after inoculation were subtracted, to account for potential carry-over of end products in the inoculum or media.

#### 2.2.6.2 Time series experiment of fermentation end product formation during growth

To assess the fermentation end product formation of WTE2008 and R-7 throughout their growth, triplicate batch 10 mL cultures in Hungate tubes growing in BY medium made with rumen fluid of a pasture-fed sheep, 0.1% (w/v) each of Bacto-peptone and casamino acids, and 0.5% (w/v) cellobiose as substrate were incubated at 39°C, and their growth and production of fermentation end products were measured at 0, 4, 8, 16, and 24 hours of incubation. At each time point, optical density of tubes were measured using a Spectronic 200 spectrophotometer (Thermo Fisher). A 300 µL aliquot of culture was removed for SCFA analyses (section 2.2.3) aseptically using a CO<sub>2</sub>-gassed needle and syringe, expelling the same volume of CO<sub>2</sub> into the tube as was taken each time. To measure hydrogen production, a 100 µL aliquot of culture headspace was also removed from each tube with a sterile needle and 1 mL luer-lock syringe using a push button valve (section 2.2.3).

#### 2.2.6.3 Screening of strains for cobalamin-dependent propionate production

To screen *Prevotella* 1 strains for their growth and fermentation end product formation in the presence and absence of cobalamin, cultures of strains were revived in M2GSC media, before being passaged three times on Strobel medium containing vitamin solution without added cyanocobalamin by transferring 0.5 mL inoculum into 9.5 mL of medium, to remove background cobalamin in the revival medium. Triplicate cultures of each strain were inoculated into tubes of 10 mL Strobel medium containing vitamin solution with and without 50 µg/L cyanocobalamin. Optical densities and samples for SCFA analyses were taken after 48 hours of incubation at 39°C, and were subtracted from measurements taken immediately after inoculation.

#### 2.2.6.4 Comparative omics' experiment of cobalamin-induced propionate production

To investigate the genes and proteins that are upregulated during cobalamin induction of the propionate pathway in KHP1, 100 mL cultures of *P. ruminicola* KHP1 were grown in the presence and absence of cobalamin (n = 6 per treatment). At two-hourly timepoints, as well as immediately after inoculation, 1 mL was taken from each culture using a sterile needle and CO<sub>2</sub>-flushed syringe, and the optical density (600 nm) was measured using a Spectronic 200 spectrophotometer (Thermo Fisher). The sample was transferred from the cuvette to an Eppendorf tube, and fermentation end product concentrations were measured (see section 2.2.3). Cultures were harvested during log phase growth after 10 hours by flash-freezing in liquid nitrogen, and were stored at -80°C for transcriptome and proteome analyses.

### 2.2.7 Molecular methods

#### 2.2.7.1 16S rRNA gene sequencing

DNA was extracted from cells harvested from 1 mL of turbid culture using either the GeneMatrix DNA preparation kit (Bio-Rad) or the Nucleospin Soil DNA extraction kit (Macherey-Nagel), following the manufacturer's instructions. A 1 µL aliquot of DNA extract was then used as template in PCR reactions using the Platinum green PCR mix kit (Thermo Fisher Scientific). The primer pair fD1 (5'-AGAGTTTGATCCTGGCTCAG-3') and rD1 (5'-AAGGAGGTGATCCAGCC-3') (Weisburg et al., 1991) were used to amplify the full-length bacterial 16S ribosomal RNA (rRNA) gene. Amplification was performed using a Mastercycler pro S thermal cycler (Eppendorf, Hamburg, Germany) under the following conditions: initial denaturation at 95°C for 4 min, followed by 25 cycles of denaturation at 95°C, annealing at 55°C each for 30 s, and extension 72°C for 1 min, followed by a final extension at 72°C for 10 min. PCR products were electrophoresed on 0.8% (w/v) agarose gels to determine the presence of the ~1.5 kb 16S rRNA gene amplicon. PCR products were purified using the QIAquick PCR purification kit (Qiagen) following the manufacturer's instructions. Purified PCR-products for each strain were then diluted in 25 µL nuclease-free H<sub>2</sub>O (Invitrogen) to 1 ng/µL in two tubes, each containing 4 pmol of either the forward or reverse primer used. Samples

were sequenced at the Massey Genome Service (Massey University, Palmerston North) using Sanger ABI sequencing. This service included fluorescent labelling of PCR products using BigDye™ Terminator (Applied Biosystems; Version 3.1), a Ready Reaction Cycle Sequencing Kit, subsequent removal of unincorporated fluorescent dideoxy NTPs (ddNTPs) by clean up and precipitation of products and capillary separation on an ABI3730 Genetic Analyser (Applied Biosystems, Foster City, CA, USA).

### 2.2.7.2 Agarose gel electrophoresis

Agarose gels were made by adding 0.2 g (for total DNA extracts (section 2.2.7.3)) or 0.4 g (for 16S rRNA amplicons (section 2.2.7.1)) of molecular-grade agarose (Fisher Biotec, Australia) to 50 mL 1× TAE buffer (section 2.1.2) in a conical flask, and boiling in a microwave until molten. Flasks were cooled briefly under running tap water, and 5 µL of SYBR safe (Thermo Fisher) was added to agarose and swirled to mix. Gels were cast and electrophoresed using the Horizon 58 system (Biometra). Amplicons were electrophoresed for approximately 1 h at 90 V, and total DNA extracts for 2-3 h at 40 V. Gels were photographed using a Nikon D700 camera with the Kodak Gel Logic 200 Imaging System (Eastman Kodak, Rochester, NY, USA).

### 2.2.7.3 DNA extraction and genome sequencing

High molecular-weight DNA was extracted from cultures using a modified phenol-chloroform extraction protocol based on that described in Palevich (2016). Cultures 50 mL in volume in serum bottles were incubated at 39°C for 1-2 days. Due to its poorer growth ( $OD_{600} < 0.3$  after 48h), three 50 mL cultures of strain XBB3002 were incubated and combined after growth. Cultures were then centrifuged at  $5,000 \times g$  for 5 min to pellet cells, and the resulting pellets were washed by resuspending in saline-EDTA solution (section 2.1.2) and centrifuging again at  $5,000 \times g$  for 5 min. Pellets were then resuspended by gentle agitation in 1 mL freshly prepared lysis buffer, and the lysate was incubated at 37°C for an hour with gentle agitation each half-hour. SDS solution (20%) was then added to give a final concentration of 1% (w/v), and proteinase K (Invitrogen) to a final concentration of 200 µg/mL, and lysates were

incubated a further 1.5 hr at 60°C. Lysates were then buffered by adding 400 µL of Tris-EDTA (section 2.1.2), and mixed in a 1:1 ratio with phenol-chloroform-isoamyl alcohol solution (25:24:1), mixing well by inversion and phase separation by centrifugation at 13,000 × *g*. The aqueous layer was then transferred to a separate tube and again mixed in a 1:1 ratio with phenol-chloroform-isoamyl alcohol solution. The same procedure was carried out mixing the resulting aqueous phase an additional two times with pure chloroform. Three volumes of 96% ethanol and 1:10 volume of 5 M ammonium acetate were added to each extract, mixed well, and stored overnight at -20°C. Precipitated DNA was pelleted by centrifugation at 16,000 × *g* for 30 min, and the resulting pellet was washed twice with 70% ethanol by centrifugation at 16,000 × *g* for 5 min. Supernatant was then carefully removed and the pellet was allowed to dry by incubating at 37°C for 30 min. The pellet was gently resuspended by pipetting an appropriate volume of elution buffer using a wide-bore pipette tip, and stored at 4°C. Genome sequencing and assembly was carried out by Nextomics Biosciences (Wuhan, China), A hybrid long/short read approach was used, with sequencing performed using PromethION (Oxford Nanopore technologies (ONT)) and MGISEQ-2000 sequencing instruments.

#### 2.2.7.4 RNA extraction, library preparation and sequencing

RNA extractions were carried out using a modified phenol-chloroform procedure with bead-beating, based on the method described in Lueders et al. (2004). Flash-frozen material was removed from -80°C storage on dry ice, and approximately 4 g of frozen culture material was transferred using a metal spatula sprayed with RNaseZAP (Sigma-Aldrich) into a 50 mL falcon tube, containing 4 mL of RNeasy Protect Bacterial Reagent (QIAGEN), and thawed on ice for approximately two hours with gentle agitation every half hour. Samples were separated into four 2 mL bead beating tubes, and were centrifuged at 20,000 × *g* for 10 minutes at 4°C and carefully removing the supernatant. To each pellet, 500 µL of RNA extraction buffer (section 2.1.2), 210 µL of 20% SDS, 500 mg of Zirconia/Silica beads (0.1:0.5mm, 1:1), and 500 µL of acid phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5) was added. The resulting mixtures were lysed by mechanical bead-beating using a Mini-Beadbeater (BioSpec products inc, OK, United States) three times for 2 min at the preset speed, with a 10 min incubation on ice between cycles to prevent samples from over-heating. Samples were then centrifuged at 10,000 × *g*,

and the aqueous top layer was transferred to a new Eppendorf tube. An equal volume of isopropanol and 0.1 volumes of 5 M ammonium acetate (pH 5.5) (Invitrogen) were added, mixed and stored at -20°C overnight. Precipitated RNA was pelleted at 10,000 × g for 30 min at 4°C. The isopropanol was carefully removed, and the pellet was washed twice with RNase-free 70% ethanol, centrifuging for 5 min at maximum speed on bench-top centrifuge. The resulting RNA extracts were then DNase treated using the TURBO DNase kit (Invitrogen), following the manufacturer's instructions. RNA was then purified using the MEGAclean Transcription Cleanup Kit (Invitrogen), following the manufacturer's instructions. Samples were stored at -80°C and sent to Novogene (Beijing, China) for sequencing. Samples were depleted of rRNA using the riboZERO magnetic kit (Illumina) and library preparation using the NEBnext Ultra II RNA Library Preparation kit (Illumina), following the manufacturers' instructions. The resulting libraries were sequenced using an Illumina NovaSeq instrument.

#### 2.2.7.5 DNA/RNA quantification and quality control

DNA samples were assessed for purity and quantified using an Implen NanoPhotometer. For samples sent for whole genome sequencing, samples were additionally quantified using the Qubit Broad Range DNA quantification kit (Thermo Fisher Scientific) following the manufacturer's instructions, and fragment size assessed by agarose gel electrophoresis (section 2.2.7.2).

RNA samples were assessed of purity using an Implen NanoPhotometer, quantified using the Qubit RNA Broad Range quantification kit (Thermo Fisher Scientific), and assessed for integrity using an RNA 6000 chip on an Agilent 2100 Bioanalyser.

#### 2.2.7.6 Protein extraction and mass-spectrometry

Frozen cultures were sent to the Norwegian University of Life Sciences (Norway) on dry ice for protein extraction and mass-spectrometry. FastPrep tubes were filled with 4 mm glass beads ( $\leq 106 \mu\text{m}$ ) (Thermo Fisher Scientific), 500  $\mu\text{L}$  of thawed culture, and 250  $\mu\text{L}$  lysis buffer

(30 mM dithiothreitol (DTT), 150 mM Tris-HCl (pH=8), 0.3% Triton X-100, 12% SDS), followed by brief vortexing and resting on ice for 30 min. Lysis was performed using a FastPrep 24 Classic Grinder (MP Biomedical, Ohio, USA) for three cycles of 60 seconds at 4.0 m/s. Samples were centrifuged at  $16,000 \times g$  at  $4^{\circ}\text{C}$ , and lysate carefully removed. Protein concentrations were measured using the Bio-Rad DC<sup>TM</sup> Protein Assay (Bio-Rad, California, USA) with bovine serum albumin as standard. Absorbance of lysates was measured at A750 on BioTek Synergy<sup>TM</sup> H4 Hybrid Microplate Reader (Fisher Scientific). An aliquot of 40-50  $\mu\text{g}$  of protein of each sample was prepared in SDS buffer, heated in a water bath for 5 min for  $99^{\circ}\text{C}$  and analysed by SDS-PAGE using Any-kD mini-PROTEAN TGX stain-free<sup>TM</sup> gels (Bio-Rad) in a 2 minute run for sample clean-up purposes, before staining with Coomassie Blue R-250. Visible bands were carefully excised from the gel and divided into  $1 \times 1$  mm pieces. Samples were then destained, reduced, alkylated, digested, and the resulting peptides extracted and desalted using the OASIS<sup>®</sup> HLB  $\mu$ Elution plate (Franz & Lee, 2012) following the manufacturers' instructions. Peptides were analysed by nano-LC-MS/MS using a Q-Exactive hybrid quadrupole Orbitrap MS (Thermo Fisher) as described in (Arntzen et al., 2015). Raw MS files were analysed using MaxQuant (Tyanova et al., 2016a) and proteins identified and quantified using the MaxLFQ algorithm (Cox & Mann, 2008).

## 2.2.8 Bioinformatics

### 2.2.8.1 16S rRNA gene sequence analysis

The quality of base-calling of the sequencing chromatograms was assessed using Geneious v10.0.9 (Kearse et al., 2012). Forward and reverse sequence reads for each isolate were assembled using the *de-novo* assembly option (default settings). The approximately 1,300-1,400 bp region of each assembled sequence where chromatogram peaks of each base-call were reliable (with unassigned bases removed) was then searched against the NCBI nr database using BLASTn (Altschul et al., 1997) under default settings to identify the best hits to each sequence.

Phylogenetic maximum-likelihood trees of 16S rRNA gene sequences were constructed in MEGA X (Kumar et al., 2018). Sequences were firstly aligned using MUSCLE (Edgar, 2004) under default settings. Maximum-likelihood trees were then built from the resulting alignments using the Tamura-Nei model (Tamura & Nei, 1993) under default settings, and assessed using 500 bootstrap replications.

### 2.2.8.2 Genome assembly and analyses

Genome sequencing and assembly was carried out by Nextomics Biosciences (Wuhan, China), as described in Mahoney-Kurpe et al. (2021) (Section 3.2). Long reads were quality filtered ( $Q > 7$ , sequence length  $> 1000$  bp) using Guppy 4.0.11. Short MGISEQ reads were quality filtered using fastp (Chen et al., 2018); following removal of adapters, reads containing N base-calls were removed. Reads had 5 bp trimmed from each end and read-pairs for which at least one read had  $>20\%$  of bases with  $Q < 20$  were removed. Quality-filtered long reads were assembled using Flye v2.7, with the `–plasmid` and `–nano-raw` settings. Assemblies were further polished with Racon v1.4.13 (Vaser et al., 2017) under default settings using alignments of the ONT data, and with Pilon v1.23 (Walker et al., 2014) and NextPolish v1.2.4 (Hu et al., 2019) using alignments of the MGISEQ data. The resulting contigs were circularised using Circlator v1.5.1 (Hunt et al., 2015) using the ‘fixstart’ parameter. Genomes were annotated using the NCBI PGAP pipeline v5.0 (Tatusova et al., 2016).

Average nucleotide identity (ANI) distances between strains were calculated using fastANI v1.32 (Jain et al., 2018) under default settings. Taxonomic classification of strains were made using GTDB-Tk v1.3.0 (Chaumeil et al., 2020) to query the Genome Taxonomy Database (GTDB) (release R5-RS95) (Parks et al., 2018), and was carried out by Dr Sandeep Kumar (AgResearch, New Zealand). Full-length 16S rRNA sequences were detected and extracted from complete genome assemblies using RNAmmer v1.2 (Lagesen et al., 2007). Draft genome sequences of the Hungate1000 collection were analysed using the Integrated Microbial Genomes server (IMG/MER) (Chen et al., 2021). Scanning of contigs for cobalamin-binding riboswitches was carried out using riboswitch scanner (Mukherjee & Sengupta, 2016),

searching for 'cobalamin', 'ado-cbl' and 'adoCbl-variant' riboswitch families. Functional COG annotations were made using eggno-mapper v2 (Huerta-Cepas et al., 2017).

### 2.2.8.3 Creation of a Hungate1000 protein database

An amino acid sequence file generated from all 501 genomes of the Hungate1000 collection was obtained from Dr Sinead Leahy (AgResearch Grasslands). To retrieve NCBI accession.version identifiers for each amino acid sequence, this file was processed using the 'blastp' command of DIAMOND (version 0.9.21.122) (Buchfink et al., 2014) against the NCBI nr database (as of the 28<sup>th</sup> of May 2018) using all the default settings, with the exception of the output file which was changed to the 12 column BLAST output. The resulting hits were filtered by Paul Maclean (AgResearch Grasslands) using a 97% similarity threshold and an E-value threshold of  $1e^{-100}$ . For each amino acid sequence passing this filter, the unique sequences of the top 25 significant hits were retrieved from the nr database and collated into a fasta file. The resulting fasta file was then used to create a DIAMOND database file using the 'makedb' command, which was used as the database file.

### 2.2.8.4 Metatranscriptome analyses

Metatranscriptomic reads of the high and low methane yield sheep from the Shi et al. (2014) dataset were aligned against the generated protein database file of the Hungate1000 collection (section 2.2.8.3) using the 'blastx' command (default settings) in DIAMOND v0.9.21.122 (Buchfink et al., 2014). Functional and taxonomic information was assigned to reads using the *daa-meganizer* command with the Top Percent option set to 0.001, using the June 2018 version of the taxonomy mapping file and the December 2017 version of the Kyoto Encyclopedia of Genes and Genomes (KEGG) mapping file, obtained from the MEGAN website (see Table 2.3). The resulting files were analysed in the MEGAN 6 (Huson et al., 2016) Ultimate Edition graphical user interface. From each file, reads assigned to KEGG Orthology (KO) (Kanehisa et al., 2016) or InterPro2GO (Mitchell et al., 2015) categories corresponding to enzymes candidates involved in propionate production from succinate (Table 2.2) were extracted into new files, which were then reanalysed in MEGAN separately to determine the



resulting taxa profiles assigned to reads of each extracted category. For all analyses, abundances of reads mapping to functional categories of interest were expressed as reads per million (RPM) to normalise differences in sequencing depth between samples. All relative abundances of different taxonomic groups were calculated as the average proportion of reads mapping to each taxon group over the total number of reads assigned to each analysed KEGG/Interpro2go category per sample.

**Table 2.2. Functional categories corresponding to enzymes involved in the conversion of succinate to propionate included in metatranscriptome analyses.**

Enzymatic step	Functional category name	KEGG/Interpro ID
succinate → succinyl-CoA	succinate CoA-transferase	IPR017821
succinyl-CoA → R-methylmalonyl-CoA	methylmalonyl-CoA mutase	K01847
S-methylmalonyl-CoA → propionyl-CoA	Na <sup>+</sup> -transporting methylmalonyl-CoA/oxaloacetate decarboxylase beta subunit	IPR005661
propionyl-CoA → propionate	propionate CoA-transferase	K01026
	phosphate acetyltransferase	K00625
	phosphate butyryltransferase	K00634
	acetate/propionate kinase	IPR004372
	butyrate kinase	IPR011245

### 2.2.8.5 Transcriptome analyses

Raw reads were trimmed of any remaining adapter sequences (default settings) and quality-filtered (-q 20) using cutadapt v3.3 (Martin, 2011). Filtered reads were then aligned against the complete *P. ruminicola* KHP1 genome using Hisat2 v2.2.1 (Kim et al., 2019) under default settings. SAM alignment files were converted to BAM format using samtools v1.11 (Li et al., 2009), and featureCounts v2.0.1 (Liao et al., 2014) was used to extract read alignment counts. To identify differentially expressed genes, the resulting matrix was input into R v4.1.1, and read counts were log<sub>2</sub>-transformed. Differentially expressed genes were identified between

treatments using the DESeq2 package (Love et al., 2014), with significance defined using false discovery rate (FDR)-adjusted  $P < 0.05$  and  $|\log_2\text{fold change}| \geq 1$  cutoffs. Principal Component Analysis (PCA) was carried out on  $\log_2$ -transformed read counts of each gene of the KHP1 genome using the 'prcomp' function in R, and plotted with the 'ggfortify' package.

#### 2.2.8.6 Proteome analyses

ProteinGroups files from MaxQuant were further processed and analysed in Perseus v1.6.15.0 (Tyanova et al., 2016b). Proteins identified as contaminants, reverse proteins and proteins identified only by site were filtered. Label-free quantification (LFQ) intensities of proteins were logarithmically normalised ( $\log_2$ ), and missing data were imputed based on the normal distribution. Differentially abundant proteins were identified by carrying out a two-sample T-test, with significance defined using FDR-adjusted  $P < 0.05$  and  $|\log_2\text{fold change}| \geq 1$  cutoffs.

**Table 2.3. Bioinformatic tools used in this thesis.**

Software name	Version	Parameters	Website and reference
BLAST	web server (accessed post-June 2018)	blastn (default) blastp (default)	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a> (Altschul et al., 1997)
circlator	1.5.1	'fixstart'	<a href="https://sanger-pathogens.github.io/circlator/">https://sanger-pathogens.github.io/circlator/</a> (Hunt et al., 2015)
cutadapt	3.3	-q 20	<a href="https://cutadapt.readthedocs.io/en/stable/index.html">https://cutadapt.readthedocs.io/en/stable/index.html</a> (Martin, 2011)
DESeq2	1.32.0	default	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a> (Love et al., 2014)
DIAMOND	0.9.21.122	default	<a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a> (Buchfink et al., 2015)
EggNOG-mapper	2.0	default	<a href="http://eggnog-mapper.embl.de/">http://eggnog-mapper.embl.de/</a> (Huerta-Capas et al., 2017)
fastANI	1.32	default	<a href="https://github.com/ParBLISS/FastANI">https://github.com/ParBLISS/FastANI</a> (Jain et al., 2018)
FeatureCounts	2.0.1	-s 2	<a href="http://subread.sourceforge.net/">http://subread.sourceforge.net/</a> (Liao et al., 2014)
Flye	3.2	--plasmid, --nano-raw	<a href="https://github.com/fenderglass/Flye">https://github.com/fenderglass/Flye</a> (Kolmogorov et al., 2019)
Geneious	10.0.9	default	<a href="https://www.geneious.com/">https://www.geneious.com/</a> (Kearse et al., 2012)
GTDB-Tk	1.3.0	default	<a href="https://github.com/GenomeTaxonomics/GTDBTk">https://github.com/GenomeTaxonomics/GTDBTk</a> (Chaumeil et al., 2020)
Hisat2	2.2.1	--rna-strandness RF	<a href="http://daehwankimlab.github.io/hisat2/">http://daehwankimlab.github.io/hisat2/</a> (Kim et al., 2019)
MEGA	10.1.5	default	<a href="https://www.megasoftware.net/">https://www.megasoftware.net/</a> (Kumar et al., 2018)
MEGAN	6.0	-Top Percent 0.001	<a href="https://software-ab.informatik.uni-tuebingen.de/download/megan6/welcome.html">https://software-ab.informatik.uni-tuebingen.de/download/megan6/welcome.html</a> (Huson et al., 2016)
MUSCLE	5.0	default	<a href="https://www.drive5.com/muscle/">https://www.drive5.com/muscle/</a> (Edgar, 2004)
NextPolish	1.2.4	default	<a href="https://github.com/Nextomics/NextPolish">https://github.com/Nextomics/NextPolish</a> (Hu et al., 2020)
Perseus	1.6.15.0	not applicable	<a href="https://maxquant.net/perseus/">https://maxquant.net/perseus/</a> (Tyanova et al., 2016)
Pilon	1.23	default	<a href="https://github.com/broadinstitute/pilon">https://github.com/broadinstitute/pilon</a> (Walker et al., 2014)
racon	1.4.13	default	<a href="https://github.com/isovic/racon">https://github.com/isovic/racon</a> (Vaser et al., 2017)
Riboswitch scanner	web server (accessed post-september 2020)	Families searched: 'adoCbl', 'adoCbl-variant', 'cobalamin'	<a href="http://service.iiserkol.ac.in/~riboscan/">http://service.iiserkol.ac.in/~riboscan/</a> (Mukherjee & Sengupta, 2016)
RNAmmmer	1.2	default	<a href="https://services.healthtech.dtu.dk/service.php?RNAmmmer-1.2">https://services.healthtech.dtu.dk/service.php?RNAmmmer-1.2</a> (Lagesen et al., 2007)
samtools	1.11	default	<a href="http://www.htslib.org/">http://www.htslib.org/</a> (Li et al., 2009)

### 3 Characterisation of the rumen R-7 bacterial group

## 3.1 Introduction

Owing to decades of microbiological research (Hungate, 1966), some of the dominant groups of culturable rumen bacteria are well characterised. However, following advances in high throughput sequencing technologies enabling comprehensive cultivation-independent identification of community members, it became apparent that some of the more abundant taxonomic groups have so far evaded characterisation (Creevey et al., 2014). One notable example is the *Clostridiales* R-7 group, which high-throughput metabarcoding studies have demonstrated is one of the most abundant bacterial groups in the rumen (Henderson et al., 2015). Strains of the group were shown to also possess a wide variety of carbohydrate-active enzymes (CAZymes) (Seshadri et al., 2018), leading to speculation that the R-7 group may have a prominent role in fibre degradation.

At the beginning of this project, the draft genome sequences of only two ruminal strains of the R-7 group were available, R-7, and WTE2008, as part of the Hungate1000 reference genome study (Seshadri et al., 2018). To enable more robust genomic analyses of this group, the complete genome sequences of an additional three rumen R-7 group strains (Noel, 2013) were generated. This work is outlined in Section 3.2, and was published in *Microbiology Resource Announcements* (American Society for Microbiology) (Appendix 7.1). While the draft genome sequences of R-7 and WTE2008 were available as part of the Hungate1000 study (Seshadri et al., 2018), neither strain had been phenotypically characterised, and they were not taxonomically classified beyond the family level (Henderson et al., 2019). To do so, phylogenetic and phenotypic characterisation of these strains were carried out. As a result of these analyses, the strains were proposed as two separate species, *Aristaeella hokkaidonensis* gen. nov. sp. nov. (R-7) and *Aristaeella lactis* sp. nov. (WTE2008), within a novel proposed family, *Aristaeellaceae* fam. nov. A manuscript describing this work (Section 3.3) has been prepared for submission to the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM).

## 3.2 Complete genome sequences of three *Clostridiales* R-7 group strains isolated from the bovine rumen in New Zealand

### 3.2.1 Abstract

The *Clostridiales* R-7 group are abundant bacterial residents of the rumen microbiome; however, they are poorly characterised. We report the complete genome sequences of three members of the R-7 group, FE2010, FE2011 and XBB3002, isolated from the ruminal contents of pasture-grazed dairy cows in New Zealand.

### 3.2.2 Announcement

The *Clostridiales* R-7 group is an abundant, but poorly characterised, group of unclassified rumen bacteria (Henderson et al., 2015). Draft genomes of only two strains are currently available (Seshadri et al., 2018). The genomic characterisation of additional members will accelerate efforts to understand the roles of the R-7 group in the rumen microbial ecosystem.

Rumen contents of fistulated dairy cows grazing ryegrass/clover pastures in Waikato, New Zealand (Noel et al., 2017) were obtained with AgResearch Grasslands Animal Ethics Committee approval (AE12174). An anaerobic dilution-to-extinction approach (Button et al., 1993) was used to isolate FE2010 and FE2011 in RM02 medium supplemented with glucose, cellobiose, xylose, L-arabinose, lactate, casamino acids, Bacto-peptone and yeast extract (Kenters et al., 2011), and XBB3002 in BY medium (Joblin, 2005) at 39°C. FE2010 and FE2011 cells were Gram-negative rods, whereas XBB3002 were coccobacilli. Partial 16S rRNA gene sequences exhibited >96% nucleotide identity to rumen strain R-7 (Noel, 2013).

High molecular-weight genomic DNA was extracted using a chemical/enzymatic lysis and phenol-chloroform extraction method (Palevich, 2016) from 1-2 day old cultures grown anaerobically in BY medium at 39°C. DNA was sequenced and assembled by Nextomics Biosciences (Wuhan, China). Long-read libraries were prepared using the native barcoding expansion (NBD-104) and SQK-LSK109 ligation sequencing kits and sequenced on a PromethION instrument, using Guppy (v4.0.11; Oxford Nanopore Technologies (ONT)) for base-calling and quality-filtering (Q >7, sequence length >1000 bp). Short-read (2x150 bp) libraries were prepared using the MGISEQ-2000RS kit and sequenced using an MGISEQ-2000 instrument. Short reads were quality filtered with fastp (v0.20.0) (Chen et al., 2018), following removal of adapters, reads containing N base-calls were removed. Reads had 5 bp trimmed from each end and read-pairs for which at least one read had >20% of bases with Q <20 were removed. Quality-filtered ONT reads were assembled using Flye (v2.7; –plasmid and –nano-raw settings) (Kolmogorov et al., 2019). Assemblies were polished with Racon (v1.4.13; default settings) (Vaser et al., 2017) using alignments of ONT data, and by Pilon (v1.23; default settings) (Walker et al., 2014) and NextPolish (v1.2.4; default settings) (Hu et al., 2019) using alignments of the short-read data. Contigs were confirmed as circular using Circlator (v1.5.1; parameter='fixstart') (Hunt et al., 2015). Annotation was performed using the NCBI PGAP pipeline (v5.0) (Tatusova et al., 2016). Sequences from each isolate assembled into circular contigs of similar size and %G+C content (Table 3.1).

Taxonomic assignments were determined using GTDB-Tk (v1.3.0) (Chaumeil et al., 2020), to query the Genome Taxonomy Database (GTDB) framework (Parks et al., 2018) (release 05-RS95). The strains were classified as members of the recently proposed order '*Christensenellales*' (Parks et al., 2018) in an undescribed family (CAG-74) and genus (GCA-900199385), which the previously sequenced R-7 group strains R-7 (species-level taxon sp900199385) and WTE2008 (sp900176495) (Seshadri et al., 2018) have also been classified as. FE2010 and FE2011 were assigned to the species-level taxon sp900322155, while XBB3002 was unassigned at the species level. These genomes expand the number of sequenced representatives of the R-7 group and will progress our understanding of their biology.

**Table 3.1. Genome details of R-7 group strains**

Parameter	FE2010	FE2011	XBB3002
BioProject accession	<a href="#">PRJNA695064</a>	<a href="#">PRJNA695064</a>	<a href="#">PRJNA695064</a>
BioSample accession	<a href="#">SAMN17600269</a>	<a href="#">SAMN17611198</a>	<a href="#">SAMN17611949</a>
GenBank accession	<a href="#">CP069593</a>	<a href="#">CP069418</a>	<a href="#">CP069419</a>
SRA accession (ONT)	<a href="#">SRX10247579</a>	<a href="#">SRX10248369</a>	<a href="#">SRX10248392</a>
SRA accession (MGISEQ)	<a href="#">SRX10247580</a>	<a href="#">SRX10248370</a>	<a href="#">SRX10248393</a>
No. of raw ONT reads	815,513	966,753	443,226
No. of filtered ONT reads	781,555	926,448	419,863
<i>N</i> <sub>50</sub> of filtered ONT reads (bp)	4,724	3,808	6,131
No. of raw MGISEQ reads	6,874,216	6,886,570	6,896,410
No. of filtered MGISEQ reads	6,866,512	6,878,218	6,888,148
Genome size (Mb)	3.51	3.56	3.26
No. of contigs	1	1	1
Sequencing coverage	650x	628x	457x
G+C content (%)	53.2	53.2	56.5

**Data availability.** Complete genomes and raw sequence reads are available in GenBank and the Sequence Read Archive under the accession numbers in Table 3.1.

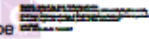
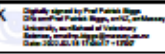
### Acknowledgements

This work was funded by the New Zealand Ministry of Business, Innovation and Employment Strategic Science Investment Fund to AgResearch. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.



## STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sam Mahoney-Kurpe
Name/title of Primary Supervisor:	Prof. Patrick Biggs
In which chapter is the manuscript /published work:	Chapter 3, Section 3.2
<p>Please select one of the following three options:</p> <p><input checked="" type="radio"/> The manuscript/published work is published or in press</p> <ul style="list-style-type: none"> <li>• Please provide the full reference of the Research Output: Mahoney-Kurpe, S. C., Palevich, N., Noel, S. J., Kumar, S., Gagic, D., Biggs, P. J., Janssen, P. H., Attwood, G. T., Moon, C. D. (2021) Complete genome sequences of three Clostridiales R-7 group strains isolated from the bovine rumen in New Zealand. <i>Microbiology Resource Announcements</i>, 10 (26), e0031021.</li> </ul> <p><input type="radio"/> The manuscript is currently under review for publication – please indicate:</p> <ul style="list-style-type: none"> <li>• The name of the journal: <i>Microbiology Resource Announcements</i></li> <li>• The percentage of the manuscript/published work that was contributed by the candidate: <span style="float: right;">85.00</span></li> <li>• Describe the contribution that the candidate has made to the manuscript/published work: Conceptualisation, Investigation, Formal analysis, Writing- Original draft.</li> </ul> <p><input type="radio"/> It is intended that the manuscript will be published, but it has not yet been submitted to a journal</p>	
Candidate's Signature:	Sam Mahoney-Kurpe 
Date:	18-Feb-2022
Primary Supervisor's Signature:	Prof Patrick Biggs 
Date:	18-Feb-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

### 3.3 Description of *Aristaeella hokkaidonensis* gen. nov. sp. nov. and *Aristaeella lactis* sp. nov., two rumen bacterial species of a novel proposed family, *Aristaeellaceae* fam. nov.

#### 3.3.1 Abstract

Two strains of Gram negative, anaerobic, rod-shaped bacteria, from an abundant but uncharacterised rumen bacterial group of the order '*Christensenellales*', were phylogenetically and phenotypically characterised. These strains, designated R-7<sup>T</sup> and WTE2008<sup>T</sup>, shared 98.6-99.0% sequence identity between their 16S rRNA gene sequences. R-7<sup>T</sup> and WTE2008<sup>T</sup> clustered together on a distinct branch within a 16S rRNA gene radiation of closely related (>97% sequence identity) '*Christensenellales*' strains that had <85.4% sequence identity to the closest type-strain, *Christensenella minuta* YIT 12065<sup>T</sup>. The genome sequences of R-7<sup>T</sup> and WTE2008<sup>T</sup> had 84% average nucleotide identity to each other, and taxonomic assignment using the Genome Taxonomy Database indicates these are separate species within a novel family of the order '*Christensenellales*'. Cells of R-7<sup>T</sup> and WTE2008<sup>T</sup> lacked any obvious appendages, and had cell wall ultra-structures characteristic of Gram-negative bacteria. The five most abundant cellular fatty acids of both strains were C<sub>16:0</sub>, C<sub>16:0</sub> iso, C<sub>17:0</sub> anteiso, C<sub>18:0</sub>, and C<sub>15:0</sub> anteiso. The strains used a wide range of the 23 soluble carbon sources tested, and grew best on cellobiose, but not on sugar-alcohols. Xylan and pectin were fermented by both strains, but not cellulose. Acetate, hydrogen, ethanol and lactate were the major fermentation end products. R-7<sup>T</sup> produced considerably more hydrogen than WTE2008<sup>T</sup>, which produced more lactate. Based on these analyses, *Aristaeellaceae* fam. nov., and *Aristaeella* gen. nov., with type species *Aristaeella hokkaidonensis* sp. nov., are proposed. Strains R-7<sup>T</sup> (= DSM 112795<sup>T</sup>) and WTE2008<sup>T</sup> (= DSM 112788<sup>T</sup>) are proposed type-strains for *Aristaeella hokkaidonensis* sp. nov. and *Aristaeella lactis* sp. nov., respectively.

### 3.3.2 Introduction

The rumen microbiota play a pivotal role in ruminant nutrition and metabolism. This complex microbial community is primarily responsible for the breakdown and fermentation of ingested plant-based feeds to generate substrates that are a primary energy source for the host (Hobson & Stewart, 1997). Despite decades of research, some of the dominant microbial groups in the rumen remain unstudied (Creevey et al., 2014). One such example is the R-7 group, which cultivation-independent studies have demonstrated is one of the most abundant rumen bacterial groups, and part of a core rumen microbiota across different host species fed different diets (Henderson et al., 2015). Taxonomic placement of the group is currently unresolved; based on 16S rRNA gene phylogeny it was predicted to fall within the order *Clostridiales* (Henderson et al., 2015) and family *Christensenellaceae* (Henderson et al., 2019), while more recent phylogenomic analyses using the Genome Taxonomy Database (GTDB) framework instead assigns the group as belonging to a novel family and genus within the recently proposed (Parks et al., 2018) order '*Christensenellales*' (Mahoney-Kurpe et al., 2021). Two strains of the R-7 group are members of the Hungate1000 reference collection of rumen microbial strains for which draft genome sequences had been generated (Seshadri et al., 2018). However, neither R-7 group strain had been formally classified taxonomically. This study reports the phylogenomic and phenotypic characterisation of R-7<sup>T</sup> and WTE2008<sup>T</sup>, and formal descriptions of the new taxa that they represent.

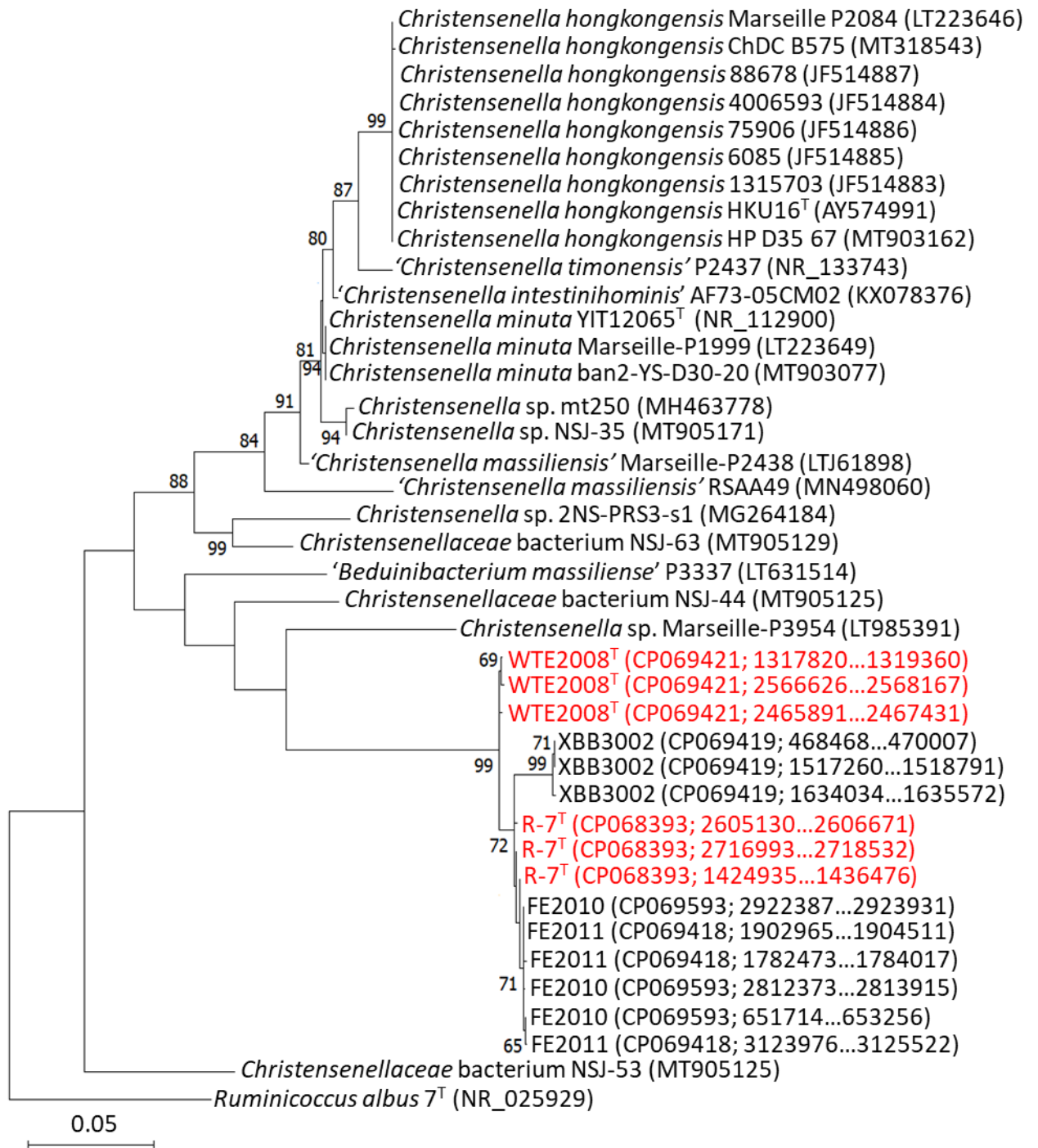
### 3.3.3 Habitat and isolation

Strain R-7<sup>T</sup> was isolated in Hokkaido, Japan, from the rumen of a sheep fed a diet of timothy hay and concentrate at a 9:1 ratio of dry matter. Rumen contents were incubated with cellulose powder for 10 min at 37 °C, before harvesting cellulose and the adherent bacteria by centrifugation (250 × *g* for 10 min). After washing the cellulose pellet five times with phosphate-buffered saline, the pellet was resuspended in anaerobic dilution solution (Bryant & Burkey, 1953). Serial dilutions of the resuspended pellet were then used as inoculum for isolations using the anaerobic roll-tube method with RGCMSA agar medium, as previously described (Minato et al., 1992).

Strain WTE2008<sup>T</sup> was isolated from rumen contents collected from a dairy cow grazing a ryegrass pasture in Waikato, New Zealand (Noel, 2013). Isolation was achieved through a dilution-to-extinction approach (Button et al., 1993), in which liquid batch cultures of RM02 medium (Kenters et al., 2011) supplemented with glucose, cellobiose, xylose, L-arabinose, casamino acids, Bacto-peptone and yeast extract, then reduced using titanium (III) nitrilotriacetic acid (Moench & Zeikus, 1983) were inoculated with serial dilutions of rumen contents.

### 3.3.4 16S rRNA gene phylogeny

Full-length 16S rRNA gene sequences for each strain were extracted from complete genome assemblies (see Genome features section 3.3.5) using RNAMmer (version 1.2) (Lagesen et al., 2007). Each genome possessed three copies of the 16S rRNA gene, and each copy was >99 % identical to each other. Within MEGA X (Kumar et al., 2018) the rRNA gene sequences per assembly were aligned with >1.2 kb 16S rRNA sequences of other strains of ‘*Christensenellales*’ (Parks et al., 2018) available in the GenBank non-redundant (nr) database (as of 21/12/2020), using MUSCLE (default settings) (Edgar, 2004). Full-length 16S rRNA gene sequences were also extracted from complete genome assemblies of additional R-7 group strains FE2010, FE2011 and XBB3002 (Mahoney-Kurpe et al., 2021). A maximum-likelihood tree was built using the default settings in MEGA X with 500 bootstrap replicates. The resulting tree (Figure 1) shows that both R-7<sup>T</sup> and WTE2008<sup>T</sup> fell within a distinct cluster with other R-7 group strains (Mahoney-Kurpe et al., 2021), all sharing >97% sequence similarity. However, beyond this cluster, the sequence of the closest type-strain, *Christensenella minuta* YIT 12065<sup>T</sup> (Morotomi et al., 2012), showed <85.4% identity to both R-7<sup>T</sup> and WTE2008<sup>T</sup> sequences.



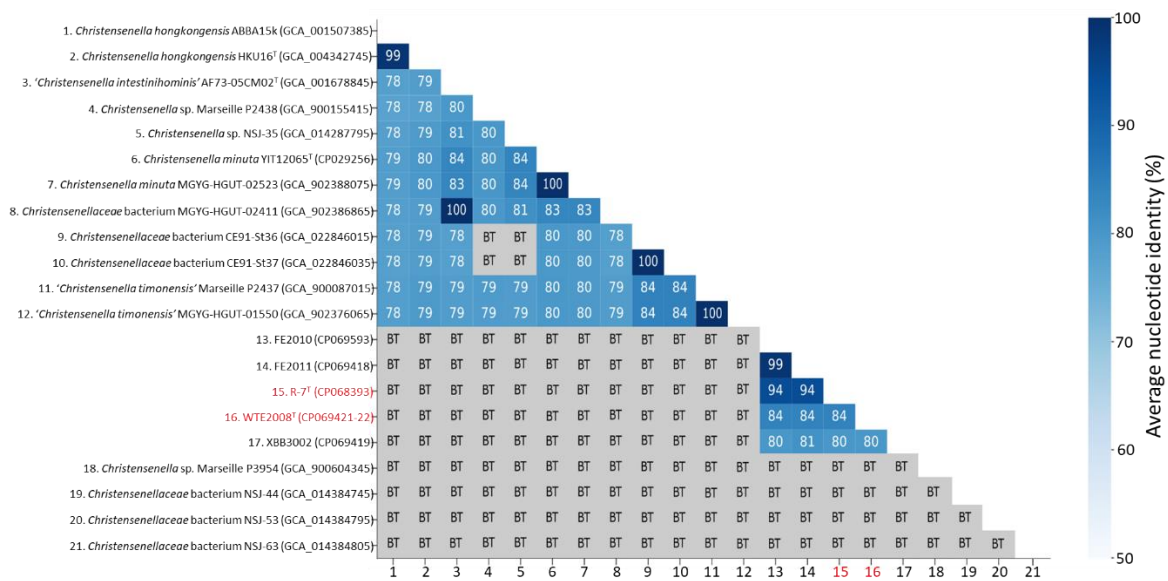
**Figure 3.1. Maximum-likelihood phylogeny based on 16S rRNA gene sequences from members of the order 'Christensenellales'.** Available 'Christensenellales' 16S rRNA gene sequences >1.2 kb were used for the phylogeny, with GenBank accession numbers given in brackets. The sequence from *Ruminococcus albus* 7<sup>T</sup> was used as an outgroup. The three sequences in each of the genomes of R-7<sup>T</sup> and WTE2008<sup>T</sup> are shown in red. The bootstrap values below the branches are shown if >60 as a percentage of 500 bootstrap replications.

### 3.3.5 Genome features

Draft genome assemblies of R-7<sup>T</sup> and WTE2008<sup>T</sup> were generated as part of the Hungate1000 project (Seshadri et al., 2018), and they exhibited an average nucleotide identity (ANI) of 84% to each other, in contrast to their highly similar 16S rRNA gene sequence identities. This unexpected observation was confirmed by generating the complete genome sequences of R-7<sup>T</sup> and WTE2008<sup>T</sup> using methods as described previously (Mahoney-Kurpe et al., 2021). The genome characteristics of the closed R-7<sup>T</sup> and WTE2008<sup>T</sup> genomes were near-identical to those of the draft assemblies (Seshadri et al., 2018), and were similar to those of the R-7 group strains XBB3002, FE2010 and FE2011 (Mahoney-Kurpe et al., 2021) (Appendix 7.2). However, WTE2008<sup>T</sup> also possessed a ca. 44 kb linear extrachromosomal element. Alignment of this fragment against the PHASTER phage database (Arndt et al., 2016) found no homology to known phage sequences, nor were any known replication origins detected using the DoriC database (Luo & Gao, 2019).

The Genome Taxonomy Database framework (release 05-RS95) (Parks et al., 2018) had classified the draft genomes (Seshadri et al., 2018) of R-7<sup>T</sup> (sp900199385) and WTE2008<sup>T</sup> (sp900176495) as separate species belonging to an unclassified family (CAG-74) and genus (GCA-900199385). These designations were confirmed using the complete genome sequences, using GTDB-Tk (Chaumeil et al., 2020). The genus GCA-900199385 also contained XBB3002, FE2010 and FE2011 (Mahoney-Kurpe et al., 2021).

To compare the genomes of R-7<sup>T</sup> and WTE2008<sup>T</sup> to those of other members of '*Christensenellales*', all available genome data in GenBank (as of the 22/12/2020) from strains of the order '*Christensenellales*' were used in pairwise ANI analyses (Figure 3.2). ANI alignments were carried out using fastANI under the default settings (Jain et al., 2018). R-7<sup>T</sup> shared 94% ANI with FE2010 and FE2011, whereas there were no matches greater than 84% to WTE2008<sup>T</sup>. This highlights the genetic distinctiveness of the R-7 group strains within the order '*Christensenellales*', and of strain R-7<sup>T</sup> and WTE2008<sup>T</sup> from each other and other R-7 group strains.



**Figure 3.2. Matrix of pairwise average nucleotide identity (ANI) values between the genomes of members of 'Christensenellales'.** Values in each box represent the ANI result for each comparison, expressed as percent identity. Comparisons showing BT (below threshold) had alignment fractions less than the 20% cutoff using the default fastANI setting and were therefore not computed.

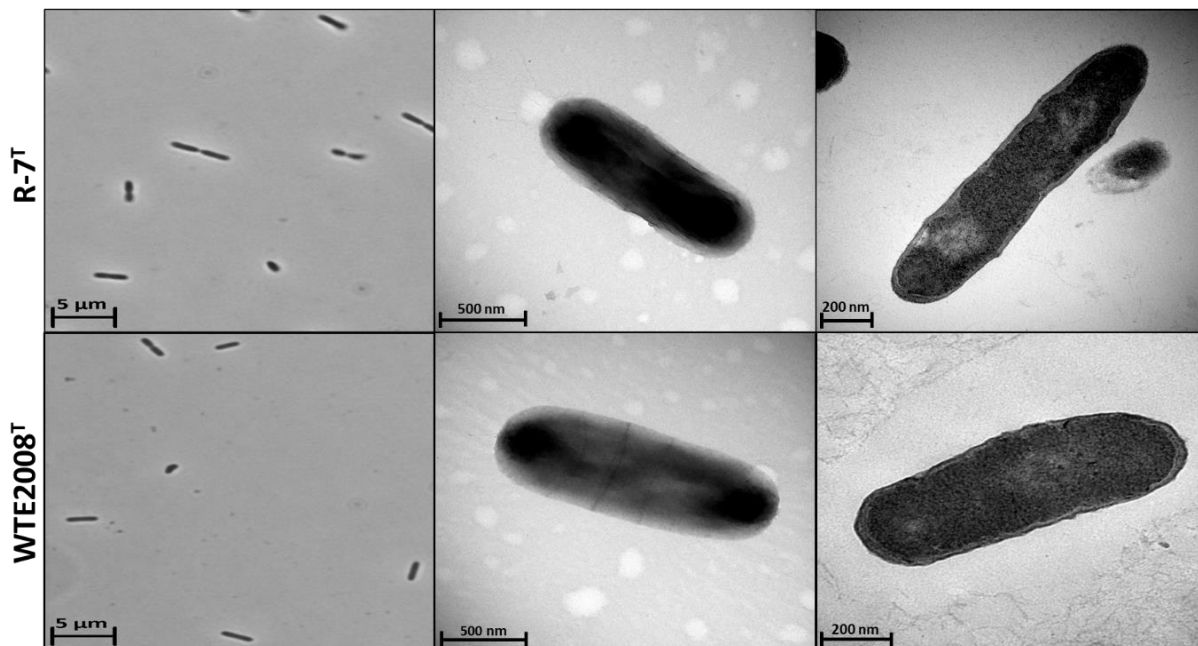
### 3.3.6 Physiology and Chemotaxonomy

Cell morphology was assessed using phase-contrast microscopy and transmission electron microscopy. In both instances, visualised cells were grown anaerobically in BY medium (Joblin, 2005) containing rumen fluid from a hay-fed cow, with 0.5% (w/v) cellobiose overnight at 39°C. For phase-contrast microscopy, cells were photographed with a Leica DM2500 microscope (Leica Microsystems, Wetzlar, Germany), and cells of both strains were 1-3 µm in length and 0.2-0.5 µm in width, often in pairs (Figure 3.3), or occasionally forming longer chains. Cells were also assessed using the Gram staining method (Coico, 2006) and both strains were Gram negative.

Transmission electron microscopy (TEM) of negatively stained cells was performed by pelleting cells by low-speed centrifugation (2800 × g), and resuspending in sterile water. Cell suspensions were processed and imaged at the Massey Microscopy and Imaging Centre (MMIC, Palmerston North, New Zealand). Cells were fixed to a Formvar grid and stained in 2% (w/v) uranyl acetate, and then viewed using a FEI Tecnai G2 Biotwin transmission electron

microscope. Cells of R-7<sup>T</sup> and WTE2008<sup>T</sup> appeared similar, lacking any obvious appendages (e.g. pili or flagella), and exhibited electron-dense poles. Both R-7<sup>T</sup> and WTE2008<sup>T</sup> had thin electron-dense rings around the mid-section of cell bodies (Figure 3.3).

Thin cross-sections of cells were prepared for TEM by washing cell pellets three times in sterile water, then resuspending in modified Karnovsky's fixative (2% paraformaldehyde and 3% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4)) for processing and imaging (MMIC). Resin-embedded thin sections of each sample were prepared using an EM UC7 ultra-microtome (Leica Microsystems, Wetzlar, Germany), and viewed using a FEI Tecnai G2 Biotwin transmission electron microscope. Both R-7<sup>T</sup> and WTE2008<sup>T</sup> had cell wall ultra-structures characteristic of Gram negative bacteria (Figure 3.3).



**Figure 3.3. Cell morphology of R-7<sup>T</sup> and WTE2008<sup>T</sup>.** Phase-contrast images (left), transmission electron microscopy (TEM) images of negatively stained cells (centre), and thin-sectioned cells (right). Scale bars are shown bottom left.

To assess the cellular fatty acid compositions of strains, cultures were anaerobically grown overnight at 39°C in 100 mL of BY medium (Joblin, 2005) containing 0.5% (w/v) cellobiose and harvested by centrifugation. Cellular fatty acid profiles (Appendix 7.3) were determined by



gas chromatography with flame-ionization detection using the MIDI Sherlock Microbial Identification System (MIS), and the Anaerobic Bacteria Library (MOORE6) for peak identification. The five most abundant cellular fatty acids of R-7<sup>T</sup> and WTE2008<sup>T</sup> were C<sub>16:0</sub>, C<sub>16:0</sub> iso, C<sub>17:0</sub> anteiso, and C<sub>18:0</sub>, and C<sub>15:0</sub> anteiso. However, C<sub>17:0</sub> anteiso and C<sub>16:0</sub> iso were more abundant in R-7<sup>T</sup>, whereas C<sub>16:0</sub> and C<sub>18:0</sub> were more abundant in WTE2008<sup>T</sup> (Table 3.2).

The growth of strains on 23 different soluble carbon substrates in BY medium (Joblin, 2005) was assessed in triplicate using a Spectronic 200 spectrophotometer (Thermo Scientific) with absorbance at 600 nm. Each strain was inoculated into anaerobic medium containing 0.5% (w/v) of the test substrate and incubated at 39°C for 48 hours. Both R-7<sup>T</sup> and WTE2008<sup>T</sup> grew well using L-arabinose, galactose, glucose, xylose, cellobiose, lactose, maltose, melibiose, sucrose, and trehalose as a carbon source. Weak growth of both strains was observed on melezitose, raffinose, and esculin. WTE2008<sup>T</sup> grew on fructose and mannose, but R-7<sup>T</sup> did not. R-7<sup>T</sup> instead grew weakly on rhamnose, whereas WTE2008<sup>T</sup> did not. Neither strain grew on ribose or any of the sugar-alcohols tested. Both strains grew best on cellobiose, however, WTE2008<sup>T</sup> grew to approximately half the optical density of R-7<sup>T</sup>.

**Table 3.2. Phenotypic characteristics of R-7<sup>T</sup> and WTE2008<sup>T</sup> compared to other related type strains.** Taxa: 1, R-7<sup>T</sup>; 2, WTE2008<sup>T</sup>; *Christensenella minuta* YIT-12065<sup>T</sup> (Morotomi et al., 2012); 4, *Christensenella hongkongensis* HKU16<sup>T</sup> (Lau et al., 2007); 5, '*Christensenella intestinhominis*' AF73-05CM02 (Zou et al., 2021).

Characteristic	1	2	3 <sup>1</sup>	4 <sup>1</sup>	5 <sup>1</sup>
<b>Isolation source</b>	Sheep rumen	Cow rumen	Human faeces	Human blood	Human faeces
<b>Gram-stain</b>	-	-	+/- <sup>2</sup>	+	-
<b>Morphology</b>	rods	rods	rods	coccobacilli/short rods	rods
<b>Cell length, width (µm)</b>	1-3, 0.2-0.5	1-3, 0.2-0.5	0.8-1.9, 0.4	0.7-1.1, 0.4-0.5	1-2, 0.5
<b>Motility</b>	-	-	-	+	-
<b>Major (&gt;5 %) cellular fatty acids<sup>3</sup></b>	C <sub>17:0</sub> anteiso (19.7) C <sub>16:0</sub> iso (18.2) C <sub>16:0</sub> (11.5) C <sub>18:0</sub> (9.1) C <sub>15:0</sub> anteiso (8.9) C <sub>18:0</sub> iso (6.6)	C <sub>18:0</sub> (19.3) C <sub>16:0</sub> (15.2) C <sub>17:0</sub> anteiso (8.7) C <sub>16:0</sub> iso (8.4) C <sub>15:0</sub> anteiso (7.7)	C <sub>15:0</sub> iso (37.8) C <sub>16:0</sub> (31.7) C <sub>14:0</sub> (14.8)	n/a	C <sub>14:0</sub> (46.6) C <sub>16:0</sub> (9.7) C <sub>10:0</sub> (7.5) Iso-C <sub>15:0</sub> (7.4) C <sub>12:0</sub> (7.2) C <sub>18:1 ω9c</sub> (6.9) Iso-C <sub>11:0</sub> (5.6)
<b>Growth on soluble substrates<sup>4</sup></b>					
<b>Arabinose</b>	+	+	+	+	+
<b>Fructose</b>	-	+	n/a	-	+
<b>Galactose</b>	+	(+)	n/a	-	+
<b>Glucose</b>	+	+	+	+	+
<b>Mannose</b>	-	(+)	(+)	+	(+)
<b>Rhamnose</b>	(+)	-	+	-	+
<b>Ribose</b>	-	-	n/a	-	+
<b>Xylose</b>	+	+	+	+	+
<b>Cellobiose</b>	+	+	-	-	-
<b>Lactose</b>	+	+	-	-	-
<b>Maltose</b>	+	+	-	-	(+)
<b>Melibiose</b>	+	+	n/a	-	-
<b>Sucrose</b>	+	+	-	-	+
<b>Melezitose</b>	(+)	(+)	-	-	(+)
<b>Raffinose</b>	+	+	-	-	(+)
<b>Trehalose</b>	+	+	-	-	-
<b>Sugar-alcohols<sup>5</sup></b>	-	-	- <sup>6</sup>	+/- <sup>7</sup>	+/- <sup>8</sup>
<b>Amygdalin</b>	-	-	n/a	n/a	-
<b>Esculin</b>	(+)	(+)	n/a	n/a	+
<b>Growth on insoluble substrates<sup>9</sup></b>					
<b>Crystalline cellulose</b>	-	-	n/a	n/a	n/a
<b>Dextrin</b>	+	+	n/a	n/a	n/a
<b>Glycogen</b>	-	-	n/a	n/a	-
<b>Inulin</b>	-	-	n/a	n/a	n/a
<b>Pectin</b>	+	+	n/a	n/a	n/a
<b>Starch</b>	+	+	n/a	n/a	n/a
<b>Xylan</b>	+	+	n/a	n/a	n/a
<b>Rutin</b>	-	-	n/a	n/a	n/a
<b>Salicin</b>	+	+	+	n/a	+
<b>Fermentation end products<sup>10</sup></b>	A, L, H, E	A, L, H, E	A, B <sup>11</sup>	n/a <sup>11</sup>	A, F, B, L <sup>11</sup>

<sup>1</sup> Values showing n/a were not assessed in previous reports of these strains.

<sup>2</sup> Described originally as Gram-negative (Morotomi et al., 2012), but later reports have shown Gram-positive staining of cells (Alonso et al., 2017; Liu et al., 2021).

<sup>3</sup> Listed in order of abundance, with abundances (%) in brackets beside each cellular fatty acid. Full profiles are shown in Appendix 7.3.

<sup>4</sup> Growth was measured by light absorbance at 600 nm. + = Growth ( $OD_{600nm} > 0.2$ ); (+) = Poor but detectable growth ( $0.1 < OD_{600nm} < 0.2$ ); - = no growth ( $OD_{600nm} < 0.1$ ). In all instances, values had measurements taken after inoculation as well as those from triplicate no-substrate controls deducted.

<sup>5</sup> Glycerol, myo-inositol, mannitol, sorbitol, and xylitol.

<sup>6</sup> Growth on myo-inositol and xylitol was not tested.

<sup>7</sup> Growth was observed on glycerol, with no growth on other sugar alcohols tested.

<sup>8</sup> Growth was observed on sorbitol and xylitol, with no growth on other sugar alcohols tested.

<sup>9</sup> Growth was assessed by fermentation end product formation. + = Production of  $> 1$  mM ethanol.

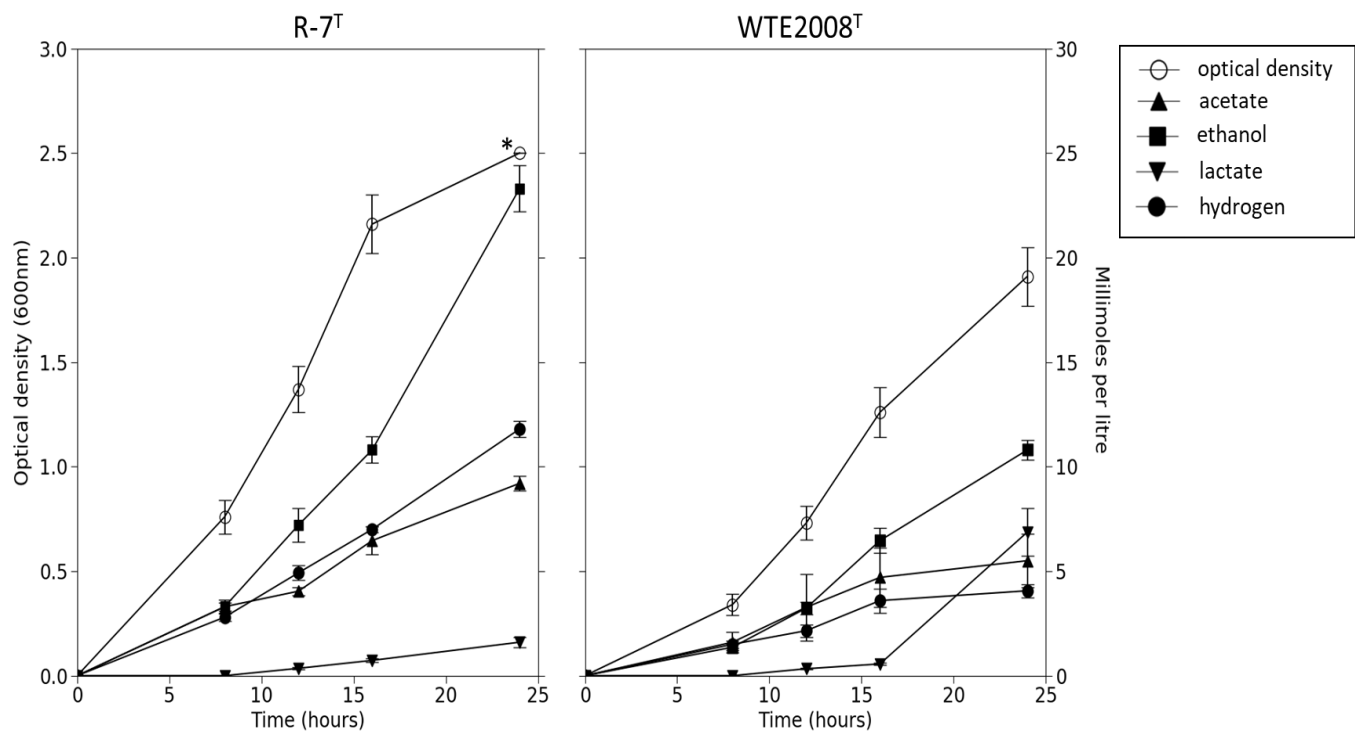
<sup>10</sup> A= acetate, L= lactate, H= hydrogen, E= ethanol, B=butyrate, F=formate.

<sup>11</sup> These studies do not report testing for the production of alcohols.

The ability of strains to degrade a range of insoluble substrates, including major plant cell wall polysaccharides, was assessed. Each strain was incubated anaerobically in BY medium (Joblin, 2005) with 0.5% (w/v) of each insoluble substrate in Hungate tubes fastened horizontally on a shaker, and gently shaken at 39°C for five days. Fermentation end products were quantified by gas chromatography as previously described (Palevich et al., 2019). Overall, short-chain fatty acid (SCFA) production was highly variable between triplicate cultures for some strain-substrate combinations, and in some instances showed net decreases in concentrations, which suggested that SCFA utilisation may have occurred (Appendix 7.4). R-7<sup>T</sup> and WTE2008<sup>T</sup> produced acetate, but not butyrate or propionate. Ethanol was also produced by both strains, with generally less variability between triplicates than SCFAs. Using ethanol production as an indicator of microbial growth, R-7<sup>T</sup> and WTE2008<sup>T</sup> could degrade and ferment xylan and pectin, suggestive of a role in ruminal fibre degradation. However, neither strain could grow on crystalline cellulose (Table 3.2), despite R-7<sup>T</sup> having been originally isolated from a cellulose-adherent fraction of the rumen microbiota. Notably, R-7<sup>T</sup> cultures yielded greater ethanol concentrations growing on dextrin and salicin than they did on cellobiose (Appendix 7.4).

To further characterise fermentation end product formation during growth of R-7<sup>T</sup> and WTE2008<sup>T</sup>, 10 mL cultures of each strain were grown in BY medium (Joblin, 2005) containing 0.5% (w/v) cellobiose, rumen fluid from a pasture-fed sheep, and 0.1% (w/v) bacto-peptone and casamino acids at 39°C for 24 h. Samples were taken at various time points to determine fermentation end product concentrations. Production of hydrogen was also measured by injecting 0.1 mL of headspace of each culture into an Aerograph 660 gas chromatograph

(Varian Associates, PaloAlto, CA, USA) fitted with a Porapak Q80/100 mesh column (WatersCorporation, Milford, MA, USA) and a thermal conductivity detector. To assess the production of formate, succinate and lactate, sample supernatants were derivatised using the method described by Richardson et al. (1989). Optical densities and production of fermentation end products over time are shown in Figure 3.4. Both strains produced acetate, ethanol, hydrogen and lactate. However, R-7<sup>T</sup> produced more hydrogen than lactate ( $11.8 \pm 0.7$  millimoles per litre hydrogen versus  $1.5 \pm 0.4$  mM lactate), whereas WTE2008<sup>T</sup> produced considerably more lactate and less hydrogen ( $4.1 \pm 0.5$  millimoles per litre hydrogen, versus  $6.5 \pm 2$  mM lactate). Notable was the apparent switch to copious lactate production and cessation of hydrogen production by WTE2008<sup>T</sup> after 18 h of growth.



**Figure 3.4. Optical density and production of fermentation end products of R-7<sup>T</sup> and WTE2008<sup>T</sup> over 24 hours of growth.** Error-bars denote SEM (n = 3). \*This data point represents the maximum absorbance limit of the spectrophotometer.

### 3.3.7 Protologue

#### 3.3.7.1 Description of *Aristaeellaceae* fam. nov.

*Aristaeellaceae* (A.ris.tae.el.la.ce'ae. N.L. fem. n. *Aristaeella* type genus of the family; *-aceae* ending to denote a family; N.L. fem. pl. n. *Aristaeellaceae* the family whose nomenclatural type is the genus *Aristaeella*).

The family is described on the basis of phylogenetic analyses of 16S rRNA gene sequences, and whole genome analyses. Cells are rod-shaped, Gram-negative and anaerobic. Belongs to the order 'Christensenellales' (Parks et al., 2018) of the phylum *Bacillota* (Oren & Garrity, 2021).

#### 3.3.7.2 Description of *Aristaeella* gen. nov.

*Aristaeella* (A.ris.tae.el'la. Gr. masc. n. *Aristaeus*; *-ella* L dim.; N.L. fem. n. *Aristaeella* named after Aristaeus (Aristaios), a Greek god associated with animal husbandry and production of alcoholic beverages).

Cells are Gram negative, obligately anaerobic, non-motile rods that do not form spores. Utilise various sugars, hemicellulose and pectin to form short-chain fatty acids, hydrogen and ethanol. The type species is *Aristaeella hokkaidonensis*.

#### 3.3.7.3 Description of *Aristaeella hokkaidonensis* sp. nov.

*Aristaeella hokkaidonensis* (hok.kai.do.nen'sis. N.L. fem. adj. *hokkaidonensis* pertaining to Hokkaido, Japan, where the type strain (R-7<sup>T</sup>) was isolated).

Cells are Gram negative, obligately anaerobic non-motile rods ranging from approximately 1-3 µm in length and 0.2-0.5 µm in width. Cells possess no obvious appendages and have cell walls characteristic of Gram-negative bacteria. Dominant cellular fatty acids are C<sub>15:0</sub> anteiso, C<sub>16:0</sub>, C<sub>16:0</sub> iso, C<sub>17:0</sub> anteiso, C<sub>18:0</sub>, and C<sub>18:0</sub> iso. Cells use arabinose, galactose, glucose, xylose,

cellobiose, lactose, maltose, melibiose, sucrose, and trehalose, with weak growth on rhamnose, melezitose, raffinose and esculin, and no growth on fructose, mannose, ribose, glycerol, myo-inositol, sorbitol, xylitol, and amygdalin. Cells degraded and used the breakdown products of dextrin, pectin, starch, xylan, and salicin, but could not degrade crystalline cellulose, glycogen, inulin, and rutin. The major fermentation end products were ethanol, hydrogen, acetate, and some lactate.

The type strain, R-7<sup>T</sup>, was isolated from the sheep rumen in Hokkaido, Japan. The genome was determined to be 3.39 Mb with a G+C content of 53%. The complete genome is deposited in GenBank under the accession CP068393. Raw reads are deposited in the Sequence Read Archive under the accessions SRR15429008 (ONT) and SRR15429007 (MGISEQ).

#### 3.3.7.4 Description of *Aristaeella lactis* sp. nov.

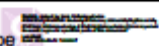
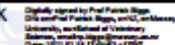
*Aristaeella lactis* (lac'tis. L. neut. n. *lac* milk, L. gen. n. *lactis* of milk, pertaining to lactate, due to the significant lactate production by the type strain WTE2008<sup>T</sup>).

Cells are Gram negative, obligately anaerobic non-motile rods ranging from approximately 1-3 µm in length and 0.2-0.5 µm in width. Cells possess no obvious appendages and have cell walls characteristic of Gram-negative bacteria. Dominant cellular fatty acids are C<sub>15:0</sub> anteiso, C<sub>16:0</sub>, C<sub>16:0</sub> iso, C<sub>17:0</sub> anteiso, and C<sub>18:0</sub>. Cells use arabinose, fructose, glucose, xylose, cellobiose, lactose, maltose, melibiose, sucrose and trehalose, with weak growth on galactose, mannose, melezitose, raffinose and esculin, and displayed no growth on amygdalin, glycerol, myo-inositol, mannitol, rhamnose, ribose, sorbitol, or xylitol. Cells degraded and used the breakdown products of dextrin, pectin, starch, xylan, and salicin, but could not degrade crystalline cellulose, glycogen, inulin, or rutin. The major fermentation end products produced were ethanol, lactate, acetate, and some hydrogen.

The type strain is WTE2008<sup>T</sup>, which was isolated from the bovine rumen. The genome of the type strain is characterised by a size of 3.45 Mb and G+C content of 53.5 mol%. The complete genome sequence of the type strain is available under the GenBank accessions CP069421-22. Raw reads are deposited in the Sequence Read Archive under the accessions SRR15428981 (ONT) and SRR15428980 (MGISEQ).

## STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sam Mahoney-Kurpe
Name/title of Primary Supervisor:	Prof. Patrick Biggs
In which chapter is the manuscript /published work:	Chapter 3, Section 3.3
<p>Please select one of the following three options:</p> <p><input type="radio"/> The manuscript/published work is published or in press</p> <ul style="list-style-type: none"> <li>• Please provide the full reference of the Research Output:</li> </ul> <p><input type="radio"/> The manuscript is currently under review for publication – please indicate:</p> <ul style="list-style-type: none"> <li>• The name of the journal: International Journal of Systematic and Evolutionary Microbiology (IJSEM)</li> <li>• The percentage of the manuscript/published work that was contributed by the candidate: 85.00</li> <li>• Describe the contribution that the candidate has made to the manuscript/published work: Conceptualisation, Investigation, Formal analysis, Visualisation, Writing- original draft.</li> </ul> <p><input checked="" type="radio"/> It is intended that the manuscript will be published, but it has not yet been submitted to a journal</p>	
Candidate's Signature:	Sam Mahoney-Kurpe 
Date:	18-Feb-2022
Primary Supervisor's Signature:	Prof Patrick Biggs 
Date:	18-Feb-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/publication or collected as an appendix at the end of the thesis.

## 3.4 Discussion and Conclusion

As shown by cultivation-independent studies of rumen microbial community composition (Cristobal-Carballo et al., 2021; Henderson et al., 2015), the R-7 group represents an abundant group of rumen bacteria that to date had remained unstudied. At the beginning of this project, the draft genomes of strains R-7 and WTE2008 were available as part of the Hungate1000 collection (Seshadri et al., 2018), though neither strain had been characterised *in vitro*, and hence remained unclassified beyond the family level (Henderson et al., 2019). The goal of this work was to therefore characterise this group, and in doing so further our understanding of their potential metabolic roles in the rumen microbiota. We firstly added to the number of genome sequences available for the R-7 group, by generating the complete genome sequences of three previously isolated (Noel, 2013) strains of the R-7 group (Mahoney-Kurpe et al., 2021), and also generated complete assemblies of R-7 and WTE2008. Phylogenetic analyses of these R-7 group genomes along with other available genome sequences of the *Christensenellaceae* family were then conducted, in conjunction with extensive phenotypic characterisation of the two R-7 group strains that were already part of the Hungate1000 collection. The composite results of these analyses allowed the proposal of R-7 and WTE2008 as type strains of the first two species (*Aristaeella hokkaidonensis* sp. nov. and *Aristaeella lactis* sp. nov., respectively) of a novel proposed bacterial genus (*Aristaeella* gen. nov.), and family (*Aristaeellaceae* fam. nov.) (Section 3.3.1).

Based on their 16S rRNA phylogeny, both R-7 and WTE2008 had previously been classified as belonging to the *Christensenellaceae* family (Henderson et al., 2019). Given that both strains share 98.7% 16S rRNA identity to each other, it was anticipated that the strains would be classified as belonging to a single species of a novel genus of the *Christensenellaceae* family. However, phylogenetic analyses revealed a surprisingly low genome-wide average nucleotide identity (ANI) value of 84% shown between these strains. Indeed, their classification using the Genome Taxonomy Database (GTDB) (Parks et al., 2018) confirmed their assignment as separate species of a novel genus and family of the recently proposed (Parks et al., 2018) order '*Christensenellales*'. Our additional genome sequencing of FE2010, FE2011 and XBB3002 further added to the robustness of the phylogenomic analyses conducted, and



showed that of the five R-7 group strains that all shared > 97% full-length 16S rRNA identity, based on the commonly used species-level ANI threshold of 95% (Jain et al., 2018), these were instead assigned to four different species-level groups. This finding highlights the importance of using genome-wide metrics to assign taxonomy, and that while ANI-based metrics may often correlate well with assignments made using 16S rRNA phylogeny (Jain et al., 2018), this does not appear to be the case for this bacterial group. As more isolated strains of this group continue to undergo characterisation and taxonomic classification, it will be interesting to see the extent to which their genomic diversity manifests at the phenotype level.

Results of the characterisation of *A. hokkaidonensis* and *A. lactis* provided insights into the potential roles of this group in ruminal metabolism. Based on its genome sequence, as well as strain R-7 having been isolated from the fibre-adherent fraction of the rumen microbiota (section 3.3.3), it was suspected that this group may contribute to fibre degradation. Characterisation of these strains found that both strains could indeed degrade hemicellulose and pectin; however, neither could degrade cellulose. Promotion of this group could therefore facilitate fibre degradation, and could thus be of particular interest in agricultural systems in New Zealand where animals predominantly graze pastures all year round.

Another notable characteristic of R-7 and WTE2008 was their copious production of ethanol and hydrogen as major fermentation end products. Both R-7 and WTE2008 produced large amounts of ethanol, to even greater concentrations than they produced acetate. Both species also produced hydrogen, particularly R-7, whereas WTE2008 instead produced more lactate. Both ethanol and hydrogen are known substrates of rumen methanogens (Greening et al., 2019). As such, despite the R-7 strains demonstrating the desirable ability of degrading some plant cell wall polysaccharides, this may potentially occur at the cost of greater methane formation. Nonetheless, given the variability seen in the production of hydrogen between R-7 and WTE2008, further characterisation may reveal strains of this group that efficiently degrade xylans and pectins to produce larger amounts of acetate and lactate, and less ethanol and hydrogen.

In conclusion, in this work we have characterised and formally named the first two species of the R-7 group, and expanded the number of genome sequences of R-7 group strains currently available. These outcomes provide a foundation for future work, which will continue to contribute towards a greater understanding of the ecological roles of this abundant but long overlooked group of rumen bacteria.

4 Unravelling the regulation of cobalamin-dependent  
propionate production by the rumen bacterium  
*Prevotella ruminicola*

## 4.1 Introduction

Given the importance of ruminal propionate in host metabolism, and its relationship with ruminant product quality and methane production, there is particular interest in promoting propionate fermentation of forage-fed ruminants, for which propionate levels are generally lower than those of animals fed concentrate-based diets (Zhang et al., 2017). Therefore, deep metatranscriptome datasets from sheep fed a forage-based diet were bioinformatically mined for reads associated with propionate production (Section 4.2) to identify prominent microbial groups in forage-fed ruminants involved in propionate production.

The results of the metatranscriptome analyses led to the selection of rumen *Prevotella 1* strains for further characterisation (Section 4.3). To better characterise the propionate production pathway in *Prevotella 1*, a selection of 14 strains spanning the phylogeny of *Prevotella 1* in the Hungate1000 collection were selected for inclusion in a phenotypic screen assessing growth and short-chain fatty acid production of strains in the presence and absence of cobalamin (vitamin B<sub>12</sub>), given the cobalamin-dependent propionate production of the type strain *P. ruminicola* 23 (Strobel, 1992). A complete genome assembly of one selected representative member of *Prevotella* was generated to serve as a reference for comparative genomics against available *Prevotella* draft genomes from the Hungate1000 study (Seshadri et al., 2018). To characterise the regulation of cobalamin-induced propionate production in ruminal *Prevotella*, and assess the potential involvement of cobalamin riboswitches, a comparative transcriptomics and proteomics approach was used in which the genome-sequenced KHP1 strain was grown in the presence and absence of cobalamin. A manuscript describing this work is currently in preparation for submission to mSystems (American Society for Microbiology).

## 4.2 Bioinformatic identification of dominant ruminal propionate producing bacteria

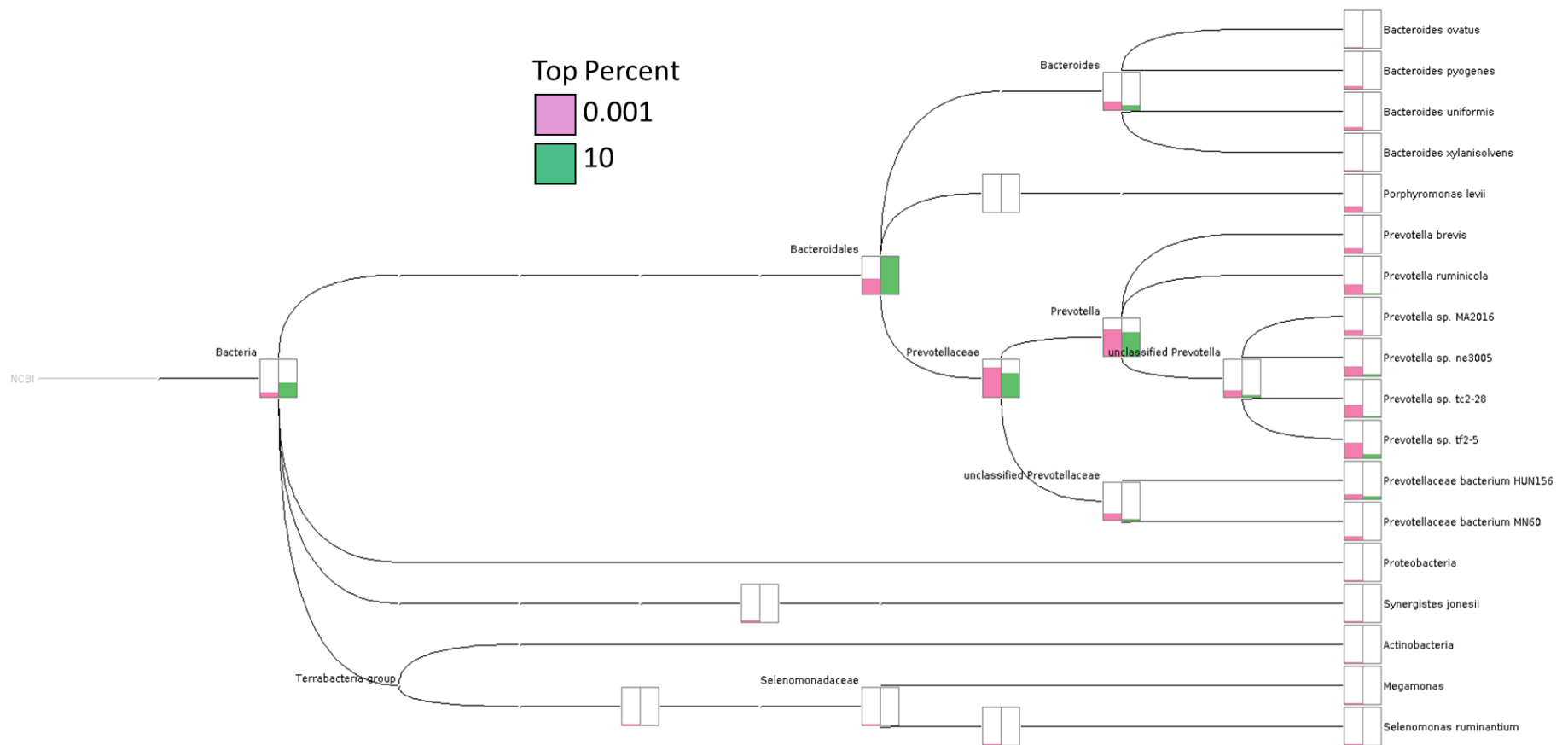
The exponential advances in sequencing and computational capabilities have enabled cultivation-independent methods of characterising microbial communities at a far greater taxonomic resolution than ever before (McCann et al., 2014). In the context of the rumen, the recent genome sequencing of many rumen microorganisms (Seshadri et al., 2018) provides a valuable resource that greatly enhances the ability to taxonomically assign rumen metagenomic and metatranscriptomic reads, enabling insights into which taxa are involved in important rumen metabolic processes in these samples. To identify the dominant microbial taxa involved in propionate production from succinate, a bioinformatic approach was used in which deeply sequenced rumen metatranscriptomic datasets were mined for reads predicted to encode enzymes involved in propionate production, to identify the likely taxonomic groups responsible. This section gives an overview of this analysis, which formed the basis of the selection of strains to characterise further (Section 4.3). The reader is referred to sections 2.2.8.3 and 2.2.8.4 for details of the methods used in this section.

There are two main approaches to metagenome/metatranscriptome analyses, referred to as assembly-based, and read-based (Shakya et al., 2019). In an assembly-based approach, the metagenomic/metatranscriptomic data are firstly assembled into larger scaffolds that are annotated for genes, and then aligned against a reference database. In a read-based approach, the sequenced short reads are instead aligned directly to a reference database without any prior assembly. Given the expansive catalogue of reference genome sequences now available of rumen microorganisms (Seshadri et al., 2018) as well as the thousands of metagenome assembled genomes (MAGs) reconstructed from rumen samples (Stewart et al., 2019), a read-based approach was chosen for use in this study.

Rather than aligning reads against all-encompassing databases such as the NCBI non-redundant (nr) or UniProt (The UniProt Consortium, 2021) databases, the building and use of environment-specific databases enhances the taxonomic resolution of read assignment

(Ritari et al., 2015; Seedorf et al., 2014). As the focus of this study was to select organisms from the Hungate1000 collection for characterisation, a database was built based on the amino acid sequences of all 501 genome sequences of the Hungate1000 collection (see section 2.2.8.3) to align the short reads of the metatranscriptome datasets against.

The metagenome analyser (MEGAN) software is a common reference-based metagenome/metatranscriptome pipeline which readily allows for the extraction of reads based on both their predicted function and taxonomy, and was therefore selected for use in this study. By default, MEGAN assigns taxonomic classification to sequence reads using a lowest common ancestor (LCA) algorithm, which assigns the lowest taxonomic rank shared between the top 10% significant hits (Huson et al., 2016). However, to identify candidate isolates of the Hungate1000 collection to select for characterisation, it was desired to classify reads at the highest taxonomic resolution possible, given that classifications at the family or even genus level in some instances still encompassed a large number of candidate Hungate1000 strains. To facilitate this, the 'Top Percent' parameter was changed to 0.001, which caused reads to be assigned to a taxonomy based only the most significant hit, or in the case of a tie, the lowest common ancestor of the top hits. Using this parameter setting greatly shifted read assignments towards lower taxonomic ranks (Figure 4.1), and was therefore used in all analyses of this study.



**Figure 4.1. Alteration of the 'Top Percent' parameter enhances the specificity of taxonomic read assignment in MEGAN.** The MEGAN tree shows a comparison between the assignment of methylmalonyl-CoA mutase (K01847)-encoding reads from a single file. The trend seen between each parameter was representative of that observed in all files for all extracted KEGG/InterPro2Go categories.

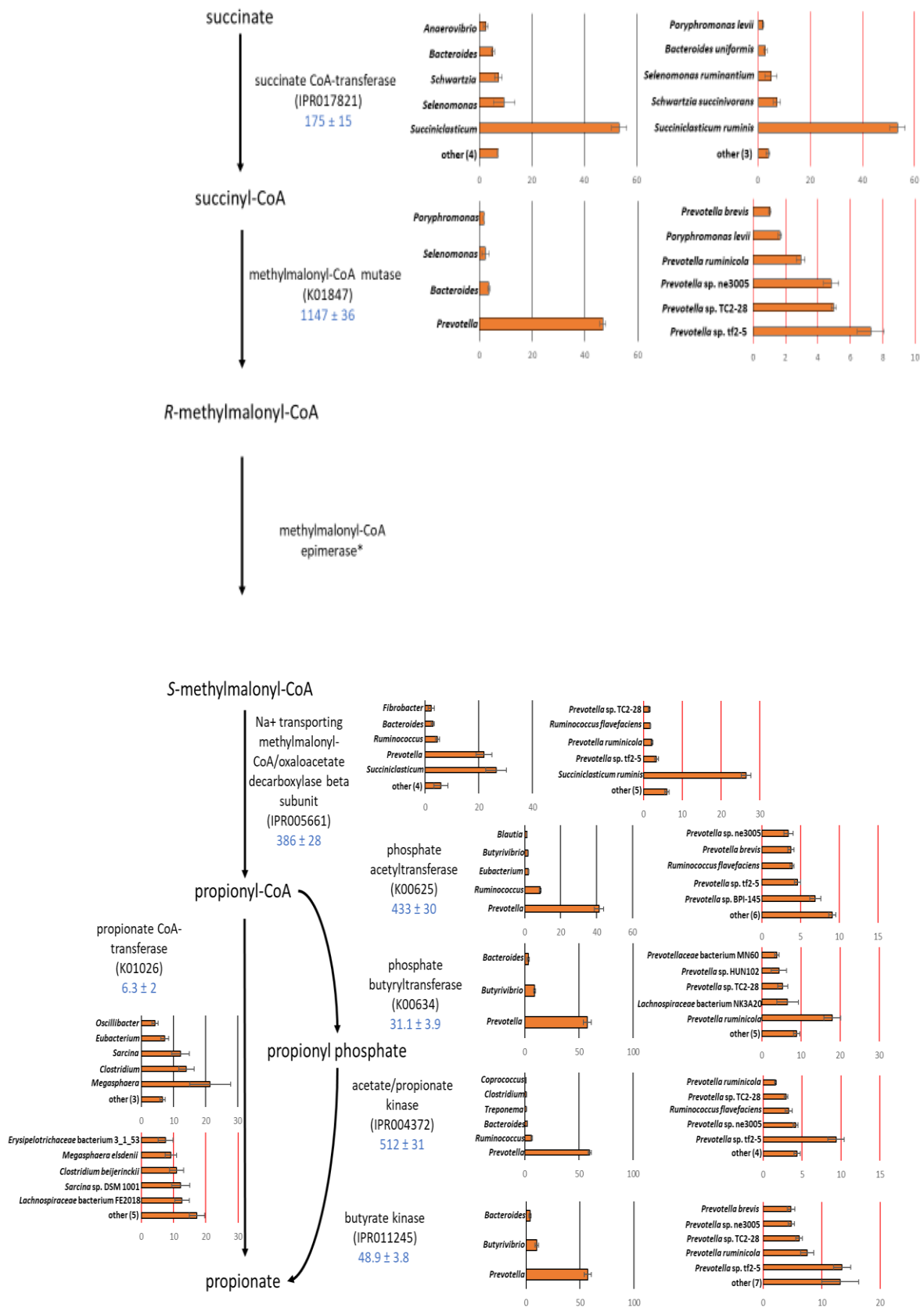
#### 4.2.1 Dominance of *Prevotella* and *Succiniclasicum* in the expression of enzymes predicted to convert succinate to propionate in metatranscriptome datasets

To determine the dominant taxa involved in propionate production via the succinate pathway in the metatranscriptome datasets from Shi et al. (2014), KEGG orthology/Interpro2GO categories corresponding to enzymes putatively involved in the pathway were extracted from each sample, and taxa profiles of the resulting reads were analysed in MEGAN. This study sampled the rumen of ten sheep of differing methane yield phenotype (4 high, 4 low and 2 intermediate methane-emitters), with each animal sampled at two different time points (Shi et al., 2014). The sequence files from the samples obtained from each animal at each time point were concatenated together and analysed as single files, to allow the consistency of observed effects across biological replicates to be assessed irrespective of sampling time. A selection of KEGG/InterPro2GO categories covering the corresponding enzymatic steps from succinate to propionate were chosen for inclusion in this analysis, aside from categories corresponding to methylmalonyl-CoA epimerase, which were absent from the datasets (see Table 2.2). Due to the cross-reactivity of different short-chain fatty acid producing transferases and kinases (Louis and Flint, 2017), some functional categories annotated as being involved in acetate or butyrate production for these steps were also included.

Relative abundances of different taxonomic groups assigned to each KEGG orthology/Interpro2GO category of interest were calculated at both the genus and species levels, as well as the normalised abundance (reads per million) of each of the extracted functional categories. A schematic figure of the dominant genus and species level assignments of reads for each assessed functional category is shown in Figure 4.2. Overall, all but one of the categories assessed were dominated by either/or a mixture of *Prevotella* and *Succiniclasicum*. All *Succiniclasicum* reads were assigned as *Succiniclasicum ruminis* (van Gylswyk, 1995) at the species level; currently the only described species of the genus. For categories dominated by *Prevotella*, at the species level these often mapped to species and strains of the *Prevotella* 1 genus-level group (Henderson et al., 2019). The only category not dominated by either *Prevotella* or *Succiniclasicum* was propionate CoA-transferase (KEGG



orthology category K01026), which was instead dominated by *Megasphaera*. However, the abundance of transcripts mapping to this category was particularly low compared to the other groups analysed (Figure 4.2). Based on these results, strains of the *Prevotella 1* genus-level group (Henderson et al., 2019) were selected from the Hungate1000 collection for further characterisation.



**Figure 4.2. Taxonomic distributions of reads mapping to functional categories involved in the pathway of propionate production from succinate.** Numbers on the x-axis of all graphs represent the average percentage of total reads mapping to each KEGG/Interpro2GO category assessed in all ten samples analysed. The top 5 categories above 1% relative abundance are shown on graphs at both genus (black gridlines) and species (red gridlines) levels, and abundances of all other categories above 1% relative abundance were collapsed into the category 'other', with the number in brackets denoting the number of categories that were collapsed. Error bars denote one standard error of the mean (SEM). Numbers in blue underneath category names represent average  $\pm$  SEM reads per million (RPM) values of each category across the 10 samples. \* Reads belonging to functional categories corresponding to methylmalonyl-CoA epimerase were not found in the datasets, potentially indicative of a low turnover due to the known high thermostability of the enzyme (McCarthy et al., 2001).

## 4.3 Transcriptomic and proteomic changes associated with cobalamin-dependent propionate production by the rumen bacterium *Prevotella ruminicola*

### 4.3.1 Abstract

*Prevotella ruminicola* is an abundant rumen bacterium that can produce propionate in a cobalamin (vitamin B<sub>12</sub>)-dependent manner; however, the underlying genes and regulatory mechanisms are poorly understood. To assess whether propionate production in *P. ruminicola* is controlled by a cobalamin-binding riboswitch, we have conducted *in silico* analyses, and compared the transcriptomes of cultures grown in defined media in the presence and absence of cobalamin (cyanocobalamin). We generated the complete genome sequence of *P. ruminicola* KHP1, and found that it contained four 'cobalamin' riboswitches. However, these were not in close proximity to genes putatively involved in propionate production *via* the succinate pathway, nor were they co-located in a single operon. Comparative genomics of a selection of screened *Prevotella* 1 strains of the Hungate1000 collection found no differences in the presence of gene candidates involved in the conversion of succinate to propionate between propionate producing strains and a strain that did not produce propionate. However, the genomes of all propionate-producing strains possessed a conserved arrangement of a putative transport protein and three subunits encoding a putative methylmalonyl-CoA decarboxylase, upstream but in the antisense orientation to both co-located subunits of methylmalonyl-CoA mutase, whereas the P6B11 propionate non-producing strain did not. Cobalamin led to the differential expression of 17.5% of the genes in the KHP1 assembly, including some of the candidate propionate pathway genes. However, the effects of cobalamin on the KHP1 proteome were less pronounced, and the only enzyme putatively involved in the conversion of succinate to propionate differentially abundant at the proteome level was both co-located subunits of the cobalamin-dependent enzyme methylmalonyl-CoA mutase, which showed increased abundance in the presence of cobalamin. While our results demonstrate the differential expression of some propionate pathway candidate genes in response to cobalamin, these effects do not appear to be due to the direct control of a cobalamin riboswitch mechanism.

### 4.3.2 Introduction

Ruminant animals are characterised by the presence of a fermentative forestomach, the rumen, which hosts a complex community of microbes. The rumen microbiota drives ruminant metabolism, and in particular, is responsible for the degradation of ingested complex plant polysaccharides, and their fermentation into short-chain fatty acids (SCFAs) that the host uses to satisfy most of its energy requirements (Dijkstra, 1994). Of the three major SCFAs produced, acetate, propionate and butyrate, propionate is of particular importance to animals with high energy requirements as it is the major precursor of host gluconeogenesis (Aschenbach et al., 2010). Propionate also acts as an alternative hydrogen sink to methane production (Janssen, 2010). Enhancement of ruminal propionate production is therefore desired, particularly in forage-fed animals where propionate levels are generally lower than in animals fed more highly digestible concentrate-based diets (Bauman et al., 1971; Zhang et al., 2017). Propionate formation from sugars mainly proceeds by either the acrylate pathway, for which lactate is an intermediate, or the succinate pathway (Reichardt et al., 2014).

The genus *Prevotella* is a diverse (Purushe et al., 2010) and abundant (Henderson et al., 2015) rumen bacterial group, particularly so in concentrate-fed animals, and contains members that produce propionate (Dehority, 1966). Using a defined growth medium it was shown that the type strain, *Prevotella ruminicola* 23<sup>T</sup>, produces propionate in a cobalamin (vitamin B<sub>12</sub>)-dependent manner, with succinate accumulating in the absence of cobalamin (Strobel, 1992). The genomes of many ruminal *Prevotella* strains contain genes associated with the succinate pathway (Reichardt et al., 2014; Seshadri et al., 2018). However, it is recognised that some enzymes in the pathway may also be involved in the production of different short-chain fatty acids, such as butyrate (Louis & Flint, 2017). Thus, based on current genome annotations alone, it is unclear precisely which genes are involved in propionate production, as well as the regulatory mechanisms that underlie cobalamin-dependent production.

In numerous bacteria, cobalamin has been shown to control gene expression through directly binding to noncoding regions of mRNAs called riboswitches, which regulate the expression of neighbouring genes (Serganov & Nudler, 2013). These highly conserved elements (Vitreschak et al., 2003) are commonly found upstream of cobalamin-related genes and form ligand-binding secondary structures, often functioning as negative regulators that repress gene transcription (Li et al., 2020), or inhibiting translation (Richter-Dahlfors & Andersson, 1992). However, cobalamin riboswitches can also induce expression of cobalamin-related genes upon ligand binding. In *Listeria monocytogenes*, cobalamin controls expression of an antisense RNA that regulates expression of genes involved in propanediol metabolism, resulting in the cobalamin-dependent pathway being fully expressed only in the presence of both propanediol and cobalamin (Mellin et al., 2013). Cobalamin riboswitches have been broadly classified into two main classes (Chan & Mondragón, 2020), based on preferential binding to either adenosylcobalamin (Cbl-I) (Peselis & Serganov, 2012) or methylcobalamin/hydroxocobalamin (Cbl-IIa) (Johnson Jr et al., 2012). However, another cobalamin riboswitch of *Bacillus subtilis* characterised more recently can bind different forms of cobalamin by altering their structure, and shares some structural features characteristic of both classes, complicating its classification using current criteria (Chan & Mondragón, 2020). Despite this array of characterised examples in other bacteria, to our knowledge, riboswitches have yet to be studied in rumen bacteria.

Given the cobalamin-dependent propionate production of *P. ruminicola* (Strobel, 1992), coupled with the multifaceted role of cobalamin riboswitches in regulating gene expression of other bacteria, we hypothesised that the genes encoding enzymes of the pathway of propionate production in ruminal *Prevotella* are under the control of cobalamin-binding riboswitches. We therefore investigated this by applying a multi-omics' approach, including genome analyses of ruminal *Prevotella* to identify pathway gene candidates and the presence of cobalamin riboswitches, and transcriptomic and proteomic analyses of cobalamin-induced propionate production, to gain greater insight into the propionate pathway and its regulation in *P. ruminicola*.

### 4.3.3 Materials and Methods

#### 4.3.3.1 Biological material and growth conditions

All strains included in this study were sourced from the Hungate1000 collection (Seshadri et al., 2018). Cultures were revived from frozen glycerol stocks or freeze-dried cells in nutrient-rich M2GSC medium (Miyazaki et al., 1997) containing 30% (v/v) centrifuged rumen fluid, and 0.2% (w/v) each of glucose, cellobiose and soluble starch. Before the commencement of experiments, cultures were passaged three times on the defined medium by Strobel (1992) in the absence of cobalamin, to prevent any potential carry-over effects of cobalamin present in rumen fluid of the revival medium. In all experiments, cultures were incubated as static batch cultures at 39°C in the dark.

#### 4.3.3.2 Screening of *Prevotella 1* strains for cobalamin-dependent propionate production

Selected strains of the Hungate1000 collection assigned to the *Prevotella 1* genus-level cluster (Henderson et al., 2019) were revived from glycerol stocks in nutrient-rich media, and their identities confirmed by Gram staining and 16S rRNA gene sequencing. Triplicate 10 mL batch cultures of each strain were cultured in defined medium (Strobel, 1992) in the presence and absence of 50 µg/L cyanocobalamin, and incubated at 39°C for 48 hr. Optical density measurements at 600 nm were taken immediately after inoculation and subtracted from end-point readings. Fermentation end products were quantified from culture supernatants by gas chromatography, as previously described (Palevich et al., 2019a).

#### 4.3.3.3 Comparative transcriptomics/proteomics experiment of cobalamin-induced propionate production

To investigate the genes and proteins that are upregulated during cobalamin induction of the propionate pathway in KHP1, 100 mL cultures were grown in the presence and absence of cobalamin (n = 6 per treatment) to generate samples for transcriptome and proteome

analyses. At two-hourly timepoints, as well as immediately after inoculation, 1 mL was taken from each culture using a sterile needle and CO<sub>2</sub>-flushed syringe, and the optical density (600 nm) was measured using a Spectronic 200 spectrophotometer (Thermo Fisher). The sample was transferred from the cuvette to an Eppendorf tube, and fermentation end product concentrations were measured as previously described (Palevich et al., 2019a). Cultures were harvested during log phase growth after 10 hours by flash-freezing in liquid nitrogen and stored at -80°C for transcriptome and proteome analyses.

#### 4.3.3.4 16S rRNA phylogeny

Phylogenetic trees of full-length 16S rRNA sequences were constructed using Geneious v10.0.9 (Kearse et al., 2012). Sequences were aligned using MUSCLE (Edgar, 2004), and a Maximum-Likelihood tree was generated using the default settings in MEGA X (Kumar et al., 2018). The resulting tree was annotated using iTOL (Letunic & Bork, 2016).

#### 4.3.3.5 Genome sequencing and analyses

The complete genome of *P. ruminicola* KHP1 was sequenced and assembled as previously described (Mahoney-Kurpe et al., 2021). Comparative genomics were carried out using the Integrated Microbial Genomes (Joint Genome Institute) server (Chen et al., 2021). Gene island maps were generated using GenomeDiagram (Pritchard et al., 2006) in Biopython (Cock et al., 2009). Contigs were scanned for cobalamin riboswitch families ('cobalamin', 'adoCbl' and 'adoCbl-variant' families) using Riboswitch Scanner (Mukherjee & Sengupta, 2016). The KHP1 assembly was scanned for genes encoding proteins involved in the conversion of succinate to propionate by blastp searches of reviewed proteins in other bacteria of the UniProt database (The UniProt Consortium, 2021) to scan for homologues (> 30% homology, > 75% query cover) in the KHP1 genome. COG functional assignments were made using eggno-mapper v2.0 (Huerta-Cepas et al., 2017).



#### 4.3.3.6 RNA-seq analyses

Total RNA was extracted from cultures using a modified acid phenol/chloroform procedure of mechanically lysed cells by bead-beating, as previously described (Palevich, 2016). RNA was DNase-treated using the Turbo DNase kit (Thermo Fisher Scientific) and purified using the MEGAclean kit (Thermo Fisher Scientific). Extracts were quantified by Qubit (Invitrogen) and assessed of integrity using a BioAnalyser 2100 with the RNA nano 6000 assay reagent kit (Agilent Technologies). RNA samples were sequenced by Novogene (Beijing, China), with ribosomal RNA depletion using the RiboZERO Magnetic kit (Illumina), and library preparation using the NEBnext Ultra II directional RNA library preparation kit (Illumina). Libraries were sequenced on an Illumina NovaSeq instrument.

Raw reads were trimmed of any remaining adapter sequences (default settings) and quality-filtered (-q 20) using cutadapt v3.3 (Martin, 2011). Filtered reads were aligned against the complete KHP1 genome using Hisat2 v2.2.1 (Kim et al., 2019) under default settings. SAM alignment files were converted to BAM format using samtools v1.11 (Li et al., 2009), and read alignment counts were extracted using FeatureCounts v2.0.1 (Liao et al., 2014). To identify differentially expressed genes, the resulting matrix was input into R, and read counts were  $\log_2$ -transformed. Differentially expressed genes were identified between treatments using the DESeq2 package (Love et al., 2014), with significance defined using FDR-adjusted  $P < 0.05$  and  $|\log_2\text{fold change}| \geq 1$  cutoffs. Principal Component Analysis (PCA) analyses was carried out on  $\log_2$ -transformed read counts of each gene of the KHP1 genome using the 'prcomp' function in R v4.1.1, and plotted with the 'ggfortify' package.

#### 4.3.3.7 Proteome analyses

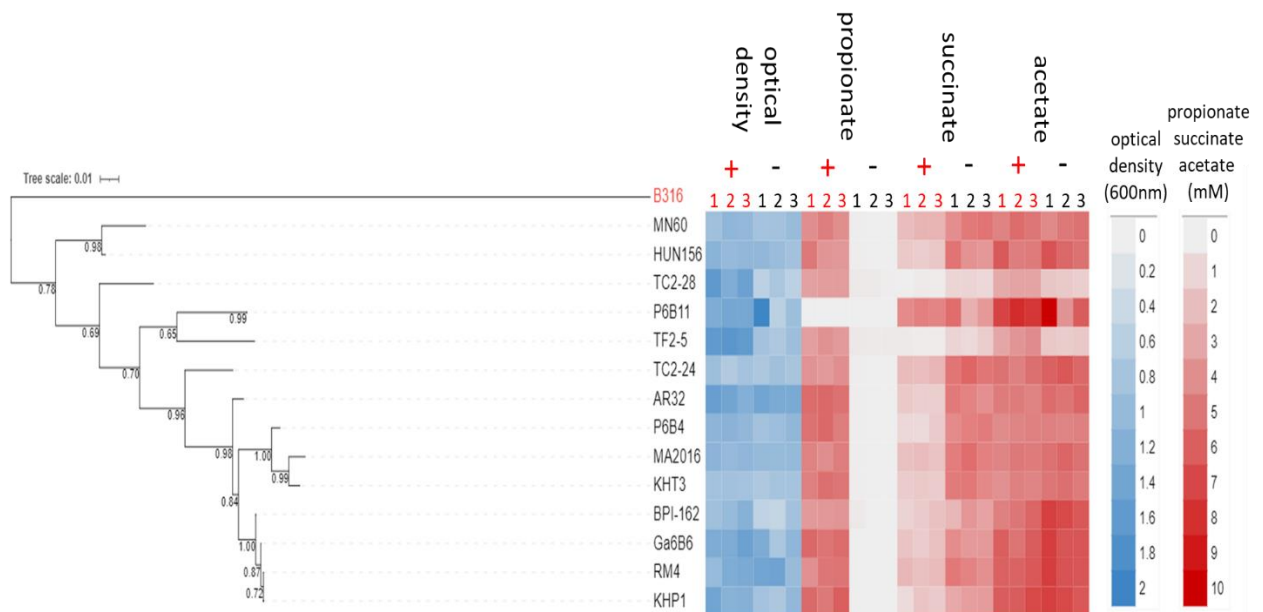
Proteins were extracted from approximately 50 g of each frozen culture, using a modified method based on that described in Delogu et al. (2020). Raw MS files were analysed using MaxQuant (Tyanova et al., 2016a) and proteins identified and quantified using the MaxLFQ algorithm (Cox & Mann, 2008). ProteinGroups files from MaxQuant were further processed

and analysed in Perseus (Tyanova et al., 2016b). Proteins identified as contaminants, reverse proteins and proteins identified only by site were filtered. Label-free quantification (LFQ) intensities of proteins were logarithmically normalised ( $\log_2$ ), and missing data were imputed based on the normal distribution. Differentially abundant proteins were identified by carrying out a two-sample T-test, with significance defined using FDR-adjusted  $P < 0.05$  and  $|\log_2\text{fold change}| \geq 1$  cutoffs. Principal component analysis was carried out as described above for transcriptome data, using the  $\log_2$ -transformed, imputed data as input.

## 4.3.4 Results

### 4.3.4.1 Most ruminal *Prevotella 1* strains produce propionate in the presence of cobalamin

To assess the effect of cobalamin supplementation on growth and fermentation of ruminal *Prevotella 1* strains, fourteen strains from the Hungate1000 culture collection (Seshadri et al., 2018) were cultured in a defined growth medium in the presence and absence of cobalamin (Strobel, 1992). The optical density and fermentation end product concentrations after 48 hours of growth showed that apart from *Prevotella brevis* P6B11, all strains produced propionate in a cobalamin-dependent manner, with succinate instead accumulating in its absence (Figure 4.3). None of the strains required cobalamin for growth, although some, such as TF2-5 and TC2-28, grew to greater optical densities when it was supplied (Figure 4.3).



**Figure 4.3. Screening of *Prevotella 1* strains for cobalamin-dependent propionate production.** A phylogenetic maximum-likelihood tree generated from alignments of full-length 16S rRNA genes of each of the screened *Prevotella 1* strains (left), is shown alongside culture growth via OD<sub>600</sub> optical density (blue heatmap) and propionate, succinate and acetate concentrations (red heatmap). Numbers 1-3 above columns represent the replicate number of cultures per treatment (red font, cobalamin-supplemented (+); black font, cobalamin-omitted (-)). Values shown were taken after 48 hrs of incubation, with initial background values subtracted. *Butyrivibrio proteoclasticus* B316<sup>T</sup> (red) was included as an outgroup.

#### 4.3.4.2 The complete KHP1 genome contains four putative cobalamin riboswitches

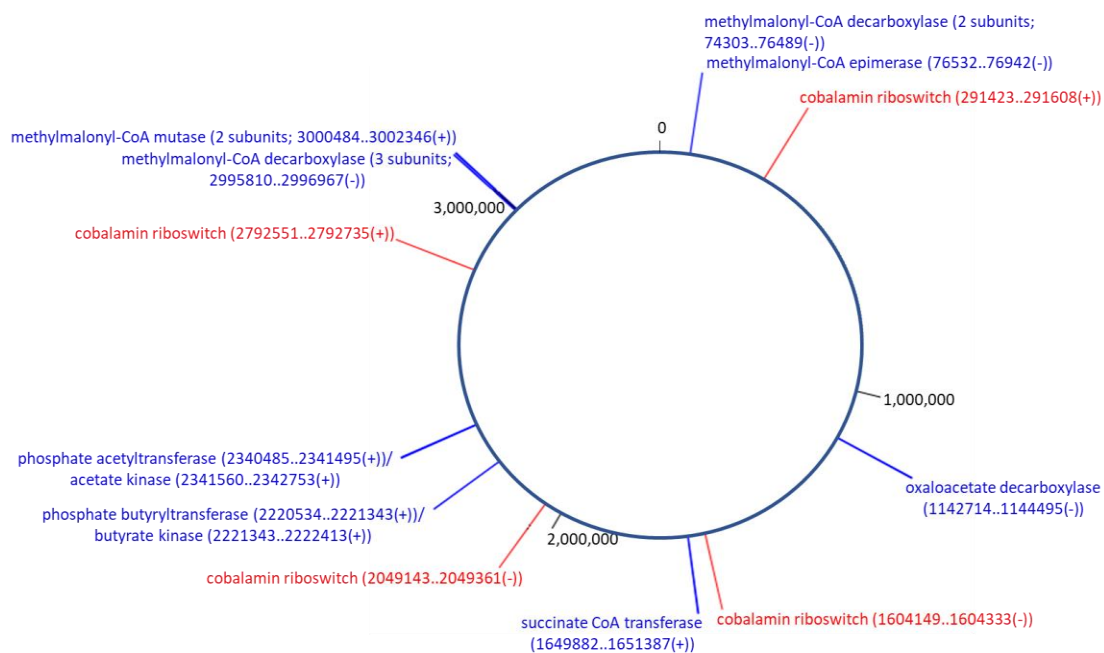
To characterise the cobalamin regulation of propionate production *via* the succinate pathway, *Prevotella ruminicola* KHP1, originally isolated from bovine rumen contents in New Zealand fed a meadow/pasture hay diet and cultured on RM02 media supplemented with vitamin K, haemin and penicillin (William J. Kelly, personal communication) was selected for further analyses, due to its close 16S rRNA gene sequence identity with the type strain *P. ruminicola* 23<sup>T</sup> (Avgustin et al., 1997). KHP1 exhibited clear cobalamin-dependent propionate production (Strobel, 1992) and cultures grew to similar optical densities in both the presence and absence of cobalamin, thus investigations into propionate pathway regulation were deemed less likely to be confounded by differences in growth rate.

To understand the putative role of riboswitches in regulating propionate production, we initially sought to scan the genome for the presence of candidate propionate pathway genes, as well as cobalamin riboswitch sequences (Mukherjee & Sengupta, 2016). However, the draft genome of KHP1 generated as part of the Hungate1000 project, resulted in six contigs (Seshadri et al., 2018), and thus the order and orientation of the contigs was unclear, as was how much (if any) of the genome was missing. We therefore re-sequenced the KHP1 genome using a hybrid long/short read approach, as previously described (Mahoney-Kurpe et al., 2021). This resulted in a circular contig 3.4 Mb in size, with a G+C content of 47.8% (GenBank accession CP071890), with additional genome information shown in Appendix 7.5. Four predicted cobalamin riboswitch sequences were identified (Appendix 7.6), which were dispersed widely around the KHP1 genome (Figure 4.4a).

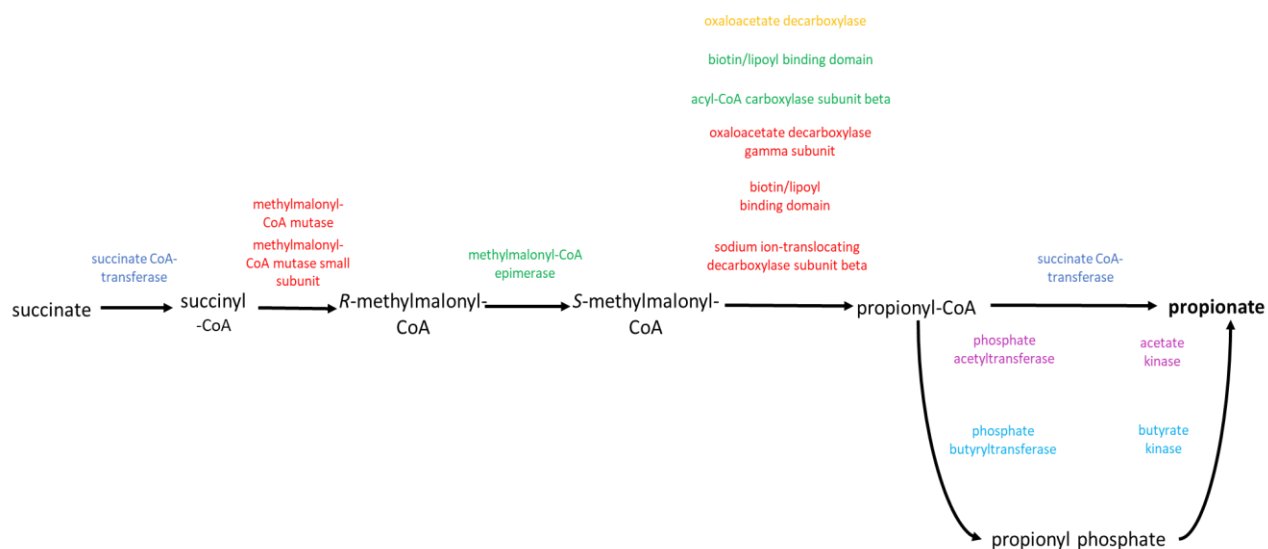
#### 4.3.4.3 Candidate succinate pathway genes are widely dispersed along the KHP1 chromosome and not in close proximity to predicted cobalamin riboswitches

The genomic locations of candidate genes involved in the conversion of succinate to propionate were determined by scanning the genome for candidate propionate pathway genes by blastp searches of numerous candidate protein sequences of other bacteria in the Swiss-Prot database (The UniProt Consortium, 2021) (Appendix 7.7). The gene candidates were found to be widely dispersed, and not in a distinct operon, nor were they located near the cobalamin riboswitches. The only occurrence of gene co-location involved those predicted to encode methylmalonyl-CoA mutase subunits near to three putative methylmalonyl-CoA decarboxylase subunits, and methylmalonyl-CoA epimerase containing immediately downstream a gene encoding a methylmalonyl-CoA decarboxylase beta subunit and a biotin binding domain, separated by a short 50 aa hypothetical protein. Multiple gene candidates for the conversion of propionyl-CoA to propionate were examined, due to the cross-reactivity of different CoA-transferases and kinases involved in short-chain fatty acid production (Louis & Flint, 2017). However, none of these genes were closely located to gene candidates for other steps of the pathway (Figure 4.4a; Appendix 7.8).

a)



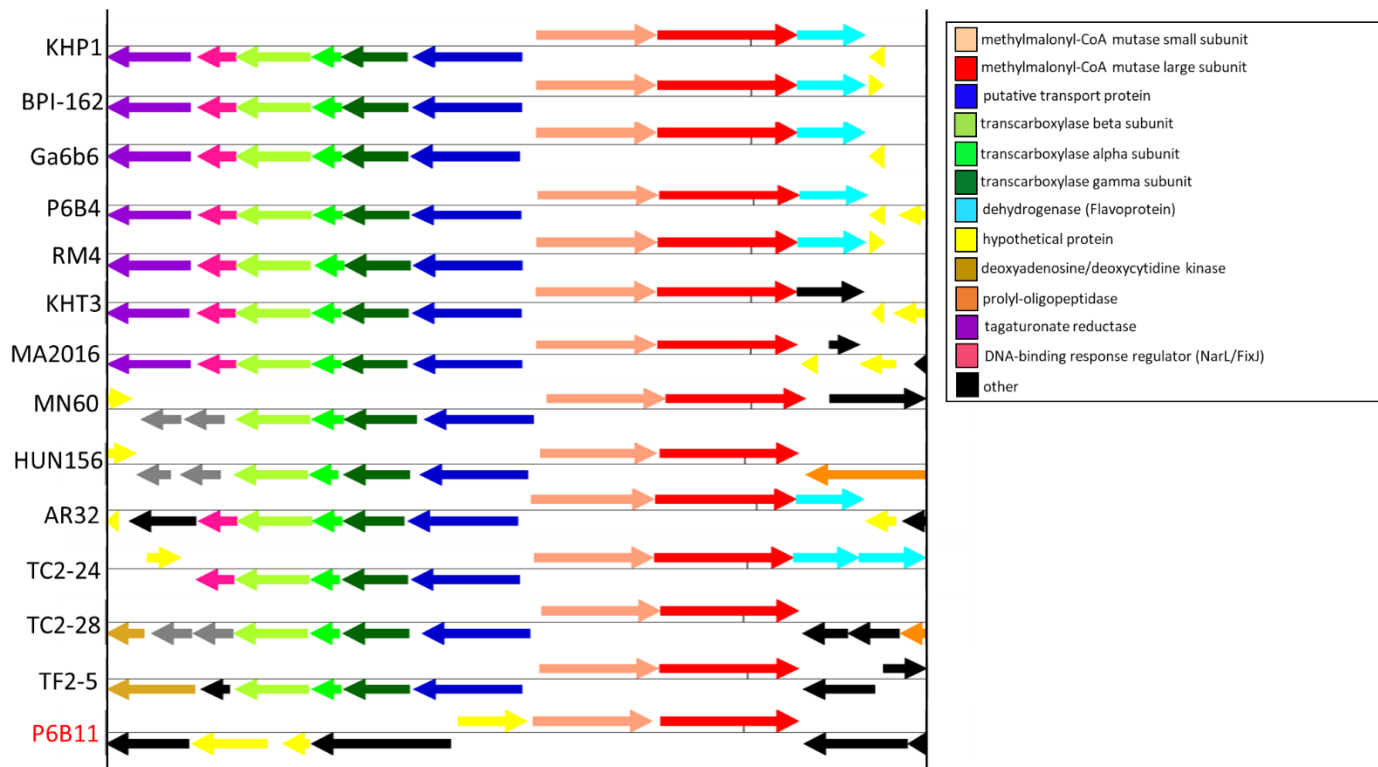
b)



**Figure 4.4. Widespread dispersal of cobalamin riboswitches and candidate propionate pathway genes on the KHP1 genome. a)** Circular map of the KHP1 chromosome showing the positioning of detected cobalamin family riboswitches (red) and candidate genes involved in converting succinate to propionate (blue). Smaller black markers represent genome coordinates (bp). **b)** Schematic diagram showing the involvement of detected candidates at each step in the conversion of succinate to propionate. Labels of candidate genes are coloured based on their co-location at each of the six loci of the KHP1 genome containing candidate genes.

#### 4.3.4.4 Comparative genomics of *Prevotella* 1 strains differing in their propionate production

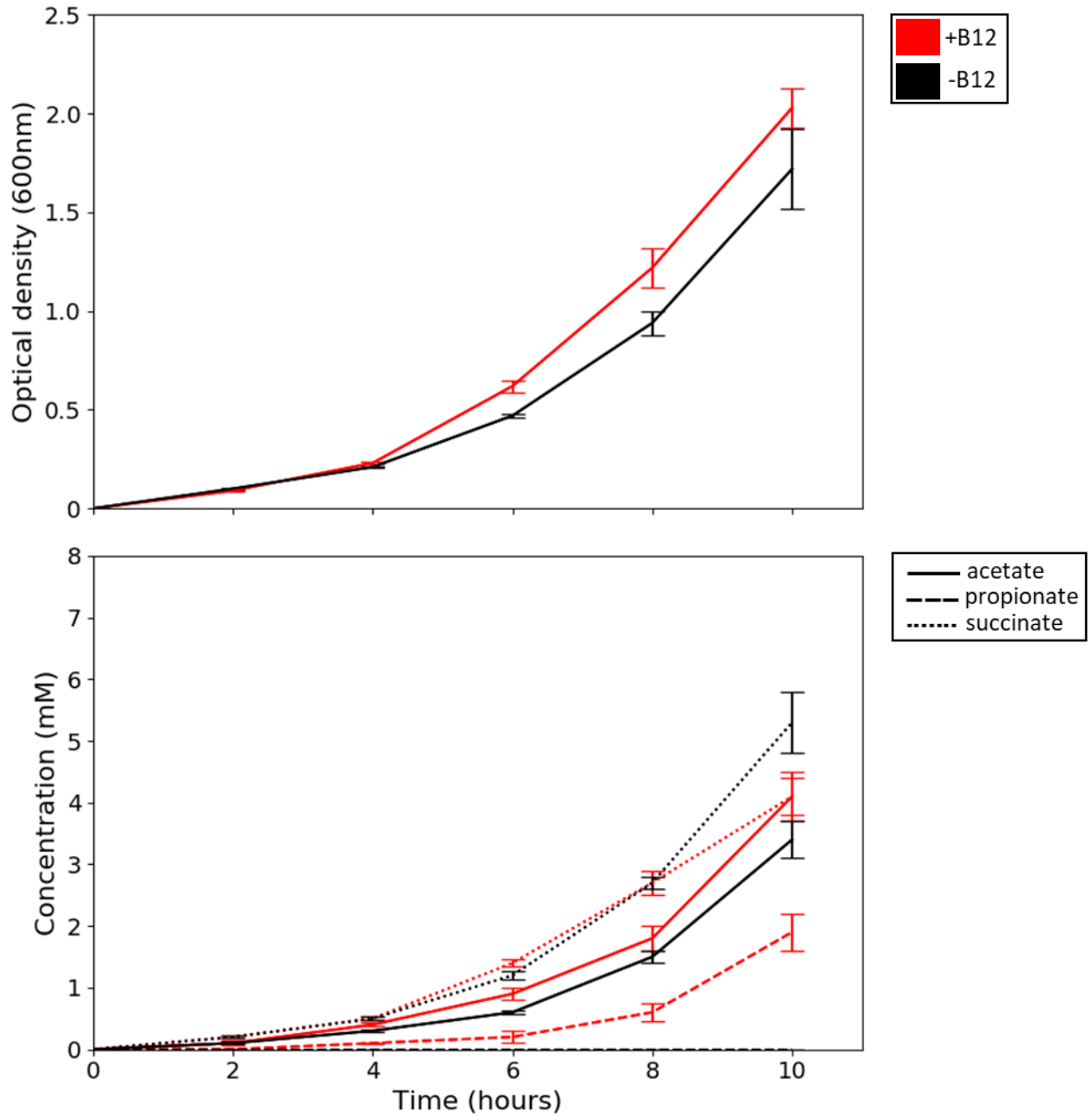
As *P. brevis* P6B11 lacked the ability to produce propionate both with and without cobalamin, we therefore wanted to compare whether any genomic signatures regarding the presence/absence of candidate propionate pathway genes and their co-location on contigs of each assembly could be identified that are conserved across the strains that produced propionate, but not in the P6B11 strain that did not. To do so, we scanned the genomes of each screened strain for homologues of each of the KHP1 candidate genes involved in the conversion of succinate to propionate. All draft genomes of screened strains possessed homologues of all the propionate pathway candidate genes identified in KHP1 (Appendix 7.9). However, in all strains that produced propionate, genes predicted to encode a putative transport protein and three subunits of a putative methylmalonyl-CoA decarboxylase were positioned upstream and in the antisense orientation, of the methylmalonyl-CoA mutase subunits. In contrast, the genes for an antisense putative transport protein and methylmalonyl-CoA decarboxylase seen at this locus across all other 13 propionate-producing strains were not in close proximity to the methylmalonyl-CoA mutase genes of the P6B11 strain (Figure 4.5) and were instead co-located together on another contig of the assembly (Appendix 7.9). Each of the contigs containing these genes were scanned for cobalamin riboswitches, although as in the KHP1 genome, none were detected in this region in any of the other strains.



**Figure 4.5. Conserved arrangement of methylmalonyl-CoA mutase, putative transporter and methylmalonyl-CoA decarboxylase genes across all propionate-producing Hungate1000 strains but absent in the non-producing P6B11 strain.** Gene island maps of regions of each screened contig containing methylmalonyl-CoA mutase subunits. Genes classified as 'other' represent annotations present on only one extracted fragment. Black markers underneath methylmalonyl-CoA mutase large subunit genes represent 10 kb markers from the left side of each panel.

#### 4.3.4.5 Comparative transcriptomics and proteomics of the impact of cobalamin on KHP1 gene expression

To investigate the genes and proteins that are upregulated during cobalamin induction of the propionate pathway in KHP1, cultures were grown in the presence and absence of cobalamin to generate samples for transcriptome and proteome analyses. Cultures were harvested during log phase growth (10 hr). Consistent with our previous observations, there was no significant difference in growth, nor acetate production between the treatments after 10 hours of growth (two sample T-test;  $P > 0.05$ ) (Figure 4.6). As expected, propionate was detected only in cobalamin-supplemented cultures, and greater concentrations of succinate tended to be observed in the absence of cobalamin and propionate ( $P = 0.07$ ) (Figure 4.6).



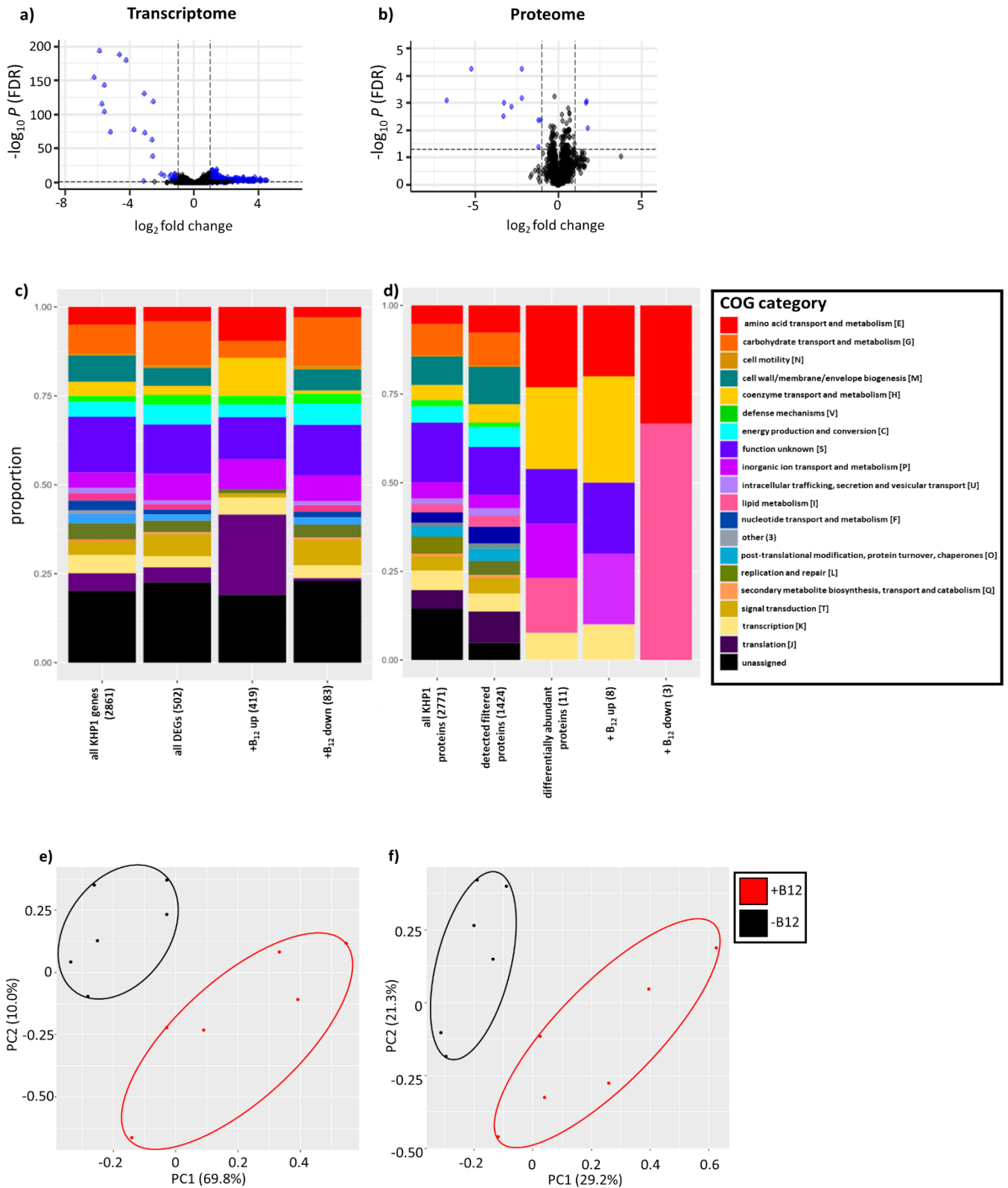
**Figure 4.6. Optical density and fermentation end product formation of KHP1 cultures throughout growth.** Error-bars denote one SEM (n = 6).

RNA-seq resulted in an average of  $8,178,217 \pm 976,705$  quality-filtered ( $Q > 20$ ) reads per sample, which aligned to the KHP1 genome at an average rate of  $98.7 \pm 0.4\%$  per sample. We identified 502 differentially expressed genes (DEGs) of which 419 were upregulated, and 83 were downregulated during cobalamin supplementation (Appendix 7.10). Despite far fewer genes being downregulated, these differences were generally more highly significant, and showed greater  $\log_2$  fold change differences in transcript abundance (Figure 4.7a). Of the



DEGs, it was most common for these to not be assigned to any COG category (22.4%) or to the category 'unknown function (S)' (13.7%). Of DEGs assigned to functionally descriptive categories, the most abundant were carbohydrate metabolism and transport (G)' (12.3%), 'inorganic ion transport and metabolism (P)' (7.5%), 'signal transduction (T)' (6.1%), and 'energy production and conversion (C)' (5.6%). Categories enriched for genes downregulated by cobalamin included 'translation (J)' (22.6%), unassigned to any COG (19%), assigned to 'function unknown (S)' (11.9%), and the 'transport and metabolism' categories of coenzymes (H) (10%), amino acids (E) (9.5%), and inorganic ions (P) (8.3%). Of the categories assigned to genes upregulated by cobalamin, the most abundant were unassigned to any COG (23%), assigned as 'function unknown (S)' (14.1%), 'carbohydrate metabolism and transport (G)' (13.7%), inorganic ion transport and metabolism (7.3%), and signal transduction (7.1%) (Figure 4.7c). Principal component analysis (PCA) plots of log<sub>2</sub>-transformed read counts showed clear separation samples in the presence and absence of cobalamin (Figure 4.7e).

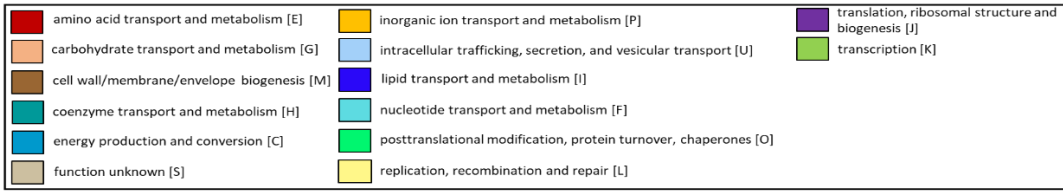
Proteome analyses of the samples using untargeted LC-MS/MS identified a total of 1,853 proteins, of which, 1,424 passed downstream quality filtering. A two-sample T-test identified only 13 differentially abundant proteins; three in the presence and ten in the absence of cobalamin (Figure 4.7b; Appendix 7.11). PCA plots also showed clear separation of the samples between each treatment (Figure 4.7f).



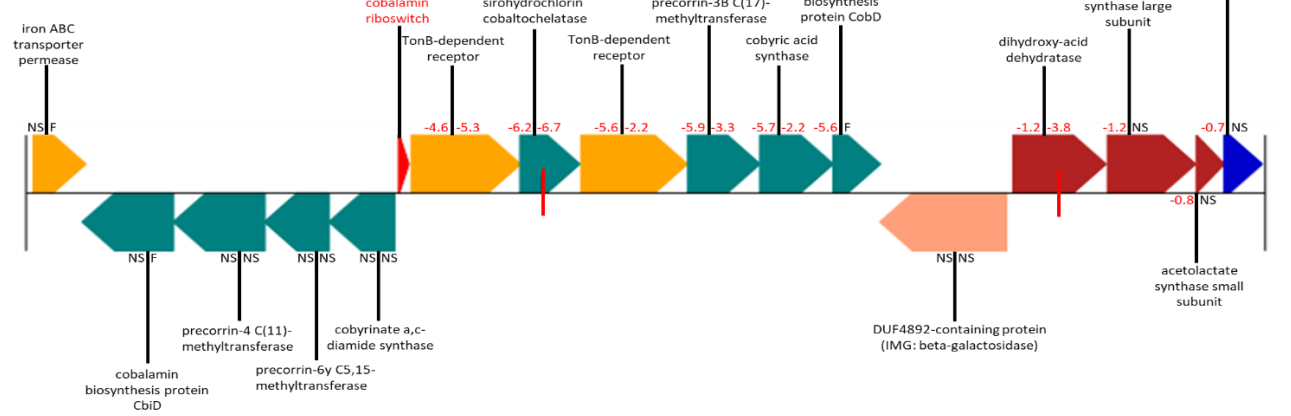
**Figure 4.7. Cobalamin-induced shifts in the KHP1 transcriptome and proteome.** Volcano plots of differentially expressed (a) transcripts and (b) proteins. Differentially expressed genes and proteins (FDR-adjusted  $P < 0.05$ ,  $|\log_2\text{fold change}| \geq 1$ ) are shown in blue. Positive  $\log_2\text{fold change}$  represents an increase abundance in the presence of cobalamin. Stacked bar plots of proportions of (c) transcriptome and (d) proteome data assigned to different COG categories. Principal component analysis (PCA) plots of  $\log_2$ -transformed (e) transcriptome data, and (f) filtered,  $\log_2$ -transformed and imputed proteome data.

#### 4.3.4.6 Differential expression of genes and proteins in close proximity to cobalamin riboswitches

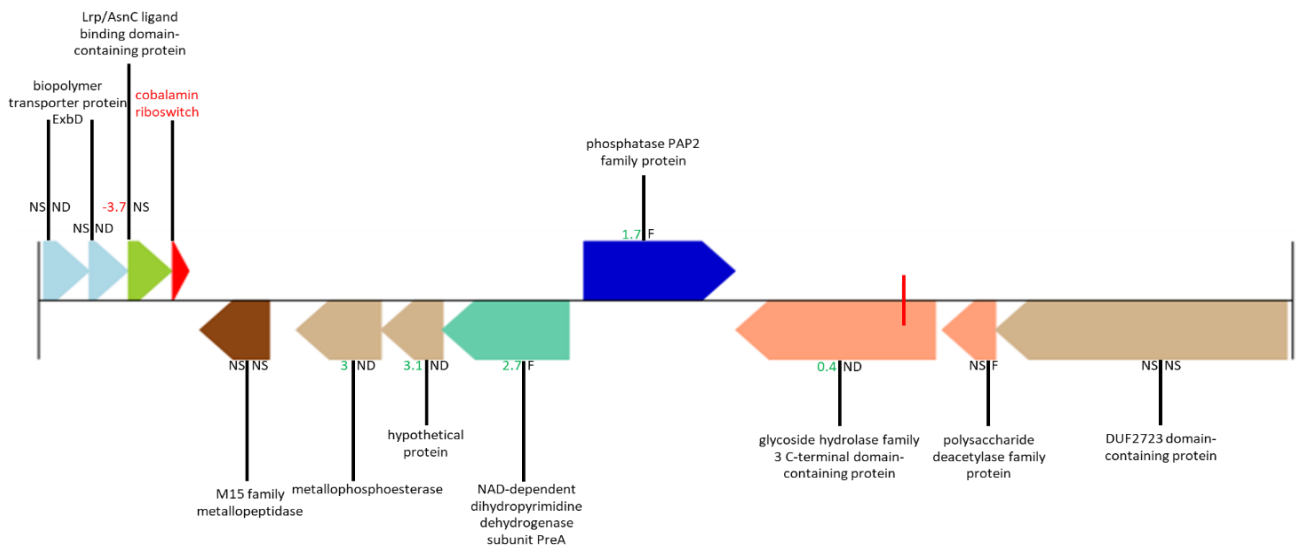
The most highly differentially expressed genes and proteins due to cobalamin supplementation (Appendices 7.10 and 7.11) were often in close proximity to the cobalamin riboswitch sequences (Figure 4.8). The riboswitch sequence at genome position 291,243 was immediately upstream of an operon of six genes involved in cobalamin biosynthesis and transport. All six genes were downregulated by cobalamin, and differences in the abundances of proteins encoded by the first five genes were also highly significant (two sample T-test,  $P \leq 0.001$ ) (Figure 4.8), and among the top seven highly differentially expressed proteins found in the proteome data between treatments (Appendix 7.11). At other riboswitch loci, some of the genes showing differential expression were positioned upstream and/or antisense to putative cobalamin riboswitches (Figure 4.8b-d). However, these genes did not exhibit significant differences in protein abundance. The cobalamin riboswitch at position 2,792,551 was unique among the four detected, given that it did not possess any genes closely downstream on the same strand, but instead had a cluster of genes downstream but in the antisense orientation, the three of which more distant from the riboswitch were also significantly upregulated by cobalamin at the transcriptome level. However, the proteins encoded by these genes were not detected in the final proteome dataset (Figure 4.8b).



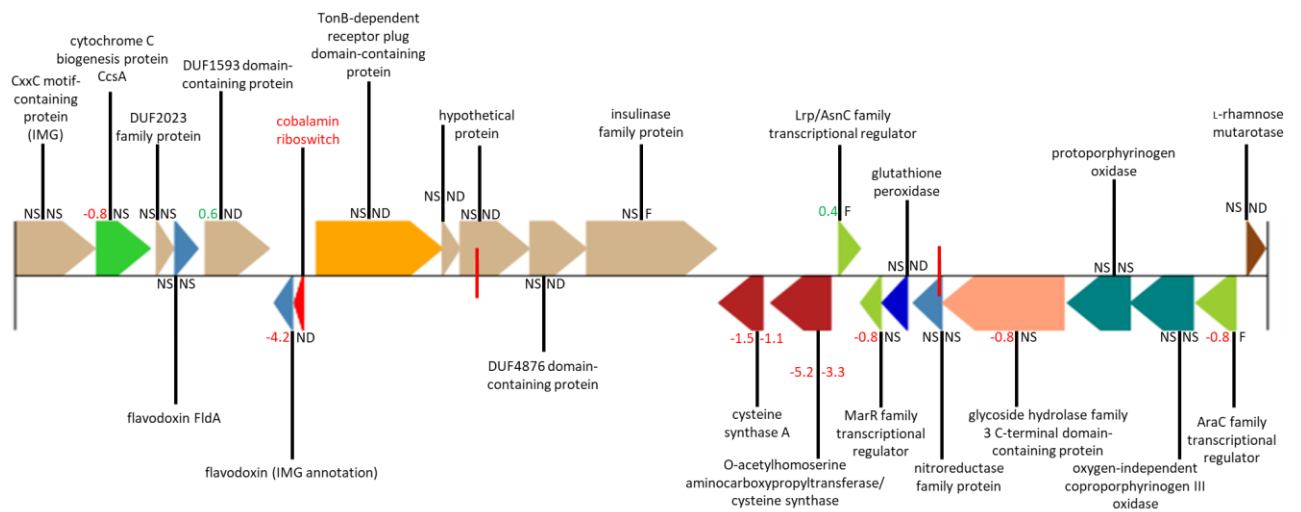
**a) 284200..308200**



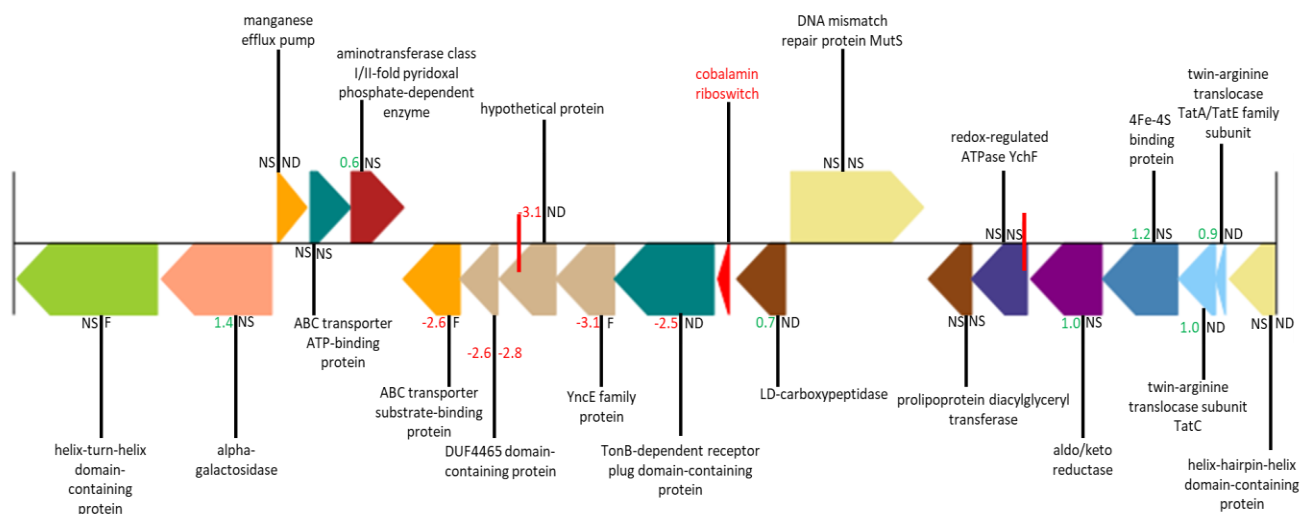
**b) 2791000..2805500**



c) 1598100..1625200



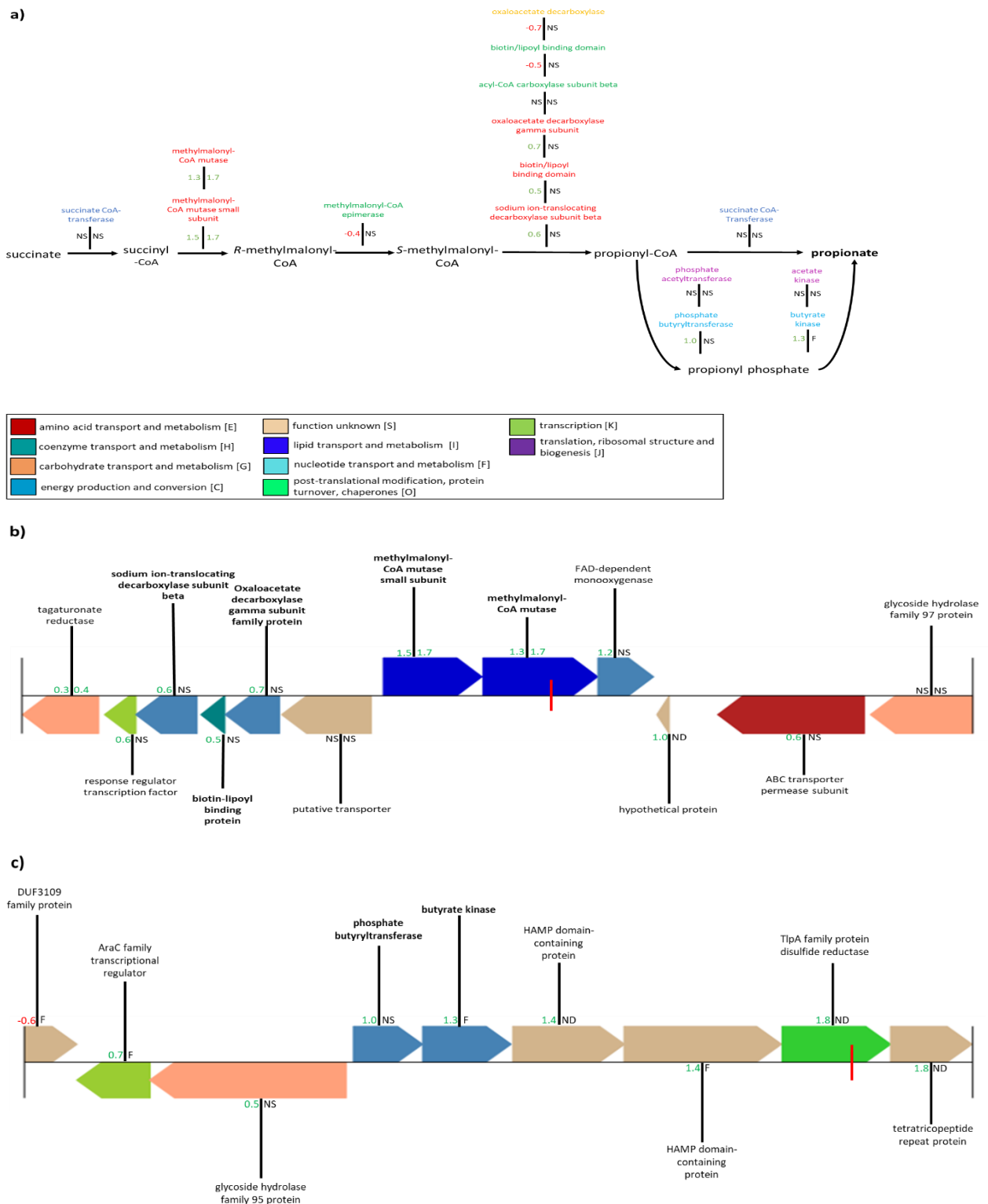
d) 2035200..2060200



**Figure 4.8. Gene and protein expression of genes in close proximity to predicted cobalamin family riboswitches.** Numbers to the left of markers for each gene represent  $\log_2$  fold changes of transcript abundances (left) and protein abundances (right) between each treatment. In both instances, positive values represent increased abundance in the vitamin B<sub>12</sub> supplemented medium. NS = not significant (two sample T-test; FDR-adjusted  $P > 0.05$ ); ND = Protein not detected in proteome data; F = protein filtered during quality control. Smaller black vertical markers represent 10 kb increments from the left of each fragment. Numbers in titles of each diagram represent genome coordinates of each fragment of the KHP1 genome. Annotation names shown are based on PGAP annotations.

#### 4.3.4.7 Differential expression of some candidate propionate pathway transcripts between cobalamin treatments, but effects on the proteome limited to overexpression of both methylmalonyl-CoA mutase subunits

A summary of the impact of cobalamin on the expression of transcripts and proteins of gene candidates involved in propionate production from succinate is shown in Figure 4.9. Irrespective of their distance from predicted cobalamin riboswitches, half of the gene candidates putatively involved in the conversion of succinate to propionate were upregulated due to the presence of cobalamin. These included both subunits of methylmalonyl-CoA mutase, as well as three co-located subunits encoding a putative methylmalonyl-CoA decarboxylase, although the upregulation of putative methylmalonyl-CoA decarboxylase subunits was minor ( $\log_2$ fold change < 0.8). Co-located genes annotated as a phosphate butyryltransferase and butyrate kinase were also both upregulated by cobalamin, despite the strain not producing any butyrate. These genes were the first two genes of an apparent operon including two downstream HAMP domain-containing proteins (Aravind & Ponting, 1999), a TlpA family disulfide reductase and a tetracopeptide repeat containing protein, all of which were upregulated by cobalamin at the transcriptome level (Figure 4.9c). In contrast, some other candidate genes such as methylmalonyl-CoA epimerase and a single subunit annotated as oxaloacetate decarboxylase at another locus were instead significantly downregulated by cobalamin, though the magnitude of these effects were minor ( $\log_2$ fold change < 0.8). At the proteome level, the only enzyme putatively involved in propionate production from succinate that was differentially abundant were the genes encoding both subunits of the cobalamin-dependent methylmalonyl-CoA mutase, which were both overexpressed in the presence of cobalamin. However, in the case of the gene encoding a butyrate kinase that was significantly upregulated at the transcriptome level, the corresponding protein group was removed during quality-filtering and could not be analysed.



**Figure 4.9. Impact of cobalamin on the expression of candidate genes putatively involved in converting succinate to propionate.** **a)** Numbers to the left/right of markers for each gene represent log<sub>2</sub>fold changes of transcript and protein abundances between treatments, respectively. In both instances, positive values represent increased abundance in the presence of cobalamin. NS = not significant (FDR-adjusted P > 0.05); F = protein filtered during sample quality control. Gene annotation names are coloured based on their clustering at the six loci of the KHP1 genome containing candidate genes. Impact of cobalamin on the expression of genes at loci containing upregulated **b)** methylmalonyl-CoA decarboxylase/methylmalonyl-CoA mutase and **c)** phosphate butyryltransferase/butyrate kinase genes.

### 4.3.5 Discussion

The aim of this study was to identify and better define the pathway and regulation of propionate production in *P. ruminicola*, with the hypothesis that the genes responsible are under the control of a cobalamin-binding riboswitch. *In silico* analyses of the completed KHP1 genome assembly revealed four cobalamin family riboswitches in the KHP1 genome; however, these were not located near identified candidate genes involved in converting succinate to propionate, nor were the candidate genes co-located in a single operon. Thus, while our results show differential expression of some candidate genes at the transcriptome level and both subunits of methylmalonyl-CoA mutase at the proteome level, the pathway of propionate production in this bacterium do not appear to be under the direct control of a cobalamin-binding riboswitch mechanism.

The mechanism(s) by which cobalamin caused the differential expression of succinate pathway candidates are not clear. In addition to their influence on gene expression by binding to riboswitches, there is increasing awareness of the roles of cobalamin as a co-factor of regulatory proteins (Klug, 2014). For example, in the phototrophic bacterium *Rhodobacter capsulatus*, cobalamin has been shown to act as a cofactor of an anti-repressor of the transcriptional regulator CrtJ, which regulates gene expression of enzymes involved in the biosynthesis of haem, carotenoids and light-harvesting proteins (Cheng et al., 2014). We found that cobalamin led to the differential expression of approximately 20% of the transcripts of the entire KHP1 genome, including some genes predicted to encode regulatory proteins. While we did not identify any predicted regulatory proteins that were differentially abundant at the proteome level, their typically low abundance can prevent their detection in proteome data (Smaczniak et al., 2012). It is also possible that post-translational modifications may be involved in the regulation of these genes (Macek et al., 2019).

Of the steps between succinate and propionate, it was of particular interest to identify the genes involved in converting propionyl-CoA to propionate, as due to redundancy between different short-chain fatty acid-producing transferases and kinases, it can be difficult to



pinpoint specifically which genes are involved in carrying out this step (Louis & Flint, 2017). In the KHP1 genome, three potential pathways for this reaction to occur were identified- one being by succinate CoA-transferase coupled to the conversion of succinate to succinyl-CoA, and the other two being *via* propionyl phosphate formation by a phosphate acetyl/butyryltransferase, and then production of propionate by an acetate/butyrate kinase. Of these, the only pathway that showed upregulation in the presence of cobalamin were the co-located genes annotated as phosphate butyryltransferase and butyrate kinase, suggesting that these genes may instead represent a putative phosphate propionyltransferase/propionate kinase. Future biochemical characterisation of the purified proteins encoded by candidate genes (Charrier et al., 2006) will shed further light on the potential involvement of these genes in propionate production.

Another particularly interesting finding of this study was the conserved arrangement of the putative transporter protein and three subunits of a putative methylmalonyl-CoA decarboxylase upstream but in the antisense orientation to methylmalonyl-CoA mutase genes that we found present across all 13 strains screened that produced propionate. However, while homologs of all of these genes were also present in the *P. brevis* P6B11 strain that did not produce propionate, the putative transporter and three subunits of a methylmalonyl-CoA decarboxylase were not in close proximity to the methylmalonyl-CoA mutase genes, and were instead co-located on a different contig of the assembly. This finding suggests that as opposed to their mere presence/absence, this specific orientation of these genes at this loci may somehow impart the ability of strains to produce propionate. While operons have traditionally been recognised as genes organised together as polycistronic mRNAs, deep RNA sequencing methods are redefining this dogma, and have revealed numerous examples of nonclassical operon architectures and mechanisms by which genes that are not immediately clustered together or in the same orientation can nonetheless be co-ordinately regulated (Bervoets & Charlier, 2019; Conway et al., 2014). Our transcriptome results showing significant upregulation of these genes also provides further evidence of their co-regulation, through a mechanism also involving cobalamin. Further targeted analyses of this locus will provide a better understanding of its regulation, and elucidate whether such a cobalamin-controlled mechanism indeed exists.

As is characteristic of bacterial genomes (Chang et al., 2016), over one third of genes of the KHP1 genome encode proteins with hypothetical annotations, some of which were among the identified differentially expressed genes between cobalamin treatments. A limitation of our approach that should therefore be emphasised is that our strategy of scanning genomes for genes that would be considered candidate genes of the succinate pathway can inherently only identify those candidate genes currently annotated as such, and we therefore cannot discount the possibility that further candidates could be among these currently hypothetical proteins that could be involved in propionate production. As genes encoding currently unannotated proteins continue to become characterised and functionally annotated, further candidates to those analysed in this study may be identified.

Despite not directly regulating candidate propionate pathway genes, our results confirm the functioning of other detected cobalamin riboswitches in the KHP1 genome in regulating the expression of other closely located genes. The riboswitch identified at position 291,243 functions by downregulating the expression of a cluster of six genes involved in cobalamin biosynthesis and transport in the presence of cobalamin, similarly to the riboswitch regulating the *cob* operon in *Salmonella typhimurium* (Richter-Dahlfors & Andersson, 1992). However, unlike these examples which have been shown to function post-transcriptionally, we instead observed a strong downregulation of the downstream operon at both the transcriptome and proteome levels. The riboswitch detected at genome position 2,792,551 did not possess any genes positioned immediately downstream on the same strand, but instead had a cluster of four genes in the antisense orientation, three of which were significantly upregulated by cobalamin at the transcriptome level. This orientation suggests that the riboswitch may regulate gene expression by a nonclassical mechanism (Mellin et al., 2013). For example, this riboswitch sequence could be in the 5' region of an antisense RNA (Thomason & Storz, 2010) spanning the downstream antisense genes that is prematurely terminated in the presence of cobalamin, thereby potentially explaining the cobalamin-associated upregulation of the three more distant of the downstream antisense genes.

In summary, while our analyses leave a number of questions unanswered regarding the pathway and regulation of propionate production by *P. ruminicola*, our results suggest that gene candidates of the pathway of propionate production are not under the direct control of a cobalamin-regulated riboswitch. The mechanisms by which methylmalonyl-CoA mutase genes and proteins are upregulated by cobalamin will be the subject of further investigation, as well as the potential formation of propionate *via* propionyl phosphate by these bacteria suggested by our transcriptome data.



## STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Sam Mahoney-Kurpe
Name/title of Primary Supervisor:	Prof. Patrick Biggs
In which chapter is the manuscript /published work:	Chapter 4 (Section 4.3)
<p>Please select one of the following three options:</p> <p><input type="radio"/> The manuscript/published work is published or in press</p> <ul style="list-style-type: none"> <li>• Please provide the full reference of the Research Output:</li> </ul> <p><input type="radio"/> The manuscript is currently under review for publication – please indicate:</p> <ul style="list-style-type: none"> <li>• The name of the journal: mSystems (American Society for Microbiology)</li> <li>• The percentage of the manuscript/published work that was contributed by the candidate: 90.00</li> <li>• Describe the contribution that the candidate has made to the manuscript/published work: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualisation, Writing-original draft</li> </ul> <p><input checked="" type="radio"/> It is intended that the manuscript will be published, but it has not yet been submitted to a journal</p>	
Candidate's Signature:	Sam Mahoney-Kurpe <small>Digitally signed by Sam Mahoney-Kurpe, DN: cn=Sam Mahoney-Kurpe, o=Massey University, ou=University of New Zealand, email=s.mahoney@massey.ac.nz, c=NZ</small>
Date:	18-Feb-2022
Primary Supervisor's Signature:	Prof Patrick Biggs <small>Digitally signed by Prof Patrick Biggs, DN: cn=Prof Patrick Biggs, o=Massey University, ou=University of New Zealand, email=p.biggs@massey.ac.nz, c=NZ</small>
Date:	18-Feb-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/publication or collected as an appendix at the end of the thesis.

## 4.4 Discussion and conclusion

In this chapter, a combined multi-omics approach was used to guide both the selection and characterisation of dominant propionate producing strains of the Hungate1000 collection. Mining of existing metatranscriptome datasets of animals fed forage-based diets for reads mapping to enzymes involved in the conversion of succinate to propionate identified that the majority of functional categories examined were dominated by *Prevotella*, particularly of the *Prevotella 1* genus-level group (Henderson et al., 2019). Following this result, strains spanning this group were selected from the Hungate1000 collection for inclusion in a phenotypic screen, and one representative strain (KHP1) that demonstrated cobalamin-dependent propionate production but grew well in both its presence and absence, was selected for complete genome sequencing and use in a comparative transcriptomics and proteomics experiment.

Based on their dominance in the functional categories analysed, strains of the *Prevotella 1* genus-level group were selected for further analysis in this study regarding their propionate production. However, the predominance of RNA-seq reads assigned to the Interpro2GO categories 'Succinate CoA-transferase' (IPR017821) mapping to *Succiniclasticum ruminis* also implies its involvement in ruminal propionate production, and therefore this species represents another interesting candidate to characterise further regarding its propionate production. Strains of this species are known to require succinate as a sole substrate for their growth, from which they produce propionate (van Gylswyk, 1995). It would be interesting to perform co-culture experiments between KHP1 and *S. ruminis*, in order to assess whether *P. ruminicola* converts the succinate it produces into propionate itself, or whether it is transferred to *S. ruminis* for further decarboxylation. Co-culture experiments of these strains could also identify potential synergism between these strains for propionate production.

One major limitation of the metatranscriptome analyses carried out is that for some steps of the pathway, the existence of multiple different orthology categories with the potential to carry out some of the reactions of the conversion of succinate to propionate complicated

their analysis. For example, cross-reactivity has been shown between different short-chain fatty acid-producing kinases (Charrier et al., 2006). This was exemplified by the *Prevotella* genomes analysed not possessing any genes explicitly annotated as a propionate CoA-transferase or phosphate propionyltransferase/propionate kinase, thereby explaining the absence of *Prevotella* reads from the propionate CoA-transferase KEGG orthology category (K01026) (see Figure 4.2). However, as the specific genes involved are better characterised, mining of reads aligning to these sequences will be able to be used to detect propionate production by dominant groups in rumen metagenome/metatranscriptome datasets with greater resolution.

For the transcriptomics carried out in this chapter (Section 4.3) a stranded library preparation kit was used which allows for the strand-specificity of the sequenced mRNA to be known, enabling insights into antisense transcription and its role in prokaryotic gene regulation (Mills et al., 2013). As such, further analyses of the RNA-seq data are planned prior to the submission of the manuscript, particularly examining loci with intriguing sense/antisense gene architectures such as that of the identified cobalamin riboswitch containing an island of downstream but antisense genes (see Figure 4.8b), and the locus containing methylmalonyl-CoA mutase genes with upstream but antisense methylmalonyl-CoA decarboxylase genes (see Figure 4.9b).

In conclusion, this chapter demonstrates how further analyses of existing meta'omic datasets can be mined as a means of selecting strains from the Hungate1000 collection predicted to have roles in key rumen metabolic processes of interest. The approach used of mining meta'omic datasets for specific functions is applicable to many other rumen metabolic processes of interest for which relevant functional categories are defined, which will facilitate future studies to better understand members involved in key rumen metabolic processes.

## 5 Conclusion and future directions

## 5.1 Conclusion

The curation of the Hungate1000 collection now provides a wide-encompassing catalogue of rumen bacteria, many of which await phenotypic characterisation and taxonomic classification. The work presented in this thesis provides two examples of approaches taken to both characterise novel groups of rumen bacteria, as well as further our understanding of the mechanisms by which an already well-studied rumen microbial group produces propionate. In the case of the R-7 group, phylogenetic and phenotypic characterisation enabled the proposal of the first two species of a novel genus (*Aristaeella* gen. nov.) and family (*Aristaeellaceae* fam. nov.), and show that the group may potentially play a role in fibre degradation, particularly in the degradation of xylan and pectin. Notably however, their concurrent production of hydrogen and ethanol suggests that enhanced fibre degradation may potentially result in increased methane production. Nonetheless, the genomic diversity exhibited by strains of this group coupled with the different rates of hydrogen production seen between R-7 and WTE2008 suggests the possibility that some R-7 group strains may be able to enhance fibre degradation without producing large amounts of methanogenic substrates.

The results of the second half of this thesis provide some insights into propionate production by *P. ruminicola*, but also raises further questions that will require more targeted analyses to answer. We show that while the *P. ruminicola* KHP1 genome possesses four ‘cobalamin’ family riboswitches that altered the expression of their neighbouring genes, these were not located near gene candidates involved in propionate production. This suggests that these riboswitches are not involved in regulating propionate production. Nonetheless, while the underlying regulatory mechanisms remain elusive, we demonstrate the upregulation of 7 of the total 14 identified candidate propionate pathway genes by cobalamin, although of these, the only genes that consistently showed increased transcript and protein abundances were both subunits of the cobalamin-dependent enzyme methylmalonyl-CoA mutase. Further analyses of these upregulated loci may provide additional insights into the regulation of propionate production in ruminal *Prevotella*, which may reveal new strategies to promote propionate production. The continued characterisation of the great variety of rumen



microorganisms that have now been isolated will inevitably further our understanding of the structure and functions of the rumen microbiota, paving the way for the development of novel strategies for its manipulation to enhance host health and productivity while mitigating damaging environmental emissions.

## 5.2 Future directions

The work carried out in this thesis demonstrates the value of the Hungate1000 culture collection as a tool to direct future studies aimed at better understanding new rumen bacterial groups, as well as to further explore the metabolism of well-characterised strains predicted to be dominant players in metabolic processes of interest. This section outlines some of the further avenues of work stemming from the outcomes of this thesis.

### 5.2.1 Future work characterising the R-7 group

At the beginning of this project, strains R-7 and WTE2008 were the only pure cultures from the R-7 group that were available for characterisation. However, following the genome sequencing of strains FE2010, FE2011 and XBB3002 (Mahoney-Kurpe et al., 2021), these represent new candidates of the R-7 group to be characterised and classified as new species. Based on their GTDB taxonomic assignments (Parks et al., 2018), FE2010 and FE2011 are representatives of the same species. Strain XBB3002 differs considerably to the previous strains and will be particularly interesting to characterise as it was unassigned to any species level clusters as defined by GTDB, indicating its uniqueness from not only other genome-sequenced isolates but also from metagenome-assembled genomes, of which many have been reconstructed from rumen samples (Stewart et al., 2018; Stewart et al., 2019). The continued characterisation of strains of this group will further our understanding of their metabolic potential in the rumen, which may identify strains displaying particularly promising characteristics that could be used to enhance ruminal fibre degradation.

Genome analyses of the R-7 group were limited to assessment of the predicted genomic relatedness of different strains, and reporting general genome characteristics. As a greater number of R-7 group isolates continue to be genome-sequenced (Mahoney-Kurpe et al., 2021), additional comparative genomic analyses of predicted functional characteristics will yield valuable insights, as has been carried out between all Hungate1000 strains (Seshadri et al., 2018). Given the ability of strains to degrade plant cell wall polysaccharides, it would be interesting to compare the carbohydrate active enzyme (CAZyme) profiles of the strains, and assess how these relate to the ability of strains to degrade various plant polysaccharides. For example, comparative genomics of 46 rumen *Butyrivibrio* and *Pseudobutyrvibrio* strains found a great diversity of different carbohydrate-active enzymes but with similar polysaccharide degradation profiles, suggestive of some level of functional redundancy between many of these enzymes. Interestingly, this analysis also showed that a number of strains lacked an enolase gene responsible for one of the steps of the Embden-Meyerhof-Parnas (EMP) pathway of conversion of glucose to pyruvate, which suggests that an alternative pathway is used for this process (Palevich et al., 2019b). Screening of R-7 group strain genomes for the presence of hydrogenase genes (Søndergaard et al., 2016) and genes that encode enzymes involved in ethanol production could also identify genes involved in these metabolic processes that may be able to differentiate strains that produce greater amounts of these end products than others.

Despite the ability of R-7 group strains to degrade the xylan and pectin, and R-7 itself being isolated from the cellulose-adherent component of rumen contents, our results show that neither R-7 nor WTE2008 could degrade cellulose. However, co-culture of these strains with cellulose-degrading rumen strains may identify synergistic interactions relating to fibre degradation. Many instances of synergism have been identified between the dominant rumen cellulolytic bacterium *Fibrobacter succinogenes* and numerous non-fibrolytic bacteria including *Prevotella ruminicola* (Fondevila & Dehority, 1996), *Selenomonas ruminantium* (Sawanon et al., 2011) and members of the uncultured U2 group (Fukuma et al., 2012). As such, similar synergistic effects could potentially exist between R-7 and cellulolytic strains when cultured together. Moreover, the rumen fungi are well recognised as potent fibre and cellulose degraders (Hagen et al., 2021). An understanding of the potential synergism of the

fungi with R-7 group strains for fibre degradation, particularly on recalcitrant fibrous substrates, may highlight new strategies for ruminants to utilise poor quality fibrous feeds.

Similarly to the strategy utilised in chapter four to mine the metatranscriptome data to identify taxa involved in particular functions of interest (see Section 4.2), the approach can also be applied whereby meta'omic reads can be mined based instead on their taxonomic classification, in order to reveal the predicted functions of the taxonomic group of interest. With the increasing number of rumen meta'omic datasets now available, meta-analyses of reads that map to R-7 group genomes could provide insights into the metabolism of the R-7 group under a variety of different conditions. Given that the R-7 group has been identified as a major group involved in the expression of hydrogenases (Greening et al., 2019), it could be particularly interesting to assess the expression of these in datasets of animals differing in their methane production (Shi et al., 2014).

Numerous examples of hydrogen transfer from a producing to a consuming species have been identified in the rumen (Stams & Plugge, 2009). Given the hydrogen and ethanol production of R-7 group strains, co-culture experiments of these strains with rumen methanogens would be of interest as some methanogens are able to use these products as substrates for growth (Leahy et al., 2010). Such experiments could elucidate whether hydrogen and ethanol production by the R-7 group promotes ruminal methanogenesis. It should also be noted that while R-7 and WTE2008 produce ethanol in batch culture, it is possible that this may not occur *in vivo* or in co-culture with methanogens that keep the hydrogen partial pressure low, as has been shown in *Ruminococcus albus* (Ianotti et al., 1973)

## 5.2.2 Future work characterising propionate production in *P. ruminicola*

Results of the comparative multi-omics experiment demonstrated that both subunits of methylmalonyl-CoA mutase showed increase abundances in the presence of cobalamin, and

were the only candidate genes involved in the conversion of succinate to propionate to be differentially abundant in both the transcriptome and proteome datasets. Our inability to identify a riboswitch associated with the differential expression begs the question as to what alternative mechanism is responsible. While both methylmalonyl-CoA mutase subunits were included in the analyses performed, other associated proteins may also be involved and should be included in future analyses, such as the GTPase MeaB that complexes with methylmalonyl-CoA mutase and functions to prevent against its inactivation (Korotkova & Lindstrom, 2003). In addition to gene regulation through riboswitch mechanisms, there is increasing knowledge of the role of cobalamin as a co-factor of regulatory proteins (Klug, 2014). For example, in the phototrophic bacterium *Rhodobacter capsulatus*, vitamin B<sub>12</sub> has been shown to act as a cofactor of an anti-repressor of the transcriptional regulator *CrtJ*, which regulates gene expression of enzymes involved in the biosynthesis of haem, carotenoids and light-harvesting proteins (Cheng et al., 2014). Similarly, long known for their roles in eukaryotic gene regulation, the roles of post-translational modifications in prokaryotes are becoming better understood (Pisithkul et al., 2015), and could also be involved in mediating some of these effects.

One of the major findings of the comparative omics' experiments was that of the three routes of propionyl-CoA conversion to propionate identified in the KHP1 genome, the only pathway upregulated in the presence of vitamin B<sub>12</sub> was mediated by phosphate butyryltransferase and butyrate kinase, for which genes are co-located on the KHP1 genome. Particularly given that KHP1 does not produce butyrate, this suggests that the conversion of propionyl-CoA to propionate by *P. ruminicola* may occur *via* the production of a propionyl phosphate intermediate by these enzymes (Louis & Flint, 2017). To better explore this possibility, biochemical characterisation of the purified enzymes of these genes could be carried out. For example, in *E. coli*, characterisation of recombinant enzymes in close proximity to methylmalonyl-CoA mutase genes identified a putative operon containing methylmalonyl-CoA mutase, methylmalonyl-CoA decarboxylase, and succinate:propionate CoA-transferase, which are together capable of converting succinate to propionate (Haller et al., 2000). Additionally, the recent development of a system enabling targeted modification in the human gut bacterium *Prevotella copri* with expected applicability in other *Prevotella* species

(Li et al., 2021) suggest that this system may enable the generation of mutant strains containing disrupted or over-expressed candidate propionate pathway genes that could be used as a means of demonstrating the contribution of specific candidate genes to propionate production (Ma et al., 2021; Zhao et al., 2019).

Despite possessing some cobalamin biosynthesis genes (see Figure 4.8a), the finding that *P. ruminicola* could not produce propionate in the absence of exogenous cobalamin suggests that the bacterium is unable to produce a form of cobalamin sufficient for methylmalonyl-CoA mutase activity, and must therefore rely on other members of the microbiota that produce cobalamin, such as *Selenomonas ruminantium* (Dryden et al., 1962). It would therefore be interesting to perform co-culture experiments with *P. ruminicola* and such cobalamin-synthesising strains, which may participate in synergistic interactions with regard to propionate production.

### 5.3 Towards the development of synthetic rumen microbial consortia

As outlined in the literature review, a major hurdle for strategies aimed at manipulating the rumen microbiota stems from its resilience in structure once established (Weimer, 2015), resulting in manipulation attempts such as inoculation with individual beneficial strains often only showing transient effects, if any, that subside shortly following discontinuing treatment. However, numerous studies have shown that multi-strain or multi-species probiotics, also referred to as synthetic microbial consortia (Che & Men, 2019) tend to show greater success than is seen when single probiotic strains are applied (Timmerman et al., 2004), and have a greater likelihood of successful establishment in the host microbiota resulting in long-term effects (Famularo et al., 1999). While it should be noted that another recent meta-analysis has disputed the greater efficacy of multi-strain probiotics (McFarland, 2021), this is likely a reflection of our current lack of understanding of the interactions between different strains to allow the successful formulation of synergistically acting consortia rather than an indication of any limitation of their potential compared with single-strain probiotic treatments.

Indeed, an emphasised outcome envisioned from the curation of the Hungate1000 collection was that it would facilitate the development of synthetic microbial consortia (Seshadri et al., 2018). As a greater number of cultured rumen bacteria are characterised, firstly in pure cultures and then as co-cultures, synthetic rumen microbial consortia of increasing complexity may be able to be formulated that together promote certain desirable metabolic outcomes. For example, in a mucosal simulator of the human digestive tract, a consortium of seven propionate-producing strains successfully restored propionate production of a simulated human colon microbiota following dysbiosis induced by antibiotic treatment (El Hage et al., 2019). Similarly, culturing of different lignocellulose-degrading soil bacteria has been used to identify consortia especially efficient in fibre degradation (Puentes-Téllez & Falcao Salles, 2018). Co-cultures formulated of a variety of specific strains that collectively complement each other with regard to a desired metabolic outcome may represent promising direct fed microbial strategies for the targeted manipulation of the rumen microbiota, particularly when inoculations are applied very early in life prior to the establishment of a stable rumen microbiota that is resilient to change (Bickhart & Weimer, 2018; Yáñez-Ruiz et al., 2015).

## 6 Bibliography

- Abbas, W., Howard, J. T., Paz, H. A., Hales, K. E., Wells, J. E., Kuehn, L. A., . . . Fernando, S. C. (2020). Influence of host genetics in shaping the rumen bacterial community in beef cattle. *Scientific Reports*, *10*(1), 15101.
- Adams, J. C., Gazaway, J. A., Brailsford, M. D., Hartman, P. A., & Jacobson, N. L. (1966). Isolation of bacteriophages from the bovine rumen. *Experientia*, *22*(11), 717-718.
- Ajmone-Marsan, P., Garcia, J. F., & Lenstra, J. A. (2010). On the origin of cattle: How aurochs became cattle and colonized the world. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*(4), 148-157.
- Akin, D. E., & Borneman, W. S. (1990). Role of rumen fungi in fiber degradation. *Journal of Dairy Science*, *73*(10), 3023-3032.
- Allison, S. D., & Martiny, J. B. H. (2008). Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences, USA*, *105*(1), 11512-11519.
- Alonso, B. L., Irigoyen von Sierakowski, A., Sáez Nieto, J. A., Rosel, A. B. (2017). First report of human infection by *Christensenella minuta*, a gram-negative, strictly anaerobic rod that inhabits the human intestine. *Anaerobe*, *44*, 124-125.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389-3402.
- Andries, J. I., Buysse, F. X., De Brabander, D. L., & Cottyn, B. G. (1987). Isoacids in ruminant nutrition: their role in ruminal and intermediary metabolism and possible influences on performances -a review. *Animal Feed Science and Technology*, *18*(3), 169-180.
- Aravind, L., & Ponting, C. P. (1999). The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiology Letters*, *176*(1), 111-116.
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16-W21.
- Arntzen, M.Ø., Karlskås, I. L., Skaugen, M., Eijsink, V. G. H., Mathiesen, G. (2015) Proteomic Investigation of the response of *Enterococcus faecalis* V583 when cultivated in urine. *PLOS One*, *10*(4), e0126694.
- Artzi, L., Bayer, E. A., & Morais, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nature Reviews Microbiology*, *15*(2), 83-95.
- Aschenbach, J. R., Kristensen, N. B., Donkin, S. S., Hammon, H. M., & Penner, G. B. (2010). Gluconeogenesis in dairy cows: the secret of making sweet milk from sour dough. *IUBMB Life*, *62*(12), 869-877.
- Avgustin, G., Wallace, R. J., & Flint, H. J. (1997). Phenotypic diversity among ruminal isolates of *Prevotella ruminicola*: proposal of *Prevotella brevis* sp. nov., *Prevotella bryantii* sp. nov., and *Prevotella albensis* sp. nov. and redefinition of *Prevotella ruminicola*. *International Journal of Systematic Bacteriology*, *47*(2), 284-288.
- Barcroft, J., McAnally, R. A., & Phillipson, A. T. (1944). Absorption of volatile acids from the alimentary tract of the sheep and other animals. *Journal of Experimental Biology*, *20*(2), 120.
- Bashiardes, S., Zilberman-Schapira, G., & Elinav, E. (2016). Use of metatranscriptomics in microbiome research. *Bioinformatics and Biology Insights*, *10*, 19-25.
- Bauman, D. E., Davis, C. L., & Bucholtz, H. F. (1971). Propionate production in the rumen of cows fed either a control or high-grain, low-fiber diet. *Journal of Dairy Science*, *54*(9), 1282-1287.
- Bayer, E. A., Morag, E., & Lamed, R. (1994). The cellulosome- a treasure-trove for biotechnology. *Trends in Biotechnology*, *12*(9), 379-386.
- Beauchemin, K. A. (1992). Effects of ingestive and ruminative mastication on digestion of forage by cattle. *Animal Feed Science and Technology*, *40*(1), 41-56.



- Beauchemin, K. A., Ungerfeld, E. M., Eckard, R. J., & Wang, M. (2020). Review: Fifty years of research on rumen methanogenesis: lessons learned and future challenges for mitigation. *Animal*, *14*(S1), s2-s16.
- Belanche, A., de la Fuente, G., & Newbold, C. J. (2015). Effect of progressive inoculation of fauna-free sheep with holotrich protozoa and total-fauna on rumen fermentation, microbial diversity and methane emissions. *FEMS Microbiology Ecology*, *91*(3).
- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289-300.
- Bera-Maillet, C., Mosoni, P., Kwasiborski, A., Suau, F., Ribot, Y., & Forano, E. (2009). Development of a RT-qPCR method for the quantification of *Fibrobacter succinogenes* S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. *Journal of Microbiological Methods*, *77*(1), 8-16.
- Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological Reviews*, *70*(2), 567-590.
- Bernalier, A., Fonty, G., Bonnemoy, F., & Gouet, P. (1993). Inhibition of the cellulolytic activity of *Neocallimastix frontalis* by *Ruminococcus flavefaciens*. *Journal of General Microbiology*, *139*(4), 873-880.
- Bervoets, I., & Charlier, D. (2019). Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews*, *43*(3), 304-339.
- Beukes, P. C., Gregorini, P., Romera, A. J., Levy, G., & Waghorn, G. C. (2010). Improving production efficiency as a strategy to mitigate greenhouse gas emissions on pastoral dairy farms in New Zealand. *Agriculture, Ecosystems & Environment*, *136*(3), 358-365.
- Bickhart, D. M., & Weimer, P. J. (2018). Symposium review: Host-rumen microbe interactions may be leveraged to improve the productivity of dairy cows. *Journal of Dairy Science*, *101*(8), 7680-7689.
- Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S., & Aldrich, H. C. (1999). The propanediol utilization (*pdu*) operon of *Salmonella enterica* serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B<sub>12</sub>-dependent 1, 2-propanediol degradation. *Journal of Bacteriology*, *181*(19), 5967-5975.
- Bonhomme, A. (1990). Rumen ciliates: their metabolism and relationships with bacteria and their hosts. *Animal Feed Science and Technology*, *30*(3), 203-266.
- Brulc, J. M., Antonopoulos, D. A., Miller, M. E., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., . . . White, B. A. (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*(6), 1948-1953.
- Bryant, M. P. (1973). Nutritional requirements of the predominant rumen cellulolytic bacteria. *Federation Proceedings*, *32*(7), 1809-1813.
- Bryant, M. P., & Burkey, L. A. (1953). Cultural methods and some characteristics of some of the more numerous groups of bacteria in the bovine rumen. *Journal of Dairy Science*, *36*(3), 205-217.
- Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*, 59.
- Buckel, W., Thauer, R. K. (2018). Flavin-based electron-bifurcation, ferredoxin, flavodoxin, and anaerobic respiration with protons (Ech) or NAD<sup>+</sup> (Rnf) as electron acceptors: a historical review. *Frontiers in Microbiology*, *9*, 401.
- Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., & Thomas, T. (2011). Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences, USA*, *108*(34), 14288-14293.

- Burnet, M. C., Dohnalkova, A. C., Neumann, A. P., Lipton, M. S., Smith, R. D., Suen, G., & Callister, S. J. (2015). Evaluating models of cellulose degradation by *Fibrobacter succinogenes* S85. *PLOS One*, *10*(12), e0143809.
- Burton, R. A., Gidley, M. J., & Fincher, G. B. (2010). Heterogeneity in the chemistry, structure and function of plant cell walls. *Nature Chemical Biology*, *6*(10), 724-732.
- Button, D. K., Schut, F., Quang, P., Martin, R., & Robertson, B. R. (1993). Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. *Applied and Environmental Microbiology*, *59*(3), 881-891.
- Calsamiglia, S., Busquet, M., Cardozo, P. W., Castillejos, L., & Ferret, A. (2007). Invited review: essential oils as modifiers of rumen microbial fermentation. *Journal of Dairy Science*, *90*(6), 2580-2595.
- Calsamiglia, S., Ferret, A., & Devant, M. (2002). Effects of pH and pH fluctuations on microbial fermentation and nutrient flow from a dual-flow continuous culture system. *Journal of Dairy Science*, *85*(3), 574-579.
- Campbell, M. M., & Sederoff, R. R. (1996). Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants). *Plant Physiology*, *110*(1), 3-13.
- Carberry, C. A., Kenny, D. A., Han, S., McCabe, M. S., & Waters, S. M. (2012). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. *Applied and Environmental Microbiology*, *78*(14), 4949-4958.
- Cassab, G. I. (1998). Plant cell wall proteins. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*, 281-309.
- Chalupa, W. (1977). Manipulating rumen fermentation. *Journal of Animal Science*, *45*(3), 585-599.
- Chan, C. W., & Mondragón, A. (2020). Crystal structure of an atypical cobalamin riboswitch reveals RNA structural adaptability as basis for promiscuous ligand binding. *Nucleic Acids Research*, *48*(13), 7569-7583.
- Chang, Y.-C., Hu, Z., Rachlin, J., Anton, B. P., Kasif, S., Roberts, R. J., & Steffen, M. (2016). COMBREX-DB: an experiment centered database of protein function: knowledge, predictions and knowledge gaps. *Nucleic Acids Research*, *44*(D1), D330-D335.
- Charrier, C., Duncan, G. J., Reid, M. D., Rucklidge, G. J., Henderson, D., Young, P., . . . Louis, P. (2006). A novel class of CoA-transferase involved in short-chain fatty acid metabolism in butyrate-producing human colonic bacteria. *Microbiology*, *152*(1), 179-185.
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925-1927.
- Che, S., & Men, Y. (2019). Synthetic microbial consortia for biosynthesis and biodegradation: promises and challenges. *Journal of Industrial Microbiology & Biotechnology*, *46*(9), 1343-1358.
- Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., . . . Kyrpides, N. C. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Research*, *49*(1), 751-763.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890.
- Cheng, Y. F., Edwards, J. E., Allison, G. G., Zhu, W. Y., & Theodorou, M. K. (2009). Diversity and activity of enriched ruminal cultures of anaerobic fungi and methanogens grown together on lignocellulose in consecutive batch culture. *Bioresource Technology*, *100*(20), 4821-4828.
- Cheng, Z., Li, K., Hammad, L. A., Karty, J. A., & Bauer, C. E. (2014). Vitamin B<sub>12</sub> regulates photosystem gene expression via the CrtJ antirepressor AerR in *Rhodobacter capsulatus*. *Molecular Microbiology*, *91*(4), 649-664.
- Christopherson, M. R., Dawson, J. A., Stevenson, D. M., Cunningham, A. C., Bramhacharya, S., Weimer, P. J., . . . Suen, G. (2014). Unique aspects of fiber degradation by the ruminal

- ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. *BMC genomics*, 15(1), 1066.
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., . . . de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422-1423.
- Coico, R. (2006). Gram staining. *Current Protocols in Microbiology*, 00(1), A.3C.1-A.3C.2.
- Coleman, G. S. (1992). The rate of uptake and metabolism of starch grains and cellulose particles by *Entodinium* species, *Eudiplodinium maggii*, some other entodiniomorphid protozoa and natural protozoal populations taken from the ovine rumen. *Journal of Applied Bacteriology*, 73(6), 507-513.
- Comtet-Marre, S., Parisot, N., Lepercq, P., Chaucheyras-Durand, F., Masoni, P., Peyretailade, E., . . . Forano, E. (2017). Metatranscriptomics reveals the active bacterial and eukaryotic fibrolytic communities in the rumen of dairy cow fed a mixed diet. *Frontiers in Microbiology*, 8, 67.
- Conway, T., Creecy, J. P., Maddox, S. M., Grissom, J. E., Conkle, T. L., Shadid, T. M., . . . Wanner, B. L. (2014). Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio*, 5(4), e01442-14.
- Cotta, M. A., & Russell, J. B. (1982). Effect of peptides and amino acids on efficiency of rumen bacterial protein synthesis in continuous culture. *Journal of Dairy Science*, 65(2), 226-234.
- Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12), 1367-1372.
- Cox, M. S., Deblois, C. L., & Suen, G. (2021). Assessing the response of ruminal bacterial and fungal microbiota to whole-rumen contents exchange in dairy cows. *Frontiers in Microbiology*, 12, 665776.
- Creevey, C. J., Kelly, W. J., Henderson, G., & Leahy, S. C. (2014). Determining the culturability of the rumen bacterial microbiome. *Microbial Biotechnology*, 7(5), 467-479.
- Cristobal-Carballo, O., McCoard, S. A., Cookson, A. L., Laven, R. A., Ganesh, S., Lewis, S. J., & Muetzel, S. (2021). Effect of divergent feeding regimes during early life on the rumen microbiota in calves. *Frontiers in Microbiology*, 12, 711040.
- Dai, X., Tian, Y., Li, J., Su, X., Wang, X., Zhao, S., . . . Huang, L. (2015). Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. *Applied and Environmental Microbiology*, 81(4), 1375-1386.
- Dale, H. E., Stewart, R. E., & Brody, S. (1954). Rumen temperature. I. Temperature gradients during feeding and fasting. *The Cornell Veterinarian*, 44(3), 368-374.
- Dassa, B., Borovok, I., Ruimy-Israeli, V., Lamed, R., Flint, H. J., Duncan, S. H., . . . Bayer, E. A. (2014). Rumen cellulosomes: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PLOS One*, 9(7), e99221.
- Dehority, B. A. (1966). Characterisation of several bovine rumen bacteria isolated with a xylan medium. *Journal of Bacteriology*, 91(5), 1724-1729.
- Dehority, B. A., & Tirabasso, P. A. (1998). Effect of ruminal cellulolytic bacterial concentrations on *in situ* digestion of forage cellulose. *Journal of Animal Science*, 76(11), 2905-2911.
- Delogu, F., Kunath, B. J., Evans, P. N. Arntzen, M. Ø., Hvidsten, T. R., Pope, P. B. (2020). Integration of absolute multi-omics reveals dynamic protein-to-RNA ratios and metabolic interplay within mixed-domain microbiomes. *Nature Communications*, 11, 4708.
- Dijkstra, J. (1994). Production and absorption of volatile fatty acids in the rumen. *Livestock Production Science*, 39(1), 61-69.
- Dill-McFarland, K. A., Breaker, J. D., & Suen, G. (2017). Microbial succession in the gastrointestinal tract of dairy cows from 2 weeks to first lactation. *Scientific Reports*, 7, 40864.
- Dodd, D., Moon, Y. H., Swaminathan, K., Mackie, R. I., & Cann, I. K. (2010). Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. *The Journal of Biological Chemistry*, 285(39), 30261-30273.

- Dryden, L. P., Hartman, A. M., Bryant, M. P., Robinson I. M., Moore, L. A. (1962). Production of vitamin B<sub>12</sub> and vitamin B<sub>12</sub> analogues by pure cultures of ruminal bacteria. *Nature*, *195*, 201-202.
- Duffield, T. F., Merrill, J. K., & Bagg, R. N. (2012). Meta-analysis of the effects of monensin in beef cattle on feed efficiency, body weight gain, and dry matter intake. *Journal of Animal Science*, *90*(12), 4583-4592.
- Eadie, J. M. (1962). Inter-relationships between certain rumen ciliate protozoa. *Microbiology*, *29*(4), 579-588.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.
- Edwards, J. E., Forster, R. J., Callaghan, T. M., Dollhofer, V., Dagar, S. S., Cheng, Y., . . . Smidt, H. (2017). PCR and omics based techniques to study the diversity, ecology and biology of anaerobic fungi: insights, challenges and opportunities. *Frontiers in Microbiology*, *8*, 1657.
- Edwards, J. E., McEwan, N. R., J., T. A., & Wallace, R. J. (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek*, *86*(3), 263-281.
- El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R., & Van de Wiele, T. (2019). Propionate-producing consortium restores antibiotic-induced dysbiosis in a dynamic *in vitro* model of the human intestinal microbial ecosystem. *Frontiers in Microbiology*, *10*.
- Ellison, M. J., Conant, G. C., Cockrum, R. R., Austin, K. J., Truong, H., Becchi, M., . . . Cammack, K. M. (2014). Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Research*, *21*(2), 115-125.
- Emerson, E. L., Weimer, P. J. (2017). Fermentation of model hemicelluloses by *Prevotella* strains and *Butyrivibrio fibrisolvens* in pure culture and in ruminal enrichment cultures. *Applied and Environmental Microbiology*, *101*, 4269-4278.
- Famularo, G., De Simone, C., Matteuzzi, D., & Pirovano, F. (1999). Traditional and high potency probiotic preparations for oral bacteriotherapy. *BioDrugs*, *12*(6), 455-470.
- FAO (2019). The state of the world's biodiversity for food and agriculture, J. Bélanger & D. Pilling (eds.). FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome. 572 pp. (<http://www.fao.org/3/CA3129EN/CA3129EN.pdf>) Licence: CC BY-NC-SA 3.0 IGO.
- Finlay, B. J., Esteban, G., Clarke, K. J., Williams, A. G., Embley, T. M., & Hirt, R. P. (1994). Some rumen ciliates have endosymbiotic methanogens. *FEMS Microbiology Letters*, *117*(2), 157-161.
- Fondevila, M., & Dehority, B. A. (1996). Interactions between *Fibrobacter succinogenes*, *Prevotella ruminicola*, and *Ruminococcus flavefaciens* in the digestion of cellulose from forages. *Journal of Animal Science*, *74*(3), 678-684.
- Fonty, G., Gouet, P., Jouany, J. P., & Senaud, J. (1987). Establishment of the microflora and anaerobic fungi in the rumen of lambs. *Microbiology*, *133*, 1835-1843.
- Fonty, G., Senaud, J., Jouany, J. P., & Gouet, P. (1988). Establishment of ciliate protozoa in the rumen of conventional and conventionalized lambs: influence of diet and management conditions. *Canadian Journal of Microbiology*, *34*(3), 235-241.
- Franz, T., Lee, X. (2012). The OASIS<sup>®</sup> HLB  $\mu$ Elution plate as a one-step platform for manual high-throughput in-gel digestion of proteins and peptide desalting. *Proteomics*, *12*, 2487-2492.
- Fukuma, N., Koike, S., & Kobayashi, Y. (2012). Involvement of recently cultured group U2 bacterium in ruminal fiber digestion revealed by coculture with *Fibrobacter succinogenes* S85. *FEMS Microbiology Letters*, *336*(1), 17-25.
- Gijzen, H. J., Lubberding, H. J., Gerhardus, M. J. T., & Vogels, G. D. (1988). Contribution of rumen protozoa to fibre degradation and cellulase activity *in vitro*. *FEMS Microbiology Letters*, *53*(1), 35-43.
- Gilbert, R. A., Kelly, W. J., Altermann, E., Leahy, S. C., Minchin, C., Ouwkerk, D., & Klieve, A. V. (2017). Toward understanding phage:host interactions in the rumen; complete genome sequences of lytic phages infecting rumen bacteria. *Frontiers in Microbiology*, *8*, 2340.

- Goopy, J. P., Donaldson, A., Hegarty, R., Vercoe, P. E., Haynes, F., Barnett, M., & Oddy, V. H. (2014). Low-methane yield sheep have smaller rumens and shorter rumen retention time. *The British Journal of Nutrition*, *111*(4), 578-585.
- Greening, C., Geier, R., Wang, C., Woods, L. C., Morales, S. E., McDonald, M. J., . . . Mackie, R. I. (2019). Diverse hydrogen production and consumption pathways influence methane production in ruminants. *The ISME journal*, *13*(10), 2617-2632.
- Gruninger, R. J., Puniya, A. K., Callaghan, T. M., Edwards, J. E., Youssef, N., Dagar, S. S., . . . Elshahed, M. S. (2014). Anaerobic fungi (phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiology Ecology*, *90*(1), 1-17.
- Guan, L. L., Nkrumah, J. D., Basarab, J. A., & Moore, S. S. (2008). Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiology Letters*, *288*(1), 85-91.
- Guzman, C. E., Wood, J. L., Egidi, E., White-Monsant, A. C., Semenec, L., Grommen, S. V. H., . . . Franks, A. E. (2020). A pioneer calf foetus microbiome. *Scientific Reports*, *10*(1), 17712.
- Hartinger, T., Zebeli, Q. (2021). The present role and new potentials of anaerobic fungi in ruminant nutrition. *Journal of Fungi*, *7*, 200.
- Hackmann, T. J., Ngugi, D. K., Firkins, J. L., Tao, J. (2017). Genomes of rumen bacteria encode atypical pathways for fermenting hexoses to short-chain fatty acids. *Environmental Microbiology*, *19*(11), 4670-4683.
- Hagen, L. H., Brooke, C. G., Shaw, C. A., Norbeck, A. D., Piao, H., Arntzen, M. Ø., . . . Hess, M. (2021). Proteome specialisation of anaerobic fungi during ruminal degradation of recalcitrant plant fiber. *The ISME Journal*, *15*, 421-434.
- Haller, T., Buckel, T., Rétey, J., & Gerlt, J. A. (2000). Discovering new enzymes and metabolic pathways: conversion of succinate to propionate by *Escherichia coli*. *Biochemistry*, *39*(16), 4622-4629.
- Hao, H., Cheng, G., Iqbal, Z., Ai, X., Hussain, H. I., Huang, L., . . . Yuan, Z. (2014). Benefits and risks of antimicrobial use in food-producing animals. *Frontiers in Microbiology*, *5*, 288-288.
- Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., GRC collaborators, & 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., Kumar, S., Forster, R. J., Kelly, W. J., Leahy, S. C., . . . Janssen, P. H. (2019). Improved taxonomic assignment of rumen bacterial 16S rRNA sequences using a revised SILVA taxonomic framework. *PeerJ*, *7*, e6496.
- Hess, M., Sczyrba, A., Egan, R., Kim, T. W., Chokhawala, H., Schroth, G., . . . Rubin, E. M. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, *331*(6016), 463-467.
- Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., . . . Zhang, N. (2007). A higher-level phylogenetic classification of the Fungi. *Mycological Research*, *111*(5), 509-547.
- Hillman, D. L. K., Yarlett, N., & Williams, A. G. (1989). Hydrogen production by rumen holotrich protozoa: effects of oxygen and implications for metabolic control by *in situ* conditions. *The Journal of Protozoology*, *36*(2), 205-213.
- Hilpert, W., Dimroth, P. (1983). Purification and characterization of a new sodium-transport decarboxylase. Methylmalonyl-CoA decarboxylase from *Veillonella alcalescens*. *European Journal of Biochemistry*, *132*, 579-587.
- Ho, Y. W., Abdullah, N., & Jalaludin, S. (1988). Penetrating structures of anaerobic rumen fungi in cattle and swamp buffalo. *Microbiology*, *134*(1), 177-181.
- Hobson, P. N., and Stewart, C. S. (1997). The rumen microbial ecosystem. 2nd ed. London, UK. Chapman and Hall.

- Hu, J., Fan, J., Sun, Z., & Liu, S. (2019). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, *36*(7), 2253-2255.
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115-2122.
- Hungate, R. E. (1944). Studies on cellulose fermentation: I. the culture and physiology of an anaerobic cellulose-digesting bacterium. *Journal of Bacteriology*, *48*(5), 499-513.
- Hungate, R. E. (1966). The rumen and its microbes. New York, USA. Academic Press.
- Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, *16*(1), 294.
- Huson, D. H., Beier, S., Flade, I., Górski, A., El-Hadidi, M., Mitra, S., . . . Tappu, R. (2016). MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Computational Biology*, *12*(6), e1004957.
- Huws, S. A., Creevey, C. J., Oyama, L. B., Mizrahi, I., Denman, S. E., Popova, M., . . . Morgavi, D. P. (2018). Addressing global ruminant agricultural challenges through understanding the rumen microbiome: past, present, and future. *Frontiers in Microbiology*, *9*(2161).
- Iannotti, E. L., Kafkewitz, D., Wolin, M. J., & Bryant, M. P. (1973). Glucose fermentation products in *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes*: changes caused by interspecies transfer of H<sub>2</sub>. *Journal of Bacteriology*, *114*(3), 1231-1240.
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114.
- Jami, E., Israel, A., Kotser, A., & Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *The ISME journal*, *7*(6), 1069-1079.
- Jami, E., & Mizrahi, I. (2012). Similarity of the ruminal bacteria across individual lactating cows. *Anaerobe*, *18*(3), 338-343.
- Janssen, P. H., Schuhmann, A., Mörschel, E., Rainey, F. A. (1997). Novel anaerobic ultramicrobacteria belonging to the *Verrucomicrobiales* lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Applied and Environmental Microbiology*, *63*(4), 1382-1388.
- Janssen, P. H. (2010). Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. *160*(1-2).
- Janssen, P. H., & Kirs, M. (2008). Structure of the archaeal community of the rumen. *Applied and Environmental Microbiology*, *74*(12), 3619-3625.
- Jewell, K. A., McCormick, C. A., Odt, C. L., Weimer, P. J., & Suen, G. (2015). Ruminal bacterial community composition in dairy cows is dynamic over the course of two lactations and correlates with feed efficiency. *Applied and Environmental Microbiology*, *81*(14), 4697-4710.
- Joblin, K. N. (2005). Methanogenic archaea. In H. P. S. Makkar & C. S. McSweeney (Eds.), *Methods in Gut Microbial Ecology for Ruminants* (pp. 47-53). Dordrecht: Springer Netherlands.
- Johnson Jr, J. E., Reyes, F. E., Polaski, J. T., & Batey, R. T. (2012). B<sub>12</sub> cofactors directly stabilize an mRNA regulatory switch. *Nature*, *492*(7427), 133-137.
- Jung, H. G., Samac, D. A., & Sarath, G. (2012). Modifying crops to increase cell wall digestibility. *Plant Science*, *185-186*, 65-77.
- Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., . . . Attwood, G. T. (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*(1), 56.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457-D462.

- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.
- Kelly, W. J., Leahy, S. C., Li, D., Perry, R., Lambie, S. C., Attwood, G. T., & Altermann, E. (2014). The complete genome sequence of the rumen methanogen *Methanobacterium formicicum* BRM9. *Standards in Genomic Sciences*, *9*, 15.
- Kelly, W. J., Pacheco, D. M., Li, D., Attwood, G. T., Altermann, E., & Leahy, S. C. (2016). The complete genome sequence of the rumen methanogen *Methanobrevibacter millerae* SM9. *Standards in Genomic Sciences*, *11*, 49.
- Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S., & 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*(1), 52-60.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907-915.
- King, E. E., Smith, R. P., St-Pierre, B., & Wright, A. 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*(16), 5682-5687.
- Kittelmann, S., Pinares-Patiño, C. S., Seedorf, H., Kirk, M. R., Ganesh, S., McEwan, J. C., & Janssen, P. H. (2014). Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLOS One*, *9*(7), e103171.
- Kittelmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I., & Janssen, P. H. (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLOS One*, *8*(2), e47879.
- Klug, G. (2014). Beyond catalysis: vitamin B<sub>12</sub> as a cofactor in gene regulation. *Molecular Microbiology*, *91*(4), 635-640.
- Koike, S., Handa, Y., Goto, H., Sakai, K., Miyagawa, E., Matsui, H., . . . Kobayashi, Y. (2010). Molecular monitoring and isolation of previously uncultured bacterial strains from the sheep rumen. *Applied and Environmental Microbiology*, *76*(6), 1887-1894.
- Koike, S., Pan, J., Kobayashi, Y., & Tanaka, K. (2003). Kinetics of *in sacco* fiber-attachment of representative ruminal cellulolytic bacteria monitored by competitive PCR. *Journal of Dairy Science*, *86*(4), 1429-1435.
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540-546.
- Korotkova, N., Lindstrom, M.E. (2004). MeaB is a component of the methylmalonyl-CoA mutase complex required for protection of the enzyme from inactivation. *Journal of Biological Chemistry*, *279*(14), 13652-13658.
- Krause, D. O., Bunch, R. J., Conlan, L. L., Kennedy, P. M., Smith, W. J., Mackie, R. I., & McSweeney, C. S. (2001). Repeated ruminal dosing of *Ruminococcus* spp. does not result in persistence, but changes in other microbial populations occur that can be measured with quantitative 16S-rRNA-based probes. *Microbiology*, *147*(7), 1719-1729.
- Krause, D. O., Denman, S. E., Mackie, R. I., Morrison, M., Rae, A. L., Attwood, G. T., & McSweeney, C. S. (2003). Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiology Reviews*, *27*(5), 663-693.
- Krause, D. O., Nagaraja, T. G., Wright, A. D. G., & Callaway, T. R. (2013). Board-invited review: rumen microbiology: leading the way in microbial ecology. *Journal of Animal Science*, *91*(1), 331-341.
- Krause, D. O., & Russell, J. B. (1996). How many ruminal bacteria are there? *Journal of Dairy Science*, *79*(8), 1467-1475.
- Kuhns, M., Trifunović, D., Huber, H., Volker, M. (2020). The Rnf complex is a Na<sup>+</sup> coupled respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*. *Communications Biology*, *3*, 431.

- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549.
- Kunath, B. J., Minniti, G., Skaugen, M., Hagen, L. H., Vaaje-Kolstad, G., Eijsink, V. G. H., . . . Arntzen, M. Ø. (2019). Metaproteomics: Sample Preparation and Methodological Considerations. In J.-L. Capelo-Martínez (Ed.), *Emerging Sample Treatments in Proteomics* (pp. 187-215). Cham: Springer International Publishing.
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100-3108.
- Lambie, S. C., Kelly, W. J., Leahy, S. C., Li, D., Reilly, K., McAllister, T. A., . . . Altermann, E. (2015). The complete genome sequence of the rumen methanogen *Methanosarcina barkeri* CM1. *Standards in Genomic Sciences*, 10, 57.
- Lau, S. K. P., McNabb, A., Woo, G. K. S., Hoang, L., Fung, A. M. Y., Chung, L. M. W., . . . Yuen, K.-Y. (2007). *Catabacter hongkongensis* gen. nov., sp. nov., isolated from blood cultures of patients from Hong Kong and Canada. *Journal of Clinical Microbiology*, 45(2), 395-401.
- Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., . . . Attwood, G. T. (2010). The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLOS One*, 5(1), e8926.
- Lehloenya, K. V., Krehbiel, C. R., Mertz, K. J., Rehberger, T. G., & Spicer, L. J. (2008). Effects of propionibacteria and yeast culture fed to steers on nutrient intake and site and extent of digestion. *Journal of Dairy Science*, 91(2), 653-662.
- Letunic, I., & Bork, P. (2016). Interactive Tree of Life (ITOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242-W245.
- Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., . . . Guan, L. L. (2019). Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome*, 7(1), 92.
- Li, F., Neves, A. L. A., Ghoshal, B., & Guan, L. L. (2017). Mining metagenomic and metatranscriptomic data for clues about microbial metabolic functions in ruminants. *Journal of Dairy Science*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Li, J., Gálvez, E. J. C., Amend, L., Almási, É, Iljazovic, A., Lesker, T. R., . . . Strowig, T. (2021). A versatile genetic toolbox for *Prevotella copri* enables studying polysaccharide utilization systems. *The EMBO Journal*, 40(23), e108287.
- Li, J., Ge, Y., Zadeh, M., Curtiss, R., & Mohamadzadeh, M. (2020). Regulating vitamin B<sub>12</sub> biosynthesis via the Cbl riboswitch in *Propionibacterium* strain UF1. *Proceedings of the National Academy of Sciences, USA*, 117(1), 602.
- Li, Z., Wright, A. G., Si, H., Wang, X., Qian, W., Zhang, Z., & Li, G. (2016). Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. *Environmental Microbiology Reports*, 8(6), 1016-1023.
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), 923-930.
- Linhardt, R. J., Galliher, P. M., & Cooney, C. L. (1986). Polysaccharide lyases. *Applied Biochemistry and Biotechnology*, 12(2), 135-176.
- Liu, X., Sutter, J. L., de la Cuesta-Zuluaga, J., Waters, J.L., Youngblut, N.D., Ley, R. E. (2021) Reclassification of *Catabacter hongkongensis* as *Christensenella hongkongensis* comb. nov. based on whole genome analysis. *International Journal of Systematic and Evolutionary Microbiology*, 71(4), 004774.
- Louis, P., & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, 19(1), 29-41.



- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.
- Lueders, T., Manefield, M., Friedrich, M. W. (2004). Enhanced sensitivity of DNA and rRNA-based isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients.
- Luo, H., & Gao, F. (2019). DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Research*, *47*(D1), D74-D77.
- Ma, C., Shi, Y., Mu, Q., Li, R., Xue, Y., & Yu, B. (2021). Unravelling the thioesterases responsible for propionate formation in engineered *Pseudomonas putida* KT2440. *Microbial Biotechnology*, *14*(3), 1237-1242.
- Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., & Mijakovic, I. (2019). Protein post-translational modifications in bacteria. *Nature Reviews Microbiology*, *17*(11), 651-664.
- Mackie, R. I., Sghir, A., Gaskins, H. R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *The American Journal of Clinical Nutrition*, *69*(5),
- Mahoney-Kurpe, S. C., Palevich, N., Noel, S. J., Kumar, S., Gagic, D., Biggs, P. J., . . . Moon, C. D. (2021). Complete genome sequences of three *Clostridiales* R-7 group strains isolated from the bovine rumen in New Zealand. *Microbiology Resource Announcements*, *10*(26), e00310-21.
- Marques, R. D. S., & Cooke, R. F. (2021). Effects of ionophores on ruminal function of beef cattle. *Animals*, *11*(10).
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10-12.
- Maxin, G., Rulquin, H., & Glasser, F. (2011). Response of milk fat concentration and yield to nutrient supply in dairy cows. *Animal*, *5*(8), 1299-1310.
- McCann, J. C., Wickersham, T. A., & Loor, J. J. (2014). High-throughput methods redefine the rumen microbiome and its relationship with nutrition and metabolism. *Bioinformatics and Biology Insights*, *8*, 109-125.
- McCarthy, A. A., Baker, H. M., Shewry, S. C., Patchett, M. L., Baker, E. N. (2001). Crystal structure of methylmalonyl-coenzyme A epimerase from *P. shermanii*: a novel enzymatic function on an ancient metal binding scaffold. *Structure*, *9*(7), 637-646.
- McFarland, L. V. (2021). Efficacy of single-strain probiotics versus multi-strain mixtures: systematic review of strain and disease specificity. *Digestive Diseases and Sciences*, *66*(3), 694-704.
- McMahon, L. R., McAllister, T. A., Berg, B. P., Majak, W., Acharya, S. N., Popp, J. D., . . . Cheng, K. J. (2000). A review of the effects of forage condensed tannins on ruminal fermentation and bloat in grazing cattle. *Canadian Journal of Plant Science*, *80*(3), 469-485.
- Mellin, J. R., Tiensuu, T., Bécavin, C., Guin, E., Johansson, J., & Cossart, P. (2013). A riboswitch-regulated antisense RNA in *Listeria monocytogenes*. *Proceedings of the National Academy of Sciences, USA*, *110*(32), 13132-13137.
- Miedes, E., Vanholme, R., Boerjan, W., & Molina, A. (2014). The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in Plant Science*, *5*, 358.
- Miettinen, H., & Huhtanen, P. (1996). Effects of the ratio of ruminal propionate to butyrate on milk yield and blood metabolites in dairy cows. *Journal of Dairy Science*, *79*(5), 851-861.
- Miller, T. L., & Jenesel, S. E. (1979). Enzymology of butyrate formation by *Butyrivibrio fibrisolvens*. *Journal of Bacteriology*, *138*(1), 99-104.
- Mills, J. D., Kawahara, Y., & Janitz, M. (2013). Strand-specific RNA-Seq provides greater resolution of transcriptome profiling. *Current Genomics*, *14*(3), 173-181.
- Minato, H., Otsuka, M., Shirasaka, S., Itabashi, H., & Mitsumori, M. (1992). Colonization of microorganisms in the rumen of young calves. *The Journal of General and Applied Microbiology*, *38*(5), 447-456.

- Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., . . . Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, *43*, D213-D221.
- Miyazaki, K., Martin, J. C., Marinsek-Logar, R., & Flint, H. J. (1997). Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp. *brevis*) B<sub>14</sub>. *Anaerobe*, *3*(6), 373-381.
- Moench, T. T., & Zeikus, J. G. (1983). An improved preparation method for a titanium (III) media reductant. *Journal of Microbiological Methods*, *1*(4), 199-202.
- Moore, K., & Jung, H. (2001). Lignin and fiber digestion. *Journal of Range Management*, *54*, 420-430.
- Moraïs, S., Mizrahi, I. (2019). Islands in the stream: from individual to communal fiber degradation in the rumen ecosystem. *FEMS Microbiology Reviews*, *43*, 362-379.
- Morgavi, D. P., Forano, E., Martin, C., & Newbold, C. J. (2010). Microbial ecosystem and methanogenesis in ruminants. *Animal*, *4*(7), 1024-1036.
- Morotomi, M., Nagai, F., & Watanabe, Y. (2012). Description of *Christensenella minuta* gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order *Clostridiales*, and proposal of *Christensenellaceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, *62*(1), 144-149.
- Mukherjee, S., & Sengupta, S. (2016). Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in genomic sequences. *Bioinformatics*, *32*(5), 776-778.
- Müller, V., Chowdhury, N. P., Basen, M. (2018). Electron bifurcation: a long-hidden energy coupling mechanism. *Annual Review of Microbiology*, *72*(1), 331-353.
- Nakamura, A. M., Nascimento, A. S., & Polikarpov, I. (2017). Structural diversity of carbohydrate esterases. *Biotechnology Research and Innovation*, *1*(1), 35-51.
- Naumoff, D. G. (2011). Hierarchical classification of glycoside hydrolases. *Biochemistry*, *76*(6), 622-635.
- Newbold, C. J., de la Fuente, G., Belanche, A., Ramos-Morales, E., & McEwan, N. R. (2015). The role of ciliate protozoa in the rumen. *Frontiers in Microbiology*, *6*, 1313-1313.
- Ng, F., Kittelmann, S., Patchett, M. L., Attwood, G. T., Janssen, P. H., Rakonjac, J., & Gagic, D. (2015). An adhesin from hydrogen-utilizing rumen methanogen *Methanobrevibacter ruminantium* M1 binds a broad range of hydrogen-producing microorganisms. *Environmental Microbiology*, *18*(9), 3010-3021.
- Noel, S. J. (2013). Cultivation and community composition analysis of plant-adherent rumen bacteria. PhD thesis, Massey University, Palmerston North, New Zealand.
- Noel, S. J., Attwood, G. T., Rakonjac, J., Moon, C. D., Waghorn, G. C., & Janssen, P. H. (2017). Seasonal changes in the digesta-adherent rumen bacterial communities of dairy cattle grazing pasture. *PLOS One*, *12*(3), e0173819.
- Nyonyo, T., Shinkai, T., Tajima, A., & Mitsumori, M. (2013). Effect of media composition, including gelling agents, on isolation of previously uncultured rumen bacteria. *Letters in Applied Microbiology*, *56*(1), 63-70.
- Oren, A., Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, *71*, 005056.
- Orpin, C. G. (1977). Invasion of plant tissue in the rumen by the flagellate *Neocallimastix frontalis*. *Journal of General Microbiology*, *98*(2), 423-430.
- Ørskov, E. R., & McDonald, I. (1979). The estimation of protein degradability in the rumen from incubation measurements weighted according to rate of passage. *The Journal of Agricultural Science*, *92*(2), 499-503.
- Palevich, N. (2016). Comparative genomics of *Butyrivibrio* and *Pseudobutyrvibrio* from the rumen. PhD thesis, Massey University, Palmerston North, New Zealand.
- Palevich, N., Kelly, W. J., Ganesh, S., Rakonjac, J., & Attwood, G. T. (2019a). *Butyrivibrio hungatei* MB2003 competes effectively for soluble sugars released by *Butyrivibrio proteoclasticus*

- B316 during growth on xylan or pectin. *Applied and Environmental Microbiology*, 85(3), e02056-02018.
- Palevich, N., Kelly, W. J., Leahy, S. C., Denman, S., Altermann, E., Rakonjac, J., & T., A. G. (2019b). Comparative genomics of rumen *Butyrivibrio* spp. uncovers a continuum of polysaccharide-degrading capabilities. *Applied and Environmental Microbiology*, 86(1), e01993-01919.
- Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R. E., & Blot, M. (1999). Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences*, 96(7), 3807-3812.
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarszewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, 36(10), 996-1004.
- Parry, M. L., Canziani, O., Palutikof, J. P., van der Linden, P., & Hanson, C. (2007). *Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.
- Paster, B. J., Russell, J. B., Yang, C. M., Chow, J. M., Woese, C. R., & Tanner, R. (1993). Phylogeny of the ammonia-producing ruminal bacteria *Peptostreptococcus anaerobius*, *Clostridium sticklandii*, and *Clostridium aminophilum* sp. nov. *International Journal of Systematic Bacteriology*, 43(1), 107-110.
- Paynter, M. J., & Elsdon, S. R. (1970). Mechanism of propionate formation by *Selenomonas ruminantium*, a rumen micro-organism. *Journal of General Microbiology*, 61(1), 1-7.
- Perez-Muñoz, M. E., Arrieta, M.-C., Ramer-Tait, A. E., & Walter, J. (2017). A critical assessment of the “sterile womb” and “in utero colonization” hypotheses: implications for research on the pioneer infant microbiome. *Microbiome*, 5(1), 48.
- Peselis, A., & Serganov, A. (2012). Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nature Structural & Molecular Biology*, 19(11), 1182-1184.
- Pisithkul, T., Patel, N. M., & Amador-Noguez, D. (2015). Post-translational modifications as key regulators of bacterial metabolic fluxes. *Current Opinion in Microbiology*, 24, 29-37.
- Pritchard, L., White, J. A., Birch, P. R. J., & Toth, I. K. (2006). GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics*, 22(5), 616-617.
- Procházka, J., Mrázek, J., Štrosová, L., Fliegerová, K., Záborská, J., & Dohányos, M. (2012). Enhanced biogas yield from energy crops with rumen anaerobic fungi. *Engineering in Life Sciences*, 12(3), 343-351.
- Puentes-Téllez, P. E., & Falcao Salles, J. (2018). Construction of effective minimal active microbial consortia for lignocellulose degradation. *Microbial Ecology*, 76(2), 419-429.
- Puniya, A. K., Salem, A. Z. M., Kumar, S., Dagar, S. S., Griffith, G. W., Puniya, M., . . . Kumar, R. (2015). Role of live microbial feed supplements with reference to anaerobic fungi in ruminant productivity: a review. *Journal of Integrative Agriculture*, 14(3), 550-560.
- Purushe, J., Fouts, D. E., Morrison, M., White, B. A., Mackie, R. I., North American Consortium for Rumen Bacteria, . . . Nelson, K. E. (2010). Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii*: insights into their biological niche. *Microbial Ecology*, 60(4), 721-729.
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833-844.
- Rae, A. L., Manners, J. M., Jones, R. J., McIntyre, C. L., & Lu, D. (2001). Antisense suppression of the lignin biosynthetic enzyme, caffeate O-methyltransferase, improves *in vitro* digestibility of the tropical pasture legume, *Stylosanthes humilis*. *Functional Plant Biology*, 28(4), 289-297.
- Ramšak, A., Peterka, M., Tajima, K., Martin, J. C., Wood, J., Johnston, M. E. A., . . . Avguštin, G. (2000). Unravelling the genetic diversity of ruminal bacteria belonging to the CFB phylum. *FEMS Microbiology Ecology*, 33(1), 69-79.

- Ranganathan, A., Smith, O. P., Youssef, N. H., Struchtemeyer, C. G., Atiyeh, H. K., & Elshahed, M. S. (2017). Utilizing anaerobic fungi for two-stage sugar extraction and biofuel production from lignocellulosic biomass. *Frontiers in Microbiology*, *8*, 635.
- Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P., . . . Louis, P. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal*, *8*(6), 1323-1335.
- Reisinger, A., Clark, H., Cowie, A. L., Emmet-Booth, J., Gonzalez Fischer, C., Herrero, M., . . . Leahy, S. C. (2021). How necessary and feasible are reductions of methane emissions from livestock to support stringent temperature goals? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2210), 20200452.
- Rey, M., Enjalbert, F., Combes, S., Cauquil, L., Bouchez, O., & Monteils, V. (2013). Establishment of ruminal bacterial community in dairy calves from birth to weaning is sequential. *Journal of Applied Microbiology*, *116*(2), 245-257.
- Richardson, A. J., Calder, A. G., Stewart, C. S., & Smith, A. (1989). Simultaneous determination of volatile and non-volatile acidic fermentation products of anaerobes by capillary gas chromatography. *Letters in Applied Microbiology*, *9*(1), 5-8.
- Richardson, L. F., Raun, A. P., Potter, E. L., Cooley, C. O., & Rathmacher, R. P. (1976). Effect of monensin on rumen fermentation *in vitro* and *in vivo*. *Journal of Animal Science*, *43*(3), 657-664.
- Richter-Dahlfors, A. A., & Andersson, D. I. (1992). Cobalamin (vitamin B<sub>12</sub>) repression of the Cob operon in *Salmonella typhimurium* requires sequences within the leader and the first translated open reading frame. *Molecular Microbiology*, *6*(6), 743-749.
- Ritari, J., Salojärvi, J., Lahti, L., & de Vos, W. M. (2015). Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC genomics*, *16*, 1056.
- Ross, E. M., Petrovski, S., Moate, P. J., & Hayes, B. J. (2013). Metagenomics of rumen bacteriophage from thirteen lactating dairy cattle. *BMC Microbiology*, *13*, 242.
- Russell, J. B., & Houlihan, A. J. (2003). Ionophore resistance of ruminal bacteria and its potential impact on human health. *FEMS Microbiology Reviews*, *27*(1), 65-74.
- Russell, J. B., & Mantovani, H. C. (2002). The bacteriocins of ruminal bacteria and their potential as an alternative to antibiotics. *Journal of Molecular Microbiology and Biotechnology*, *4*(4), 347-355.
- Russell, J. B., & Rychlik, J. L. (2001). Factors that alter rumen microbial ecology. *Science*, *292*(5519), 1119-1122.
- Russell, J. B., & Wallace, R. J. (1997). Energy-yielding and energy-consuming reactions. In P. N. Hobson & C. S. Stewart (Eds.), *The Rumen Microbial Ecosystem* (pp. 246-282). Dordrecht: Springer Netherlands.
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., . . . Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*, 87.
- Sangubotla, R., & Kim, J. (2021). Chapter 28 - Advances in mass spectrometry for microbial proteome analysis. In B. Viswanath (Ed.), *Recent Developments in Applied Microbiology and Biochemistry* (pp. 299-308): Academic Press.
- Sangwan, N., Xia, F., & Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, *4*(1), 8.
- Sasson, G., Kruger Ben-Shabat, S., Seroussi, E., Doron-Faigenboim, A., Shterzer, N., Yaacoby, S., . . . Mizrahi, I. (2017). Heritable bovine rumen bacteria are phylogenetically related and correlated with the cow's capacity to harvest energy from its feed. *mBio*, *8*(4).
- Sawanon, S., Koike, S., & Kobayashi, Y. (2011). Evidence for the possible involvement of *Selenomonas ruminantium* in rumen fiber digestion. *FEMS Microbiology Letters*, *325*(2), 170-179.
- Saxena, R. K., Anand, P., Saran, S., Isar, J., & Agarwal, L. (2010). Microbial production and applications of 1,2-propanediol. *Indian Journal of Microbiology*, *50*(1), 2-11.

- Schaefer, D. M., Davis, C. L., & Bryant, M. P. (1980). Ammonia saturation constants for predominant species of rumen bacteria. *Journal of Dairy Science*, *63*(8), 1248-1263.
- Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology*, *61*, 263-289.
- Seedorf, H., Kittelmann, S., Henderson, G., & 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.
- Seedorf, H., Kittelmann, S., & Janssen, P. H. (2015). Few highly abundant operational taxonomic units dominate within rumen methanogenic archaeal species in New Zealand sheep and cattle. *Applied and Environmental Microbiology*, *81*(3), 986-995.
- Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell*, *152*(1-2), 17-24.
- Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., . . . Kelly, W. J. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nature Biotechnology*, *36*, 359-367.
- Shabat, S. K., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Berg Miller, M. E., . . . Mizrahi, I. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *The ISME journal*, *10*(12), 2958-2972.
- Shade, A., & Handelsman, J. (2012). Beyond the Venn diagram: the hunt for a core microbiome. *Environmental Microbiology*, *14*(1), 4-12.
- Shakya, M., Lo, C. C., & Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics*, *10*, 904.
- Shi, W., Moon, C. D., Leahy, S. C., Kang, D., Froula, J., Kittelmann, S., . . . Rubin, E. M. (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Research*, *24*(9), 1517-1525.
- Siciliano-Jones, J., & Murphy, M. R. (1989). Production of volatile fatty acids in the rumen and cecum-colon of steers as affected by forage:concentrate and forage physical form. *Journal of Dairy Science*, *72*(2), 485-492.
- Skillman, L. C., Evans, P. N., Naylor, G. E., Morvan, B., Jarvis, G. N., & Joblin, K. N. (2004). 16S ribosomal DNA-directed PCR primers for ruminal methanogens and identification of methanogens colonising young lambs. *Anaerobe*, *10*(5), 277-285.
- Smaczniak, C., Li, N., Boeren, S., America, T., van Dongen, W., Goerdayal, S. S., . . . Kaufmann, K. (2012). Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues. *Nature Protocols*, *7*(12), 2144-2158.
- Solden, L. M., Hoyt, D. W., Collins, W. B., Plank, J. E., Daly, R. A., Hildebrand, E., . . . Wrighton, K. C. (2017). New roles in hemicellulosic sugar fermentation for the uncultivated *Bacteroidetes* family BS11. *The ISME Journal*, *11*(3), 691-703.
- Söllinger, A., Tveit, A. T., Poulsen, M., Noel, S. J., Bengtsson, M., Bernhardt, J., . . . Urich, T. (2018). Holistic assessment of rumen microbiome dynamics through quantitative metatranscriptomics reveals multifunctional redundancy during key steps of anaerobic feed degradation. *mSystems*, *3*(4), e00038-18.
- Solomon, R., Wein, T., Levy, B., Eshed, S., Dror, R., Reiss, V., . . . Jami, E. (2022). Protozoa populations are ecosystem engineers that shape prokaryotic community structure and function of the rumen microbial ecosystem. *The ISME Journal*
- Somerville, C. (2006). Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology*, *22*, 53-78.
- Søndergaard, D., Pedersen, C. N. S., & Greening, C. (2016). HydDB: A web tool for hydrogenase classification and analysis. *Scientific Reports*, *6*(1), 34212.
- Springmann, M., Mason-D'Croz, D., Robinson, S., Garnett, T., Godfray, H. C., Gollin, D., . . . Scarborough, P. (2016). Global and regional health effects of future food production under climate change: a modelling study. *Lancet*, *387*(10031), 1937-1946.
- Stams, A. J. M., Plugge, C. M. (2009) Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Reviews Microbiology*, *7*, 568-577.

- Stein, D. R., Allen, D. T., Perry, E. B., Bruner, J. C., Gates, K. W., Rehberger, T. G., . . . Spicer, L. J. (2006). Effects of feeding propionibacteria to dairy cows on milk yield, milk components, and reproduction. *Journal of Dairy Science*, *89*(1), 111-125.
- Statistics New Zealand (2019). 'Livestock numbers' and 'Agricultural and Horticultural land use'. Retrieved from <https://www.stats.govt.nz/>.
- Statistics New Zealand (2020). 'Overseas merchandise trade datasets'. Retrieved from <https://www.stats.govt.nz/>.
- Stevenson, D. M., & Weimer, P. J. (2007). Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. *Applied Microbiology and Biotechnology*, *75*(1), 165-174.
- Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., & Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature Biotechnology*, *37*(8), 953-961.
- Strobel, H. J. (1992). Vitamin B<sub>12</sub>-dependent propionate production by the ruminal bacterium *Prevotella ruminicola* 23. *Applied and Environmental Microbiology*, *58*(7), 2331-2333.
- Sullivan, M. B. (2015). Viromes, not gene markers, for studying double-stranded DNA virus communities. *Journal of Virology*, *89*(5), 2459.
- Swift, C. L., Louie, K. B., Bowen, B. P., Hooker, C. A., Solomon, K. V., Singan, V., . . . O'Malley, M. A. (2021). Cocultivation of anaerobic fungi with rumen bacteria establishes an antagonistic relationship. *mBio*, *12*(4), e01442-21.
- Tamura, K., Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*, 512-526.
- Tapio, I., Fischer, D., Blasco, L., Tapio, M., Wallace, R. J., Bayat, A. R., . . . Vilkki, J. (2017). Taxon abundance, diversity, co-occurrence and network analysis of the ruminal microbiota in response to dietary changes in dairy cows. *PLOS One*, *12*(7), e0180260.
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., . . . Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, *44*(14), 6614-6624.
- Taxis, T. M., Wolff, S., Gregg, S. J., Minton, N. O., Zhang, C., Dai, J., . . . Conant, G. C. (2015). The players may change but the game remains: network analyses of ruminal microbiomes suggest taxonomic differences mask functional similarity. *Nucleic Acids Research*, *43*(20), 9600-9612.
- The UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480-D489.
- Thomason, M. K., & Storz, G. (2010). Bacterial antisense RNAs: how many are there, and what are they doing? *Annual Review of Genetics*, *44*, 167-188.
- Tilman, D., & Clark, M. (2014). Global diets link environmental sustainability and human health. *Nature*, *515*, 518-522.
- Timmerman, H. M., Koning, C. J., Mulder, L., Rombouts, F. M., & Beynen, A. C. (2004). Monostrain, multistrain and multispecies probiotics- a comparison of functionality and efficacy. *International Journal of Food Microbiology*, *96*(3), 219-233.
- Tjardes, K. E., Buskirk, D. D., Allen, M. S., Ames, N. K., Bourquin, L. D., & Rust, S. R. (2000). Brown midrib-3 corn silage improves digestion but not performance of growing beef steers. *Journal of Animal Science*, *78*(11), 2957-2965.
- Trchsel, J., Bayles, D. O., Looft, T., Levine, U. Y., & Allen, H. K. (2016). Function and phylogeny of bacterial butyryl coenzyme A: acetate transferases and their diversity in the proximal colon of swine. *Applied and Environmental Microbiology*, *82*(22), 6788-6798.
- Tyanova, S., Temu, T., & Cox, J. (2016a). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, *11*(12), 2301-2319.

- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., . . . Cox, J. (2016b). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, *13*(9), 731-740.
- Urrutia, N. L., & Harvatine, K. J. (2017). Acetate dose-dependently stimulates milk fat synthesis in lactating dairy cows. *Journal of Nutrition*, *147*(5), 763-769.
- van Gylswyk, N. O. (1995). *Succiniclasticum ruminis* gen. nov., sp. nov., a ruminal bacterium converting succinate to propionate as the sole energy-yielding mechanism. *International Journal of Systematic Bacteriology*, *45*(2), 297-300.
- Van Maanen, R. W., Herbein, J. H., McGilliard, A. D., & Young, J. W. (1978). Effects of monensin on in vivo rumen propionate production and blood glucose kinetics in cattle. *The Journal of Nutrition*, *108*(6), 1002-1007.
- van Nevel, C. J., Prins, R. A., & Demeyer, D. I. (1974). On the inverse relationship between methane and propionate in the rumen. *Zeitschrift für Tierphysiologie, Tierernährung und Futtermittelkunde*, *33*(3), 121-125.
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., & Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiology*, *153*(3), 895-905.
- Varel, V. H., Yen, J. T., & Kreikemeier, K. K. (1995). Addition of cellulolytic clostridia to the bovine rumen and pig intestinal tract. *Applied and Environmental Microbiology*, *61*(3), 1116-1119.
- Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737-746.
- Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., & Gelfand, M. S. (2003). Regulation of the vitamin B<sub>12</sub> metabolism and transport in bacteria by a conserved RNA structural element. *RNA*, *9*(9), 1084-1097.
- Voragen, A. G. J., Coenen, G., Verhoef, R. P., & Schols, H. A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Structural Chemistry*, *20*(2), 263-275.
- Vyas, D., McGeough, E. J., McGinn, S. M., McAllister, T. A., & Beauchemin, K. A. (2014). Effect of *Propionibacterium* spp. on ruminal fermentation, nutrient digestibility, and methane emissions in beef heifers fed a high-forage diet. *Journal of Animal Science*, *92*(5), 2192-2201.
- Waghorn, G. C., & Clark, D. A. (2004). Feeding value of pastures for ruminants. *New Zealand Veterinary Journal*, *52*(6), 320-331.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., . . . Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS One*, *9*(11), e112963.
- Wallace, R. J., Onodera, R., & Cotta, M. A. (1997). Metabolism of nitrogen-containing compounds. In P. N. Hobson & C. S. Stewart (Eds.), *The Rumen Microbial Ecosystem* (pp. 283-328). Dordrecht: Springer Netherlands.
- Weimer, P. J. (1996). Why don't ruminal bacteria digest cellulose faster? *Journal of Dairy Science*, *79*(8), 1496-1502.
- Weimer, P. J. (1998). Manipulating ruminal fermentation: a microbial ecological perspective. *Journal of Animal Science*, *76*(12), 3114-3122.
- Weimer, P. J. (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. *Frontiers in Microbiology*, *6*, 296.
- Weimer, P. J., Cox, M. S., Vieira de Paula, T., Lin, M., Hall, M. B., & Suen, G. (2017). Transient changes in milk production efficiency and bacterial community composition resulting from near-total exchange of ruminal contents between high- and low-efficiency Holstein cows. *Journal of Dairy Science*, *100*(9), 7165-7182.
- Weimer, P. J., Stevenson, D. M., Mantovani, H. C., & Man, S. L. (2010). Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents. *Journal of Dairy Science*, *93*(12), 5902-5912.

- Weimer, P. J., Waghorn, G. C., Odt, C. L., & Mertens, D. R. (1999). Effect of diet on populations of three species of ruminal cellulolytic bacteria in lactating dairy cows. *Journal of Dairy Science*, 82(1), 122-134.
- Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, 173(2), 697-703.
- Welch, D. F. (1991). Applications of cellular fatty acid analysis. *Clinical Microbiology Reviews*, 4(4), 422-438.
- Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 21(2/3), 213-251.
- Widdel, F., & Pfennig, N. (1981). Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. *Archives of Microbiology*, 129(5), 395-400.
- Williams, A. G., & Coleman, G. S. (1997). The rumen protozoa. In P. N. Hobson & C. S. Stewart (Eds.), *The Rumen Microbial Ecosystem* (pp. 73-139). Dordrecht: Springer Netherlands.
- Wilson, C. A., & Wood, T. M. (1992). The anaerobic fungus *Neocallimastix frontalis*: isolation and properties of a cellulosome-type enzyme fraction with the capacity to solubilize hydrogen-bond-ordered cellulose. *Applied Microbiology and Biotechnology*, 37(1), 125-129.
- Wolin, M. J., Miller, T. L., & Stewart, C. S. (1997). Microbe-microbe interactions. In P. N. Hobson & C. S. Stewart (Eds.), *The Rumen Microbial Ecosystem* (pp. 467-491). Dordrecht: Springer Netherlands.
- Xie, F., Jin, W., Si, H., Yuan, Y., Tao, Y., Liu, J., . . . Mao, S. (2021). An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants. *Microbiome*, 9(1), 137.
- Yáñez-Ruiz, D. R., Abecia, L., & Newbold, C. J. (2015). Manipulating rumen microbiome and fermentation through interventions during early life: a review. *Frontiers in Microbiology*, 6, 1133.
- Yang, C. M., & Russell, J. B. (1992). Resistance of proline-containing peptides to ruminal degradation *in vitro*. *Applied and Environmental Microbiology*, 58(12), 3954-3958.
- Yarlett, N., Hann, A. C., Lloyd, D., & Williams, A. G. (1983). Hydrogenosomes in a mixed isolate of *Isotricha prostoma* and *Isotricha intestinalis* from ovine rumen contents. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 74(2), 357-364.
- Yarlett, N., Lloyd, D., & Williams, A. G. (1985). Butyrate formation from glucose by the rumen protozoon *Dasytricha ruminantium*. *Biochemical Journal*, 228(1), 187-192.
- Zhang, B., Lingga, C., Bowman, C., & Hackmann, T. J. (2021). A new pathway for forming acetate and synthesizing ATP during fermentation in bacteria. *Applied and Environmental Microbiology*, 87(14), e0295920.
- Zhang, J., Shi, H., Wang, Y., Li, S., Cao, Z., Ji, S., . . . Zhang, H. (2017). Effect of dietary forage to concentrate ratios on dynamic profile changes and interactions of ruminal microbiota and metabolites in Holstein heifers. *Frontiers in Microbiology*, 8, 2206.
- Zhao, C., Dong, H., Zhang, Y., & Li, Y. (2019). Discovery of potential genes contributing to the biosynthesis of short-chain fatty acids and lactate in gut microbiota from systematic investigation in *E. coli*. *npj Biofilms and Microbiomes*, 5(1), 19.
- Zhong, R., & Ye, Z. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant and Cell Physiology*, 56(2), 195-214.
- Zhou, M., Hünerberg, M., Beauchemin, K. A., McAllister, T. A., Okine, E. K., & Guan, L. L. (2012). Individuality of ruminal methanogen/protozoa populations in beef cattle fed diets containing dried distillers' grain with solubles. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 62(4), 273-288.
- Zou, Y., Xue, W., Lin, X., Hu, T., Liu, S.-W., Sun, C.-H., . . . Xiao, L. (2021). Taxonomic description and genome sequence of *Christensenella intestinhominis* sp. nov., a novel cholesterol-lowering bacterium isolated from human gut. *Frontiers in Microbiology*, 12(182).



## 7 Appendix

## 7.1 Complete genome sequences of three *Clostridiales* R-7 group strains isolated from the bovine rumen in New Zealand



GENOME SEQUENCES



### Complete Genome Sequences of Three *Clostridiales* R-7 Group Strains Isolated from the Bovine Rumen in New Zealand

Sam C. Mahoney-Kurpe,<sup>a,b</sup> Nikola Palevich,<sup>a</sup> Samantha J. Noel,<sup>a,b\*</sup> Sandeep Kumar,<sup>a</sup> Dragana Gagic,<sup>b,c</sup> Patrick J. Biggs,<sup>b,c</sup> Peter H. Janssen,<sup>a</sup> Graeme T. Attwood,<sup>a</sup> Christina D. Moon<sup>a</sup>

<sup>a</sup>AgResearch Ltd, Grasslands Research Centre, Palmerston North, New Zealand

<sup>b</sup>School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

<sup>c</sup>School of Veterinary Science, Massey University, Palmerston North, New Zealand

**ABSTRACT** Members of the *Clostridiales* R-7 group are abundant bacterial residents of the rumen microbiome; however, they are poorly characterized. We report the complete genome sequences of three members of the R-7 group, FE2010, FE2011, and XBB3002, isolated from the ruminal contents of pasture-grazed dairy cows in New Zealand.

The *Clostridiales* R-7 group is an abundant, but poorly characterized, group of unclassified rumen bacteria (1). Draft genome sequences of only two strains are currently available (2). The genomic characterization of additional members will accelerate efforts to understand the roles of the R-7 group in the rumen microbial ecosystem.

Rumen contents of fistulated dairy cows grazing ryegrass/clover pastures in Waikato, New Zealand (3), were obtained with AgResearch Grasslands Animal Ethics Committee approval (AE12174). An anaerobic dilution-to-extinction approach (4) was used to isolate FE2010 and FE2011 in RM02 medium supplemented with glucose, cellobiose, xylose, L-arabinose, lactate, Casamino Acids, Bacto peptone, and yeast extract (5), and XBB3002 was isolated in basal medium with yeast extract (BY) (6) at 39°C. FE2010 and FE2011 cells were Gram-negative rods, whereas XBB3002 cells were coccobacilli. Partial 16S rRNA gene sequences exhibited >96% nucleotide identity to rumen strain R-7 (7).

High-molecular-weight genomic DNA was extracted using a chemical/enzymatic lysis and phenol-chloroform extraction method (8) from 1- to 2-day-old cultures grown anaerobically in BY at 39°C. DNA was sequenced and assembled by Nextomics Biosciences (Wuhan, China). Long-read libraries were prepared using the native barcoding expansion (NBD-104) and SQK-LSK109 ligation sequencing kits and sequenced on a PromethION instrument, using Guppy v4.0.11 (Oxford Nanopore Technologies [ONT]) for base-calling and quality-filtering (Q, >7; sequence length, >1,000 bp). Short-read (2 × 150-bp) libraries were prepared using the MGISEQ-2000RS kit and sequenced using an MGISEQ-2000 instrument. Short reads were quality-filtered with fastp v0.20.0 (9); following removal of adapters, reads containing N base calls were removed. Reads had 5 bp trimmed from each end, and read pairs for which at least one read had >20% of bases with Q of <20 were removed. Quality-filtered ONT reads were assembled using Flye v2.7 (–plasmid and –nano-raw settings) (10). Assemblies were polished with Racon v1.4.13 (default settings) (11) using alignments of ONT data and with Pilon v1.23 (default settings) (12) and NextPolish v1.2.4 (default settings) (13) using alignments of the short-read data. Contigs were confirmed as circular using Circlator v1.5.1 (with the “fixstart” parameter) (14). Annotation was performed using the NCBI PGAP pipeline v5.0 (15). Sequences from each isolate assembled into circular contigs of similar size and G+C content (Table 1).

Citation Mahoney-Kurpe SC, Palevich N, Noel SJ, Kumar S, Gagic D, Biggs PJ, Janssen PH, Attwood GT, Moon CD. 2021. Complete genome sequences of three *Clostridiales* R-7 group strains isolated from the bovine rumen in New Zealand. *Microbiol Resour Announc* 10:e00310-21. <https://doi.org/10.1128/MRA.00310-21>.

Editor Catherine Putonti, Loyola University Chicago

Copyright © 2021 Mahoney-Kurpe et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Christina D. Moon, [christina.moon@agresearch.co.nz](mailto:christina.moon@agresearch.co.nz).

\* Present address: Samantha J. Noel, Department of Animal Science, Aarhus University, Aarhus, Denmark.

Received 24 March 2021

Accepted 1 June 2021

Published 1 July 2021

TABLE 1 Genome details of R-7 group strains

Parameter	Data for strain		
	FE2010	FE2011	XBB3002
BioProject accession no.	PRJNA695064	PRJNA695064	PRJNA695064
BioSample accession no.	SAMN17600269	SAMN17611198	SAMN17611949
GenBank accession no.	CP069593	CP069418	CP069419
SRA accession no. (ONT)	SRX10247579	SRX10248369	SRX10248392
SRA accession no. (MGISEQ)	SRX10247580	SRX10248370	SRX10248393
No. of raw ONT reads	815,513	966,753	443,226
No. of filtered ONT reads	781,555	926,448	419,863
$N_{50}$ of filtered ONT reads (bp)	4,724	3,808	6,131
No. of raw MGISEQ reads	6,874,216	6,886,570	6,896,410
No. of filtered MGISEQ reads	6,866,512	6,878,218	6,888,148
Genome size (Mb)	3.51	3.56	3.26
No. of contigs	1	1	1
Sequencing coverage (x)	650	628	457
G+C content (%)	53.2	53.2	56.5

Taxonomic assignments were determined using GTDB-Tk v1.3.0 (16) to query the Genome Taxonomy Database (GTDB) framework (17) (release 05-RS95). The strains were classified as members of the recently proposed order "Christensenellales" (17) in an undescribed family (CAG-74) and genus (GCA-900199385), which the previously sequenced R-7 group strains, R-7 (species-level taxon sp900199385) and WTE2008 (sp900176495) (2), have also been classified as. FE2010 and FE2011 were assigned to the species-level taxon sp900322155, while XBB3002 was unassigned at the species level. These genomes expand the number of sequenced representatives of the R-7 group and will progress our understanding of their biology.

**Data availability.** The complete genomes and raw sequence reads are available in GenBank and the Sequence Read Archive under the accession numbers in Table 1.

#### ACKNOWLEDGMENT

This work was funded by the New Zealand Ministry of Business, Innovation and Employment Strategic Science Investment Fund to AgResearch. The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

#### REFERENCES

- Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH, Global Rumen Census Collaborators. 2015. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep* 5:14567. <https://doi.org/10.1038/srep14567>.
- Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Elze-Fadros EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D, Hungate1000 project collaborators, Perry R, Henderson G, Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyriakides NC, Henrissat B, Woyske T, Ivanova NN, Kelly WJ. 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol* 36:359–367. <https://doi.org/10.1038/nbt.4110>.
- Noel SJ, Attwood GT, Rakonjac J, Moon CD, Waghorn GC, Janssen PH. 2017. Seasonal changes in the digesta-adherent rumen bacterial communities of dairy cattle grazing pasture. *PLoS One* 12:e0173819. <https://doi.org/10.1371/journal.pone.0173819>.
- Button DK, Schut F, Quang P, Martin R, Robertson BR. 1993. Viability and isolation of marine bacteria by dilution culture: theory, procedures, and initial results. *Appl Environ Microbiol* 59:881–891. <https://doi.org/10.1128/aem.59.3.881-891.1993>.
- Kenters N, Henderson G, Jeyanathan J, Kittelmann S, Janssen PH. 2011. Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *J Microbiol Methods* 84:52–60. <https://doi.org/10.1016/j.mimet.2010.10.011>.
- Joblin KN. 2005. Methanogenic archaea, p 47–53. In Makkar HPS, McSweeney CS (ed), *Methods in gut microbial ecology for ruminants*. Springer, Dordrecht, The Netherlands.
- Noel SJ. 2013. Cultivation and community composition analysis of plant-adherent rumen bacteria. PhD thesis. Massey University, Palmerston North, New Zealand.
- Palevich N. 2016. Comparative genomics of *Butyrvibrio* and *Pseudobutyrvibrio* from the rumen. PhD thesis. Massey University, Palmerston North, New Zealand.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34:884–890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.

13. Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36:2253–2255. <https://doi.org/10.1093/bioinformatics/btz891>.
14. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
15. Tatusova T, DiCuccio M, Badretdin A, Chetverin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
16. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
17. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>.

## 7.2 Genome statistics of resequenced R-7<sup>T</sup> and WTE2008<sup>T</sup> genomes

**Table 7.1. Genome characteristics of R-7 group strains.**

Parameter	1	2	3	4	5	6	7
GenBank accession	CP068393	CP069421-22	CP069593	CP069418	CP069419	CP029256	GCA_004342745
No. of contigs	1	2	1	1	1	1	38
Genome size (Mb)	3.39	3.41 (0.044) <sup>1</sup>	3.51	3.56	3.27	2.97	3.15
G+C content (%)	53	53.7 (39.9) <sup>1</sup>	53.2	53.2	56.5	51.5	48.5

Taxa: 1, R-7<sup>T</sup>; 2, WTE2008<sup>T</sup>; 3, FE2010; 4, FE2011; 5, XBB3002; 6, *Christensenella minuta* YIT12065<sup>T</sup>; 7, *Christensenella hongkongensis* HKU16<sup>T</sup>.

<sup>1</sup>Values in brackets represent statistics of the second contig of the WTE2008 assembly.

## 7.3 Cellular fatty acid profiles

**Table 7.2. Complete cellular fatty acid profiles of R-7<sup>T</sup> and WTE2008<sup>T</sup>.**

Cellular fatty-acid	R-7 <sup>T</sup>	WTE2008 <sup>T</sup>
10:0	0.11	0.11
11:0	-	0.05
11:0 anteiso	0.05	-
12:0	0.78	0.82
12:0 iso	0.18	0.33
13:0	0.20	0.38
13:0 iso	0.26	0.48
13:0 anteiso	0.42	0.94
13:0 iso 3OH	0.13	0.19
14:0	3.08	3.61
14:0 iso	1.80	3.19
14:0 DMA	-	0.19
15:0	3.23	3.44
15:0 iso	0.90	1.79
15:0 anteiso	8.92	7.70
15:0 iso DMA	0.43	0.67
15:0 3OH	0.17	0.67
16:0	11.54	15.17
16:0 iso	18.19	8.40
16:0 2OH	0.17	0.20
16:0 3OH	0.97	2.20
16:1 $\omega$ 7c	0.08	0.20
17:0	4.30	2.99
17:0 iso	0.48	0.70
17:0 anteiso	19.74	8.72
17:0 3OH	0.44	0.95
17:0 anteiso 3OH	0.78	2.39
18:0	9.11	19.32
18:0 iso	6.63	2.57
18:1 $\omega$ 6c	0.22	0.48
18:1 $\omega$ 9c	0.72	1.72
18:2 $\omega$ 6, 9c	0.75	1.27
19:0	0.50	0.35
19:0 anteiso	1.85	0.78
20:0	0.80	1.52
Summed feature 1	-	0.10
Summed feature 2	0.26	0.62
Summed feature 3	0.61	1.09
Summed feature 4	0.77	2.25
Summed feature 5	0.20	0.66
Summed feature 6	0.23	0.43
Summed feature 7	-	0.35

Summed features: 1, C12:0 3OH and C13:0 dimethylacetal (DMA); 2, C15:0 DMA and C14:0 3OH; 3, C13:0 anteiso 3OH and C16:1 $\omega$ 9c DMA; 4, C16:0 iso 3OH and unknown 17.157 DMA; 5, C18:1 $\omega$ 7c and unknown 17.834; 6, C17:0 iso 3OH and C18:2 DMA; 7, unknown 18.622 and C:19:0 iso. Categories showing dashes were not detected.

## 7.4 Ethanol and SCFA production on insoluble substrates

**Table 7.3. Production of ethanol and total SCFAs by R-7<sup>T</sup> and WTE2008<sup>T</sup> on insoluble substrates.**

Substrate	WTE2008 <sup>T</sup>		R-7 <sup>T</sup>	
	ethanol	SCFAs	ethanol	SCFAs
Cellobiose*	5.2 ± 0.6	1.2 ± 0.3	4.8 ± 0.9	3.8 ± 7.5
Cellulose	0.5 ± 0.1	-1.1 ± 0.2	0	-0.3 ± 1
Dextrin	2.7 ± 0.9	1.3 ± 0.8	6.6 ± 0.7	1.7 ± 4.2
Glycogen	0.5 ± 0.2	-0.1 ± 1.3	0	-3.5 ± 4.4
Inulin	0.5 ± 0.02	-0.5 ± 0.6	0.8 ± 0.1	0.7 ± 0.6
Pectin	2.3 ± 0.5	0.3 ± 1.1	1.2 ± 0.3	1.9 ± 0.3
Starch	2.1 ± 0.5	1.5 ± 0.9	3.3 ± 0.3	7.7 ± 0.8
Xylan	3.1 ± 0.1	1.3 ± 0.4	4.2 ± 0.6	7.5 ± 1.2
Rutin	0.6 ± 0.1	-0.1 ± 0.4	0	0.1 ± 0.1
Salicin	2.1 ± 0.3	0.3 ± 1.8	8.8 ± 0.8	7.3 ± 8.3

\*Optimal soluble substrate included as a reference. All numbers represent production of each end product (in mM), with the average concentrations of ethanol and total short-chain fatty-acids (SCFAs) seen in triplicate tubes with triplicate no-substrate controls and time 0 values subtracted, ± 1 standard deviation of the mean of each triplicate.

## 7.5 Genome details of the resequenced *P. ruminicola* KHP1 assembly

**Table 7.4. Genome statistics of *P. ruminicola* KHP1.**

Parameter	Data for strain KHP1 <sup>1</sup>
BioProject accession no.	<a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA715253">PRJNA715253</a>
BioSample accession no.	<a href="https://www.ncbi.nlm.nih.gov/biosample/SAMN18341379">SAMN18341379</a>
GenBank accession no.	<a href="https://www.ncbi.nlm.nih.gov/genbank/CP071890">CP071890</a>
SRA accession no. (ONT)	<a href="https://www.ncbi.nlm.nih.gov/sra/SRR16277835">SRR16277835</a>
SRA accession no. (MGISEQ)	<a href="https://www.ncbi.nlm.nih.gov/sra/SRR16277834">SRR16277834</a>
No. of raw ONT reads	617,812
No. of filtered ONT reads	577,710
<i>N</i> <sub>50</sub> of filtered ONT reads (bp)	5,042
No. of raw MGISEQ reads	6,862,268
No. of filtered MGISEQ reads	6,854,260
Genome size (Mb)	3.43
No. of contigs	1
Sequencing coverage (x)	514
G+C content	47.76
<b>Genes (total)</b>	2,861
<b>CDSs (total)</b>	2,793
<b>CDSs (with protein)</b>	2,771
<b>Genes (RNA)</b>	68
<b>rRNAs (5S, 16S, 23S)</b>	4, 4, 4
<b>tRNAs</b>	54
<b>Noncoding RNAs</b>	2
<b>Pseudogenes</b>	22

<sup>1</sup> Data are based on PGAP annotations.



## 7.6 Sequences and locations of cobalamin riboswitches

**Table 7.5. Predicted 'cobalamin' family riboswitches in the KHP1 genome.**

Sequence (5'-3')	Genome coordinates
CAAAATAAGGTCATCTGGTGGCCGATGCCA CTACGATGAAAAGGGAATACGGTGAGAATC CGTAACTGTACCTGCAGCTGTAATCCTCGCA AAAAGGGTTTGCCTGTATAACGCCACTGAG CCACAGGCTCGGGAAGGTAAGGCAGACTG AGGAAAGTCAGAAGACCTGCCGAATGTCAA TTGAAG	291423..291608 (+)
GTAATATCACCCGCGAAGGTGCACGTAAC GTGCTTAATTGGGAATGTGAGTGAGAATCT CCGACTGTCCCGCAGCAGTGAACCTCATTAT GGCTGTCCGACTATAGGCCATTGCCCGTTT TGGGCGAGAAGGCGTCCGATAGTGGAGGA AAGTCTGAAGACCAGCCTTCTGCGATTTTGT TGC	2792551..2792735 (+)
TTTAGCAAATACCGAAAAAGGCTGGTCTTCA GACTTTCCTCCACTCCCAAGCGCCTTCTCACC CAAATGGGCAATGGCATTGATGCTTAGGAG CCATAATGGAGTTCAGTCTGCGGGACAGT CGGAGATTCTCACTCACATTCCCAATTAATC GCGGCTAAGCGAACCTTTACGGGCGAAGGT C	1604149-1604333 (-)
AAGCGGCTAAAGTTTCATCGGCAGGTCTTCT GACTTATCGTCTGGCAGCAAACGTCTTCCCA AAAATCTCAGTGACATACATGTTTGCGCCTC ATAAAGGACGACTTACAGCAGCAGGAAAG GCCACGGGTAGGTGCCTTTCGGCGGGATTTC TGTTCAGGATTCTACCCGATTCCCTATTAAT CTCCACATCAGGAGAACCGATTACAGCTGC AAAGA	2049143..2049361 (-)

## 7.7 Protein homologues, encoded by the KHP1 genome, of characterised proteins/enzymes involved in converting succinate to propionate in other bacteria

**Table 7.6. Homologues in the KHP1 genome of characterised genes converting succinate to propionate in other bacteria.**

Enzymatic step	UniProt ID	Annotation	Host organism	Matches ( $\geq 30\%$ identity across $\geq 70\%$ query cover) to KHP1 <sup>1</sup>
succinate $\rightarrow$ succinyl-CoA	P52043	propionyl-CoA:succinyl-CoA transferase	<i>Escherichia coli</i>	succinate CoA-transferase (J4031_07180)- 49%
succinyl-CoA $\rightarrow$ methylmalonyl-CoA	P11652	methylmalonyl-CoA mutase small subunit	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i>	no matches
	P11653	methylmalonyl-CoA mutase large subunit	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i>	methylmalonyl-CoA mutase (J4031_12565)- 65%
	Q59676	methylmalonyl-CoA mutase small subunit	<i>Poryphromonas gingivalis</i>	methylmalonyl-CoA mutase small subunit (J4031_12560)- 56%
	Q59677	methylmalonyl-CoA mutase large subunit	<i>Poryphromonas gingivalis</i>	methylmalonyl-CoA mutase (J4031_12565)- 81%
R-methylmalonyl-CoA $\rightarrow$ S-methylmalonyl-CoA	Q8VQN0	methylmalonyl-CoA epimerase	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i>	methylmalonyl-CoA epimerase (J4031_00330)- 33%
	O58010	methylmalonyl-CoA epimerase	<i>Pyrococcus horokoshii</i>	methylmalonyl-CoA epimerase (J4031_00330)- 47%
	Q70AC7	Methylmalonyl-CoA carboxyltransferase 5S subunit	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i>	oxaloacetate decarboxylase (J4031_05015)- 30%
	Q8GBW6	Methylmalonyl-CoA carboxyltransferase 12S subunit	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i>	acyl-CoA carboxylase subunit beta (J4031_00325)- 52%
	P02904	Methylmalonyl-CoA carboxyltransferase 1.3S subunit	<i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i>	biotin/lipoyl binding domain containing protein (J4031_12545)- 35%
	Q57079	Methylmalonyl-CoA carboxyltransferase subunit alpha	<i>Veillonella parvula</i>	acyl-CoA carboxylase subunit beta (J4031_00325)- 58%
	Q57286	Methylmalonyl-CoA carboxyltransferase subunit beta	<i>Veillonella parvula</i>	oxaloacetate decarboxylase beta subunit (J4031_12540)- 54%
	Q57111	Methylmalonyl-CoA carboxyltransferase subunit gamma	<i>Veillonella parvula</i>	biotin/lipoyl binding domain containing protein (J4031_12545)- 33%
	Q56724	Methylmalonyl-CoA carboxyltransferase subunit delta	<i>Veillonella parvula</i>	no matches
	Q57490	Methylmalonyl-CoA carboxyltransferase subunit epsilon	<i>Veillonella parvula</i>	no matches

S-methylmalonyl-CoA → propionyl-CoA	O54028	Methylmalonyl-CoA decarboxylase subunit alpha	<i>Propionigenium modestum</i>	acyl-CoA carboxylase subunit beta (J4031_00325)- 54%
	O54031	Methylmalonyl-CoA decarboxylase subunit beta	<i>Propionigenium modestum</i>	oxaloacetate decarboxylase beta subunit (J4031_12540)- 53%
	O54030	Methylmalonyl-CoA decarboxylase subunit gamma	<i>Propionigenium modestum</i>	biotin/lipoyl binding domain containing protein (J4031_00315)- 37% biotin/lipoyl binding domain containing protein (J4031_12545)- 33%
	O54029	methylmalonyl-CoA decarboxylase subunit delta	<i>Propionigenium modestum</i>	no matches
propionyl-CoA → propionate	P52043	propionyl-CoA:succinyl-CoA transferase	<i>Escherichia coli</i>	Succinate CoA-transferase (J4031_07180)- 49%
	Q8Y7T9	phosphate propanoyltransferase	<i>Listeria monocytogenes</i>	no matches
	A0Q2W0	phosphate propanoyltransferase	<i>Clostridium novyi</i>	no matches
	Q187N2	phosphate propanoyltransferase	<i>Clostridioides difficile</i>	no matches
	Q834M4	phosphate propanoyltransferase	<i>Enterococcus faecalis</i>	no matches
	Q9XDN5	phosphate propanoyltransferase	<i>Salmonella typhimurium</i>	no matches
	P11868	propionate kinase	<i>Escherichia coli</i>	acetate kinase (J4031_09710)- 44%
	P59244	propionate kinase	<i>Escherichia coli</i> 06:H1	acetate kinase (J4031_09710)- 44%
	Q8Z3K5	propionate kinase	<i>Salmonella typhi</i>	acetate kinase (J4031_09710)- 44%
	B5XVZ8	propionate kinase	<i>Klebsiella pneumoniae</i>	acetate kinase (J4031_09710)- 42%
	O06961	propionate kinase	<i>Salmonella typhimurium</i>	acetate kinase (J4031_09710)- 44%
	Q05624	phosphate butyryltransferase	<i>Clostridium beijerinckii</i>	phosphate butyryltransferase (J4031_09245)- 35%
	P58255	phosphate butyryltransferase	<i>Clostridium acetobutylicum</i>	phosphate butyryltransferase (J4031_09245)- 36%
	Q49829	butyrate kinase	<i>Clostridium acetobutylicum</i>	butyrate kinase (J4031_09250)- 50%
	Q97II1	butyrate kinase 2	<i>Clostridium acetobutylicum</i>	butyrate kinase (J4031_09250)- 51%
	Q0TMV4	butyrate kinase	<i>Clostridium perfringens</i>	butyrate kinase (J4031_09250)- 50%
	Q8A4P5	probable butyrate kinase	<i>Bacteroides thetaiotaomicron</i>	butyrate kinase (J4031_09250)- 64%
	Q818T1	probable butyrate kinase	<i>Bacillus cereus</i>	butyrate kinase (J4031_09250)- 51%
	Q8Y7B6	probable butyrate kinase	<i>Listeria monocytogenes</i>	butyrate kinase (J4031_09250)- 48%

	Q9I5A5	phosphate acetyltransferase (pta)	<i>Pseudomonas aeruginosa</i>	phosphate acetyltransferase (J4031_09705)- 45%
	Q5LI41	acetate kinase ackA	<i>Bacteroides fragilis</i>	acetate kinase (J4031_09710)- 72%
	P0A6A3	acetate kinase	<i>Escherichia coli</i>	acetate kinase (J4031_09710)- 44%
	P63411	acetate kinase ackA	<i>Salmonella typhimurium</i>	acetate kinase (J4031_09710)- 44%
	P37877	acetate kinase	<i>Bacillus subtilis</i>	acetate kinase (J4031_09710)- 55%
	Q8XJN2	acetate kinase 2	<i>Clostridium perfringens</i>	acetate kinase (J4031_09710)- 54%
	P71104	acetate kinase	<i>Clostridium acetobutylicum</i>	acetate kinase (J4031_09710)- 54%

## 7.8 Genome location of candidate KHP1 succinate pathway genes between succinate and propionate

Table 7.7. Positioning of candidate succinate pathway genes between succinate and propionate in the KHP1 genome.

Reaction	PGAP annotation (Gene ID)	Genome coordinates	Top homology to <sup>1</sup>
<b>Succinate → succinyl-CoA</b>	succinate CoA-transferase (J4031_07180)	1649882..1651387 (-)	propionyl-CoA:succinate CoA transferase ( <i>E. coli</i> K12)- 49%
<b>Succinyl-CoA → methylmalonyl-CoA</b>	methylmalonyl-CoA mutase small subunit (J4031_12560)	3000484..3002346 (+)	methylmalonyl-CoA mutase small subunit ( <i>Poryphromonas gingivalis</i> )- 56%
	methylmalonyl-CoA mutase (J4031_12565)	3002366..3004531 (+)	methylmalonyl-CoA mutase large subunit ( <i>Poryphromonas gingivalis</i> )-81%
<b>R-methylmalonyl-CoA → S-methylmalonyl-CoA</b>	methylmalonyl-CoA epimerase (J4031_00330)	76532..76942 (-)	methylmalonyl-CoA epimerase ( <i>Pyrococcus horikoshii</i> )- 49%
<b>Methylmalonyl-CoA → propionyl-CoA</b>	oxaloacetate decarboxylase (J4031_05015)	1142714..1144495 (-)	methylmalonyl-CoA carboxyltransferase 5S subunit ( <i>Propionibacterium freudenreichii</i> subsp. <i>shermanii</i> )- 30%
	acyl-CoA carboxylase subunit beta (J4031_00325)	74924..76489 (-)	methylmalonyl-CoA decarboxylase subunit alpha ( <i>Veillonella parvula</i> ) 58%
	biotin/lipoyl binding protein (J4031_00315)	74303..74731 (-)	Glutaconyl-CoA decarboxylase subunit gamma ( <i>Acidaminococcus fermentans</i> )- 41%
	sodium ion-translocating decarboxylase subunit beta (J4031_12540)	2995810..2996967 (-)	oxaloacetate decarboxylase beta chain ( <i>Propionigenium modestum</i> )-53%
	biotin/lipoyl-binding protein (J4031_12545)	2997039..2997491 (-)	methylmalonyl-CoA decarboxylase subunit gamma ( <i>Veillonella parvula</i> )- 33%
	OadG (oxaloacetate decarboxylase gamma chain) family protein (J4031_12550)	2997497..2998522 (-)	n/a: selected as candidate based on gene annotation name and co-location with other selected candidates
<b>Propionyl-CoA → propionate</b>	(succinate CoA transferase) J4031_07180	1649882..1651387 (-)	propionyl-CoA:succinate CoA transferase ( <i>E. coli</i> K12)- 49%
	phosphate acetyltransferase (J4031_09705)	2340485..2341495 (+)	phosphate acetyltransferase ( <i>Clostridium acetobutylicum</i> )-53%
	acetate kinase (J4031_09710)	2341560..2342753 (+)	acetate kinase ( <i>Bacteroides fragilis</i> )- 73%
	phosphate butyryltransferase (J4031_09245)	2220507..2221343 (+)	phosphate butyryltransferase ( <i>Clostridium acetobutylicum</i> )- 35%
	butyrate kinase (J4031_09250)	2221343..2222413 (+)	butyrate kinase ( <i>Clostridium acetobutylicum</i> )- 50%

<sup>1</sup>Homology shown is to the best hit of each candidate to reviewed bacterial proteins in the UniProt (swiss-prot) database.

# 7.9 Homologues of candidate KHP1 propionate pathway genes in Hungate1000 assemblies of screened strains

**Table 7.8. Searching the Hungate1000 genomes of screened strains for homologues of candidate propionate pathway genes identified in the KHP1 genome.**

	succinate CoA-transferase (J4031_07180)	methylmalonyl-CoA mutase small subunit (J4031_12560)	methylmalonyl-CoA mutase large subunit (J4031_12565)	Oxaloacetate decarboxylase gamma family protein (J4031_12550)	biotin/lipoyl binding domain (J4031_12545)	sodium ion-translocating decarboxylase subunit beta (J4031_12540)	methylmalonyl-CoA epimerase (J4031_00330)	acyl-CoA carboxylase subunit beta (J4031_00325)	biotin/lipoyl binding domain (J4031_00315)	oxaloacetate decarboxylase (J4031_05015)	phosphate acetyltransferase (J4031_09705)	acetate kinase (J4031_09710)	phosphate butyryltransferase (J4031_09245)	butyrate kinase (J4031_09250)
AR32	Ga0104419_109/546 16..56181(-)/94%	Ga0104419_103/3 73631..375520(-) /96%	Ga0104419_103/371 446..373611(-)/94%	Ga0104419_103/ 377481..378422 (+)/79%	Ga0104419_103/3 78440..378895 (+)/80%	Ga0104419_103/378 899..380056 (+)/99%	Ga0104419_105/ 72188..72598(-) /100%	Ga0104419_105/ 70583..72148(-) /97%	Ga0104419_105/699 75..70403(-)/89%	Ga0104419_101/4 38545..440326(+) /99%	Ga0104419_110/ 57312..58502 (-) /95%	Ga0104419_110/ 57312..58502 (-) /96%	Ga0104419_102/752 16..76052(-)/96%	Ga0104419_102/74 149..75219 (-)/92%
BPI-162	Ga0070636_101/131 151..132656(-)/99%	Ga0070636_103/5 2475..54337(+) /98%	Ga0070636_103/494 54357..56522(+) /100%	Ga0070636_103/494 54..50479 (-)/98%	Ga0070636_103/478 8996..49448 (-) /98%	Ga0070636_103/478 07..48964 (-)/100%	Ga0070636_102/432 588..433043 (+)/100%	Ga0070636_102/434 33086..434651 (+)/100%	Ga0070636_102/434 843..435271(+)/100%	Ga0070636_113/5 8684..60465 (+)/99%	Ga0070636_101/ 803657..804667(+) /100%	Ga0070636_101/ 804699..805892 (+)/99%	Ga0070636_101/677 298..678134 (+)/100%	Ga0070636_101/67 8134..679204 (+)/99%
Ga686	T500DRAFT_scaffold0 00002.2/311087..31 2592 (+)/99%	T500DRAFT_scaffold ld00005.5/ 46970..48832 (+)/99%	T500DRAFT_scaffold 00005.5/48852..510 17 (+)/99%	T500DRAFT_scaffold 00005.5/43949..4497 4 (-)/99%	T500DRAFT_scaffo ld00005.5/43491.. 43943 (-)/100%	T500DRAFT_scaffold 00005.5/42303..434 60 (-)/99%	T500DRAFT_scaffold 00028.28/21876..223 31(-)/100%	T500DRAFT_scaffo ld00028.28/20272.. 21837 (-)/100%	T500DRAFT_scaffo ld00028.28/19672..201 00(-)/100%	T500DRAFT_scaffo ld00007.7/85671.. 87452(-)/99%	T500DRAFT_scaff ld000001.1/35652 0..357530 (+)/99%	T500DRAFT_scaff ld000001.1/35756 3..358756 (+)/99%	T500DRAFT_scaffo ld00001.1/243625..24 4461 (+)/99%	T500DRAFT_scaffo ld00001.1/244461..24 5531 (+)/99%
HUN156	IE21DRAFT_scaffold0 0001.1/353396..354 901(+)/96%	IE21DRAFT_scaffo ld00002.2/ 196859..198730(+) /70%	IE21DRAFT_scaffo ld0002.2/198733..200 898(+)/94%	IE21DRAFT_scaffo ld0002.2/193772..1948 09(-)/59%	IE21DRAFT_scaffo ld00002.2/192058..19 3209 (-)/93%	IE21DRAFT_scaffo ld00006.6/264673..2650 83 (+)/97%	IE21DRAFT_scaffo ld00006.6/265122.. 266687 (+)/95%	IE21DRAFT_scaffo ld00006.6/266928..2673 62 (+)/70%	IE21DRAFT_scaffo ld00002.2/218792.. 220564(-)/83%	IE21DRAFT_scaffo ld00001.1/228677 ..230913 (+)/92%	IE21DRAFT_scaffo ld00001.1/229723 ..229687 (+)/93%	IE21DRAFT_scaffo ld00005.5/212494..21 3318 (+)/51%	IE21DRAFT_scaffo ld00001.1/244461..24 5531 (+)/80%	IE21DRAFT_scaffo ld00005.5/213315..21 3382 (+)/80%
KHT3	Ga0104360_102/334 52..34957(+)/97%	Ga0104360_114/2 5348..27210 (+)/95%	Ga0104360_114/272 30..29395 (+)/97%	Ga0104360_114/223 39..23364 (-)/92%	Ga0104360_114/2 1869..22321(-) /92%	Ga0104360_114/206 81..21838(-)/99%	Ga0104360_120/134 98..13908(+)/97%	Ga0104360_120/1 3944..15509(+)/98 %	Ga0104360_120 - 15681..16109(+) /92%	Ga0104360_105/6 7077..68858 (+)/97%	Ga0104360_124/ 45234..46244 (-) /96%	Ga0104360_124 - 44012..45202(-) /97%	Ga0104360_121/356 85..36521 (+)/96%	Ga0104360_121/36 518..37588 (+)/91%
MA2016	T360DRAFT_scaffold 00002.2/465515..46 7020 (+)/98%	T360DRAFT_scaffo ld00004.4/135506.. 137368(-)/95%	T360DRAFT_scaffo ld00004.4/133318..13 5486 (-)/97%	T360DRAFT_scaffo ld00004.4/139380..140 405(+)/95%	T360DRAFT_scaffo ld00004.4/140412.. 140864 (+)/94%	T360DRAFT_scaffo ld00001.6/72021..672 431(+)/96%	T360DRAFT_scaffo ld00001.6/72508.. 674073 (+)/99%	T360DRAFT_scaffo ld00001.6/72508.. 674073 (+)/99%	T360DRAFT_scaffo ld00001.6/74248..674 676 (+)/92%	T360DRAFT_scaffo ld00008.8/87153.. 88934(-)/98%	T360DRAFT_scaff ld00005.5/64165 ..65175(-)/99%	T360DRAFT_scaff ld00005.5/62940 ..64133 (-)/99%	T360DRAFT_scaffo ld00005.5/193922..19 4758 (-)/98%	T360DRAFT_scaffo ld00005.5/192852..19 3922 (-)/97%
mn60	Ga0066892_116 - 34703..36208 (+) 95%	Ga0066892_101/4 41380..443224 (-) /71%	Ga0066892_101/439 209..441377(-)/94%	Ga0066892_101/445 248..446375(+)/59%	Ga0066892_101/446 46380..446877 (+)/55%	Ga0066892_101/446 903..448054 (+)/93%	Ga0066892_102/270 140..270550 (+)/97%	Ga0066892_102/2 70601..272166 (+)/95%	Ga0066892_102/272 364..272798 (+)/70%	Ga0066892_101/4 18086..419858 (+)/83%	Ga0066892_107/ 32056..33066 (-) /92%	Ga0066892_107/ 30830..32020 (-) /93%	Ga0066892_101/202 471..203295 (+)/53%	Ga0066892_101/20 3292..204452 (+)/73%
P6B4	T491DRAFT_scaffo ld00002.2/95462..969 67 (+)/97%	T491DRAFT_scaffo ld00007.7/46128.. 47990 (+)/95%	T491DRAFT_scaffo ld00007.7/48010..501 75 (+)/99%	T491DRAFT_scaffo ld00007.7/43143..4416 8 (-)/92%	T491DRAFT_scaffo ld00007.7/42685.. 43137 (-)/92%	T491DRAFT_scaffo ld00007.7/41496..426 53 (-)/99%	T491DRAFT_scaffo ld00011.11/108851..10 9261(-)/100%	T491DRAFT_scaffo ld00011.11/107247 ..108812 (-)/99%	T491DRAFT_scaffo ld00011.11/106651..10 7079 (-)/97%	T491DRAFT_scaffo ld000020/19988.. 21769(-)/99%	T491DRAFT_scaff ld000018.18/5431 9..55329(-)/99%	T491DRAFT_scaff ld000018.18/5309 4..54287(-)/99%	T491DRAFT_scaffo ld00006.6/85291..861 27 (+)/99%	T491DRAFT_scaffo ld00006.6/86127..871 97(+)/96%
P6B11	T496DRAFT_scaffo ld00014.14/5107..661 2 (-)/95%	T496DRAFT_scaffo ld00006.6/96554.. 98419 (+)/93%	T496DRAFT_scaffo ld00006.6/98552..100 717 (+)/96%	T496DRAFT_scaffo ld00015.15/25809..267 89 (+)/57%	T496DRAFT_scaffo ld00015.15/26797.. 27264 (+)/57%	T496DRAFT_scaffo ld00015.15/27298..28 449(+)/92%	T496DRAFT_scaffo ld00008.8/99749..1001 89 (+)/88%	T496DRAFT_scaffo ld00008.8/100198.. 101763 (+)/97%	T496DRAFT_scaffo ld00008.8/101947..102 381(+)/77%	T496DRAFT_scaffo ld00008.8/157803.. 159581(-)/84%	T496DRAFT_scaff ld00001.1/36669 ..37679(-)/95%	T496DRAFT_scaff ld00001.1/35344 ..36534 (-)/97%	T496DRAFT_scaffo ld00001.1/99577..100 473(+)/52%	T496DRAFT_scaffo ld00001.1/100554..10 1621 (+)/79%
RM4	T499DRAFT_scaffo ld00002.2/363638..36 5143 (+)/99%	T499DRAFT_scaffo ld00001.1/640564.. 642426 (+)/99%	T499DRAFT_scaffo ld00001.1/642446..64 4611 (+)/100%	T499DRAFT_scaffo ld00001.1/637577..638 602(-)/100%	T499DRAFT_scaffo ld00001.1/637119.. 637571(-)/99%	T499DRAFT_scaffo ld00001.1/635890..63 7047(-)/100%	T499DRAFT_scaffo ld00001.1/1132949..11 33404 (-)/100%	T499DRAFT_scaffo ld00001.1/1131341 ..1132906 (-)/100%	T499DRAFT_scaffo ld00001.1/1130720..11 31148 (-)/100%	T499DRAFT_scaffo ld00001.1/67968.. 69749 (+)/99%	T499DRAFT_scaff ld00003.3/32589 ..326906 (+)/99%	T499DRAFT_scaff ld00003.3/32697 ..328164 (+)/99%	T499DRAFT_scaffo ld00003.3/215664..21 6500 (+)/100%	T499DRAFT_scaffo ld00003.3/216500..21 7570 (+)/100%
TC2-24	Ga0066887_16/9162 2..93127 (-)/94%	Ga0066887_11/84 0336..842198 (-) /95%	Ga0066887_11/84 0336..842198 (-) /96%	Ga0066887_11/8441 79..845198 (+)/82%	Ga0066887_11/8451 79..845198 (+)/82%	Ga0066887_11/8457 18..846875(+)/98%	Ga0066887_11/5140 86..514496 (+)/96%	Ga0066887_11/51 4521..516086 (+)/98%	Ga0066887_11/5162 73..516704 (+)/89%	Ga0066887_14/21 6534..218312(+) /91%	Ga0066887_17/4 9518..50528 (+) /90%	Ga0066887_17/5 0538..51728 (-) /93%	Ga0066887_12/8648 6..87322(-)/95%	Ga0066887_12/853 65..86435 (-)/85%
TC2-28	Ga0066886_106 - 48054..49559 (-) 93%	Ga0066886_121 - 8400..10262 (-) /71%	Ga0066886_121/623 2..8397 (-)/94%	Ga0066886_121/123 47..13384 (+)/60%	Ga0066886_121 - 13440..13895 (+)/66%	Ga0066886_121/139 42..15099(+)/92%	Ga0066886_103/225 194..225604 (+)/92%	Ga0066886_103/2 25643..227208 (+)/95%	Ga0066886_103/227 403..227834 (+)/67%	Ga0066886_101/1 33078..134850(-) /83%	Ga0066886_106/ 64091..65101 (-) /92%	Ga0066886_106/ 62870..64060 (-) /95%	Ga0066886_122/222 85..23121 (+)/90%	Ga0066886_122/23 118..24185 (+)/77%
TF2-5	Ga0066888_105 - 231732..233237 (+) 91%	Ga0066888_102/1 14087..115943 (+)/84%	Ga0066888_102/115 960..118128 (+)/96%	Ga0066888_102 - 111021..112055 (+) /59%	Ga0066888_102/1 10527..110982(-) /67%	Ga0066888_102/109 329..110480(-)/93%	Ga0066888_109/441 41..44551(-)/91%	Ga0066888_109/4 2535..44100 (-) /95%	Ga0066888_109/418 81..42315 (-)/63%	Ga0066888_108/1 54367..156139 (+) /84%	Ga0066888_101/ 428169..429179(-) /92%	Ga0066888_101/ 426953..428146 (-) /93%	Ga0066888_102/267 373..268209 (+)/88%	Ga0066888_102/26 6309..267376 (-) /80%

For each strain (left column), data are shown as scaffold name(as shown in IMG)/gene coordinates on the scaffold/percentage homology to each corresponding candidate gene of the KHP1 genome, as calculated using blastp (default settings). Font is coloured based on the co-location of genes at loci of the KHP1 genome/scaffolds of the draft genomes, as shown in figure 4.4b.

## 7.10 Differentially abundant KHP1 genes

**Table 7.9. Significantly differentially expressed KHP1 genes between cobalamin treatments.**

gene ID	baseMean	Log <sub>2</sub> FC (+B <sub>12</sub> /-B <sub>12</sub> )	lfcSE	padj	PGAP annotation
J4031_01230	5676.6	-6.22	0.23	1.58E-155	sirohydrochlorin cobaltochelataase
J4031_01240	918.6	-5.87	0.20	3.67E-194	precorrin-3B C(17)-methyltransferase
J4031_01245	356.9	-5.72	0.25	2.33E-116	cobyric acid synthase
J4031_01235	1167.2	-5.58	0.22	1.66E-143	TonB-dependent receptor
J4031_01250	233.6	-5.58	0.25	1.56E-104	cobalamin biosynthesis protein CobD
J4031_07030	2409.7	-5.22	0.28	2.82E-75	O-acetylhomoserine aminocarboxypropyltransferase/cysteine synthase
J4031_01225	3355.9	-4.64	0.16	6.15E-189	TonB-dependent receptor
J4031_06995	1502.0	-4.22	0.15	1.47E-180	hypothetical protein cmsearch.
J4031_11665	8541.7	-3.75	0.20	2.13E-78	Lrp/AsnC ligand binding domain-containing protein
J4031_06475	3.0	-3.12	0.90	2.45E-03	tRNA-Ala
J4031_08510	348.0	-3.11	0.13	1.21E-131	hypothetical protein
J4031_08515	218.3	-3.05	0.17	4.82E-74	YncE family protein
J4031_08500	396.7	-2.59	0.15	2.62E-63	ABC transporter substrate-binding protein
J4031_08505	2664.4	-2.56	0.19	1.33E-39	DUF4465 domain-containing protein
J4031_08520	757.9	-2.55	0.11	1.16E-119	TonB-dependent receptor plug domain-containing protein
J4031_03120	102.1	-2.03	0.25	8.00E-14	hypothetical protein
J4031_07910	156.9	-1.86	0.27	1.56E-10	hypothetical protein
J4031_06940	260.0	-1.53	0.22	3.07E-10	response regulator transcription factor
J4031_02045	7.5	-1.47	0.51	1.25E-02	tRNA-Lys
J4031_07025	1295.0	-1.45	0.19	1.02E-12	cysteine synthase A
J4031_12275	90.8	-1.38	0.35	4.80E-04	AraC family transcriptional regulator
J4031_00215	4197.4	-1.37	0.35	6.63E-04	formate C-acetyltransferase
J4031_02090	138.2	-1.36	0.31	1.17E-04	hypothetical protein
J4031_01915	655.8	-1.31	0.33	5.47E-04	tRNA-Leu
J4031_13005	1521.8	-1.27	0.21	4.48E-08	hypothetical protein
J4031_05470	1426.6	-1.25	0.38	4.42E-03	RagB/SusD family nutrient uptake outer membrane protein

J4031_00405	31432.7	-1.23	0.23	1.71E-06	30S ribosomal protein S1
J4031_03995	367.6	-1.23	0.19	9.43E-09	AzIC family ABC transporter permease
J4031_00950	1171.0	-1.21	0.20	4.82E-08	4Fe-4S binding protein
J4031_09205	1155.5	-1.21	0.49	3.05E-02	class I mannose-6-phosphate isomerase
J4031_00790	308.1	-1.20	0.24	7.18E-06	YccF domain-containing protein
J4031_01265	3841.8	-1.20	0.15	6.35E-13	biosynthetic-type acetolactate synthase large subunit
J4031_10975	72.4	-1.19	0.43	1.55E-02	hypothetical protein
J4031_07795	8942.6	-1.19	0.24	1.03E-05	50S ribosomal protein L13
J4031_07775	338.5	-1.19	0.23	7.36E-06	carbohydrate kinase
J4031_03375	5580.1	-1.18	0.26	6.62E-05	50S ribosomal protein L9
J4031_03270	11.9	-1.18	0.48	3.06E-02	tRNA-Ala
J4031_01260	4866.1	-1.18	0.15	2.12E-13	dihydroxy-acid dehydratase
J4031_01920	8037.6	-1.18	0.30	5.92E-04	tRNA-Leu
J4031_03990	46.1	-1.18	0.29	3.94E-04	AzID domain-containing protein
J4031_06455	11111.2	-1.18	0.23	4.26E-06	DUF177 domain-containing protein
J4031_03275	54.8	-1.17	0.32	1.37E-03	tRNA-Ile
J4031_10215	1954.5	-1.16	0.26	9.89E-05	tRNA-Thr
J4031_11015	59.6	-1.16	0.21	6.54E-07	diaminopimelate epimerase
J4031_08200	532.4	-1.16	0.22	1.53E-06	TonB-dependent receptor
J4031_10350	47.7	-1.15	0.37	6.99E-03	helix-turn-helix transcriptional regulator
J4031_07780	10801.8	-1.15	0.24	1.92E-05	elongation factor Ts
J4031_10250	7457.5	-1.14	0.26	1.25E-04	50S ribosomal protein L10
J4031_10460	2656.4	-1.13	0.25	6.43E-05	30S ribosomal protein S5
J4031_07785	7551.2	-1.13	0.26	1.25E-04	30S ribosomal protein S2
J4031_03280	304.0	-1.13	0.41	1.76E-02	16S ribosomal RNA
J4031_11010	193.9	-1.13	0.24	3.04E-05	glutamate synthase subunit beta
J4031_10285	11862.6	-1.13	0.28	4.92E-04	30S ribosomal protein S10
J4031_05660	30087.5	-1.12	0.29	5.83E-04	30S ribosomal protein S15
J4031_05050	19146.1	-1.12	0.23	1.71E-05	50S ribosomal protein L25/general stress protein
J4031_05045	675.1	-1.11	0.30	1.08E-03	aminoacyl-tRNA hydrolase
J4031_03330	1890.5	-1.10	0.36	8.41E-03	tRNA methylthiotransferase MtaB



J4031_00015	1682.6	-1.09	0.33	4.37E-03	TonB-dependent receptor
J4031_07790	3114.2	-1.09	0.26	2.56E-04	30S ribosomal protein S9
J4031_10470	3032.8	-1.09	0.26	2.22E-04	50S ribosomal protein L15
J4031_12485	3138.1	-1.08	0.39	1.52E-02	sulfate permease
J4031_06240	352.4	-1.08	0.29	9.83E-04	MATE family efflux transporter
J4031_02580	771.6	-1.08	0.30	1.56E-03	tRNA-Glu
J4031_10970	148.5	-1.08	0.39	1.70E-02	asparagine synthase B
J4031_11005	853.2	-1.08	0.22	1.70E-05	glutamate synthase large subunit
J4031_10015	11.4	-1.06	0.40	2.23E-02	tRNA-Ile
J4031_03400	4826.2	-1.06	0.23	5.80E-05	HAMP domain-containing histidine kinase
J4031_00945	261.0	-1.06	0.22	2.61E-05	tetratricopeptide repeat protein
J4031_08205	498.7	-1.05	0.20	2.93E-06	TonB-dependent receptor
J4031_03460	3720.0	-1.05	0.29	1.51E-03	ABC-F family ATP-binding cassette domain-containing protein
J4031_03325	924.8	-1.05	0.36	1.04E-02	glycosyltransferase family 2 protein
J4031_10240	16316.2	-1.05	0.23	7.58E-05	50S ribosomal protein L11
J4031_10255	22134.3	-1.03	0.28	1.51E-03	50S ribosomal protein L7/L12
J4031_13115	7848.1	-1.03	0.30	2.77E-03	30S ribosomal protein S16
J4031_05745	22.7	-1.02	0.41	2.93E-02	hypothetical protein
J4031_07150	2284.1	-1.02	0.39	2.46E-02	pyridoxal 5'-phosphate synthase glutaminase subunit PdxT
J4031_04815	3182.6	-1.01	0.21	2.22E-05	Smr/MutS family protein
J4031_00210	377.6	-1.01	0.36	1.45E-02	pyruvate formate lyase-activating protein
J4031_09280	8492.7	-1.01	0.22	6.62E-05	30S ribosomal protein S20
J4031_10245	13752.4	-1.01	0.25	4.35E-04	50S ribosomal protein L1
J4031_11560	8446.3	-1.01	0.33	8.36E-03	efflux RND transporter permease subunit
J4031_04375	530.8	-1.00	0.21	1.94E-05	sulfur carrier protein ThiS analysis using gene prediction method: cmsearch.
J4031_03195	2292.4	-1.00	0.19	4.78E-06	putative transporter
J4031_07990	804.9	1.00	0.32	6.56E-03	ABC transporter substrate-binding protein
J4031_04330	766.0	1.01	0.34	9.28E-03	carboxylesterase family protein
J4031_12660	2234.5	1.01	0.40	2.96E-02	D-xylose transporter XylE
J4031_11835	762.9	1.01	0.22	4.46E-05	glycosyltransferase family 2 protein
J4031_12475	1016.0	1.01	0.34	9.11E-03	exodeoxyribonuclease III

J4031_14120	71.0	1.02	0.26	6.09E-04	nucleotidyltransferase family protein
J4031_03980	701.8	1.02	0.22	4.63E-05	hypothetical protein
J4031_12470	636.8	1.02	0.30	3.38E-03	hypothetical protein
J4031_00440	629.1	1.02	0.38	2.05E-02	restriction endonuclease subunit S
J4031_11125	108.1	1.02	0.38	1.96E-02	hypothetical protein
J4031_12125	2052.1	1.02	0.24	1.85E-04	hypothetical protein
J4031_10080	931.6	1.03	0.37	1.64E-02	DUF5110 domain-containing protein
J4031_03595	131.4	1.03	0.32	4.37E-03	AmmeMemoRadiSam system radical SAM enzyme
J4031_04710	458.3	1.03	0.39	2.03E-02	GNAT family N-acetyltransferase
J4031_02285	1119.1	1.03	0.40	2.66E-02	type II restriction endonuclease subunit M
J4031_10365	89.7	1.03	0.34	7.80E-03	DNA-binding protein
J4031_07480	110.6	1.03	0.44	4.23E-02	family 43 glycosylhydrolase
J4031_12315	3983.3	1.03	0.36	1.33E-02	ROK family protein
J4031_06610	120.8	1.04	0.31	3.99E-03	DUF2029 domain-containing protein
J4031_05465	287.4	1.04	0.41	2.76E-02	NAD(P)/FAD-dependent oxidoreductase
J4031_00450	1474.5	1.04	0.28	1.29E-03	type I restriction endonuclease subunit R
J4031_13305	1370.7	1.04	0.36	1.13E-02	hypothetical protein
J4031_12575	694.2	1.04	0.42	3.31E-02	hypothetical protein
J4031_10195	11832.6	1.04	0.39	1.91E-02	ribosome-associated translation inhibitor RaiA
J4031_03150	622.4	1.04	0.44	3.99E-02	Z1 domain-containing protein
J4031_07960	1151.2	1.04	0.29	1.62E-03	hypothetical protein
J4031_09245	962.1	1.04	0.39	1.97E-02	phosphate butyryltransferase
J4031_10950	25.5	1.04	0.39	2.07E-02	DUF3836 domain-containing protein
J4031_13320	1346.5	1.05	0.22	3.16E-05	hypothetical protein
J4031_11910	57.3	1.05	0.30	2.33E-03	N-acetylmuramoyl-L-alanine amidase
J4031_08545	1593.7	1.05	0.24	1.51E-04	aldo/keto reductase
J4031_11930	53.6	1.05	0.26	4.32E-04	AAA family ATPase
J4031_00465	125.8	1.05	0.46	4.85E-02	hypothetical protein
J4031_10075	334.4	1.05	0.36	1.13E-02	SusF/SusE family outer membrane protein
J4031_06050	1143.3	1.05	0.32	4.52E-03	TonB-dependent receptor plug domain-containing protein
J4031_11855	282.1	1.05	0.21	8.53E-06	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase

J4031_13720	2520.5	1.06	0.13	3.35E-14	DUF4842 domain-containing protein
J4031_11950	819.0	1.06	0.17	8.10E-09	molecular chaperone Tir
J4031_04440	76.9	1.06	0.38	1.60E-02	leucine-rich repeat protein
J4031_04390	571.4	1.06	0.28	1.05E-03	histidine acid phosphatase
J4031_05680	369.5	1.06	0.28	8.83E-04	M6 family metalloprotease domain-containing protein
J4031_00260	2628.8	1.06	0.16	1.54E-09	type II CRISPR RNA-guided endonuclease Cas9
J4031_13285	2587.1	1.06	0.42	2.73E-02	DNA-binding protein
J4031_10065	1905.2	1.06	0.30	2.25E-03	hypothetical protein
J4031_10725	2367.2	1.07	0.28	6.75E-04	glycine cleavage system aminomethyltransferase
J4031_09895	273.6	1.07	0.29	1.42E-03	hypothetical protein
J4031_11130	122.3	1.07	0.39	1.64E-02	hypothetical protein
J4031_09785	71.2	1.07	0.32	3.98E-03	hypothetical protein
J4031_12100	1728.7	1.07	0.18	1.45E-07	hypothetical protein
J4031_08005	1403.2	1.07	0.39	1.62E-02	right-handed parallel beta-helix repeat-containing protein
J4031_13710	294.3	1.07	0.26	3.21E-04	hypothetical protein
J4031_10730	990.6	1.07	0.25	1.74E-04	glycine cleavage system protein GcvH
J4031_10945	71.4	1.08	0.39	1.63E-02	FtsX-like permease family protein
J4031_12770	661.5	1.08	0.32	3.32E-03	endonuclease/exonuclease/phosphatase family protein
J4031_13325	70.0	1.08	0.26	2.72E-04	hypothetical protein
J4031_05485	24.8	1.08	0.45	3.67E-02	hypothetical protein
J4031_02545	1304.6	1.08	0.20	9.49E-07	AAA family ATPase
J4031_12390	4138.3	1.08	0.37	1.01E-02	large-conductance mechanosensitive channel protein MscL
J4031_11965	335.5	1.09	0.17	2.50E-09	hypothetical protein
J4031_11515	277.5	1.09	0.26	2.47E-04	hypothetical protein
J4031_13605	795.2	1.10	0.26	2.45E-04	DUF4422 domain-containing protein
J4031_10525	1012.8	1.10	0.39	1.33E-02	septal ring lytic transglycosylase RipA family protein
J4031_09885	549.0	1.10	0.29	6.63E-04	PD40 domain-containing protein
J4031_09850	414.2	1.11	0.31	1.76E-03	alpha-xylosidase
J4031_12075	689.9	1.11	0.21	3.93E-06	hypothetical protein
J4031_12090	970.7	1.11	0.22	1.23E-05	MBL fold metallo-hydrolase
J4031_08220	1052.2	1.11	0.26	1.38E-04	transporter substrate-binding domain-containing protein

J4031_03580	128.4	1.12	0.31	1.45E-03	DUF3874 domain-containing protein
J4031_03535	2001.8	1.12	0.27	3.21E-04	phage holin family protein
J4031_11955	995.9	1.12	0.17	1.11E-09	hypothetical protein
J4031_00460	173.4	1.12	0.27	2.18E-04	hypothetical protein
J4031_10690	561.7	1.12	0.23	1.96E-05	hypothetical protein
J4031_12005	1069.1	1.12	0.12	3.81E-19	polysaccharide pyruvyl transferase family protein
J4031_12000	1441.1	1.12	0.14	2.35E-14	Coenzyme F420 hydrogenase/dehydrogenase, beta subunit C-terminal domain
J4031_08085	1716.7	1.13	0.37	8.69E-03	GH92 family glycosyl hydrolase
J4031_09770	217.4	1.13	0.27	2.35E-04	peptidase C13
J4031_11960	1025.7	1.13	0.13	5.25E-16	hypothetical protein
J4031_09910	339.2	1.13	0.28	3.79E-04	hypothetical protein
J4031_12045	603.0	1.13	0.13	5.97E-16	O-antigen ligase family protein
J4031_08970	422.6	1.14	0.42	1.87E-02	thioredoxin-dependent thiol peroxidase
J4031_10085	1323.2	1.14	0.40	1.40E-02	discoidin domain-containing protein
J4031_13595	699.1	1.14	0.25	6.00E-05	glycosyltransferase family 4 protein
J4031_07405	13.0	1.14	0.43	1.91E-02	glycosyl hydrolase family 76
J4031_07640	635.9	1.15	0.43	2.11E-02	helix-turn-helix domain-containing protein
J4031_13755	749.5	1.15	0.24	2.43E-05	glycosyltransferase
J4031_09780	39.1	1.15	0.28	2.60E-04	hypothetical protein
J4031_07230	1856.6	1.15	0.21	1.29E-06	transporter substrate-binding domain-containing protein
J4031_04450	92.1	1.16	0.42	1.75E-02	hypothetical protein
J4031_06855	620.6	1.16	0.40	1.25E-02	aldo/keto reductase
J4031_11075	1563.2	1.16	0.37	6.19E-03	DUF853 family protein
J4031_02280	37.5	1.16	0.37	6.97E-03	DUF262 domain-containing protein analysis using gene prediction method: Protein Homology.
J4031_11820	1829.3	1.16	0.16	1.86E-11	polysaccharide pyruvyl transferase family protein
J4031_12030	694.3	1.16	0.12	5.96E-19	PIG-L family deacetylase
J4031_07600	901.4	1.16	0.42	1.53E-02	glycosyl hydrolase 53 family protein
J4031_13725	2161.4	1.16	0.12	1.57E-19	DUF4842 domain-containing protein
J4031_01375	283.6	1.16	0.33	2.09E-03	TIGR02172 family protein
J4031_13315	316.0	1.16	0.26	1.10E-04	phage antirepressor KilAC domain-containing protein

<b>J4031_13310</b>	370.2	1.17	0.24	1.16E-05	Bro-N domain-containing protein
<b>J4031_09625</b>	2356.6	1.17	0.25	2.69E-05	substrate-binding domain-containing protein
<b>J4031_10955</b>	70.3	1.17	0.31	1.09E-03	FtsX-like permease family protein
<b>J4031_12460</b>	2413.7	1.17	0.36	4.25E-03	glycosyltransferase family 4 protein
<b>J4031_11425</b>	1421.2	1.18	0.28	2.40E-04	DNA-binding protein
<b>J4031_07315</b>	7.1	1.18	0.50	4.17E-02	SusC/RagA family TonB-linked outer membrane protein
<b>J4031_11990</b>	917.8	1.18	0.20	2.21E-07	polysaccharide biosynthesis protein
<b>J4031_00860</b>	905.3	1.19	0.51	4.49E-02	GH92 family glycosyl hydrolase
<b>J4031_12320</b>	2824.9	1.19	0.41	1.11E-02	ROK family protein
<b>J4031_02315</b>	809.8	1.19	0.20	1.27E-07	DUF488 domain-containing protein
<b>J4031_06860</b>	275.4	1.19	0.38	5.90E-03	amidohydrolase
<b>J4031_04615</b>	352.9	1.19	0.35	2.98E-03	tetratricopeptide repeat protein
<b>J4031_12465</b>	4574.6	1.19	0.32	1.23E-03	glycoside hydrolase family 57 protein
<b>J4031_02540</b>	482.4	1.20	0.14	1.60E-14	hypothetical protein
<b>J4031_12570</b>	732.2	1.20	0.18	1.49E-09	FAD-dependent monooxygenase
<b>J4031_11830</b>	743.7	1.20	0.20	4.05E-08	glycosyltransferase family 2 protein
<b>J4031_08550</b>	1466.5	1.20	0.22	1.12E-06	4Fe-4S binding protein
<b>J4031_12845</b>	86.1	1.20	0.52	4.41E-02	RagB/SusD family nutrient uptake outer membrane protein
<b>J4031_02310</b>	617.6	1.21	0.17	3.09E-11	hypothetical protein
<b>J4031_12455</b>	6478.7	1.21	0.34	1.84E-03	glycogen debranching enzyme family protein
<b>J4031_10070</b>	1395.0	1.21	0.39	7.41E-03	DUF5110 domain-containing protein
<b>J4031_10740</b>	2438.1	1.21	0.37	4.16E-03	aminomethyl-transferring glycine dehydrogenase subunit GcvPB
<b>J4031_11735</b>	1309.7	1.21	0.39	7.41E-03	HAMP domain-containing histidine kinase
<b>J4031_10965</b>	55.8	1.21	0.39	7.63E-03	ABC transporter ATP-binding protein
<b>J4031_09765</b>	60.7	1.21	0.23	3.65E-06	hypothetical protein
<b>J4031_11845</b>	427.3	1.22	0.17	1.71E-10	glycosyltransferase family 2 protein
<b>J4031_00385</b>	469.2	1.23	0.40	7.43E-03	hypothetical protein
<b>J4031_07430</b>	28.4	1.23	0.34	1.78E-03	glycoside hydrolase family 125 protein
<b>J4031_01735</b>	906.3	1.24	0.36	2.67E-03	ABC transporter substrate-binding protein
<b>J4031_07985</b>	1252.0	1.24	0.30	2.56E-04	SpoIIE family protein phosphatase
<b>J4031_03145</b>	182.8	1.25	0.42	1.00E-02	PD-(D/E)XK motif protein

<b>J4031_03405</b>	9309.4	1.25	0.42	9.94E-03	elongation factor G
<b>J4031_12855</b>	25.1	1.25	0.47	2.10E-02	glycoside hydrolase family 30 protein
<b>J4031_12990</b>	2103.0	1.26	0.39	4.43E-03	hypothetical protein
<b>J4031_09250</b>	899.6	1.26	0.45	1.41E-02	butyrate kinase
<b>J4031_02530</b>	400.5	1.27	0.19	1.49E-09	hypothetical protein
<b>J4031_10595</b>	8.7	1.27	0.50	2.77E-02	family 43 glycosylhydrolase
<b>J4031_06045</b>	408.2	1.27	0.34	9.41E-04	DUF4249 domain-containing protein
<b>J4031_10560</b>	1537.8	1.27	0.40	5.47E-03	hypothetical protein
<b>J4031_02265</b>	21.9	1.27	0.45	1.39E-02	ATP-binding protein
<b>J4031_02275</b>	29.9	1.27	0.45	1.28E-02	DUF262 domain-containing protein
<b>J4031_10700</b>	298.7	1.27	0.20	1.46E-08	TlpA family protein disulfide reductase
<b>J4031_12290</b>	221.6	1.28	0.34	9.45E-04	MGMT family protein
<b>J4031_12850</b>	37.7	1.28	0.43	1.03E-02	xylanase
<b>J4031_02270</b>	83.1	1.28	0.36	1.56E-03	hypothetical protein
<b>J4031_12840</b>	169.8	1.28	0.44	1.08E-02	TonB-dependent receptor
<b>J4031_13280</b>	132.0	1.28	0.32	3.61E-04	smalltalk protein
<b>J4031_10710</b>	3002.7	1.28	0.47	1.72E-02	MIP family channel protein
<b>J4031_11920</b>	16.2	1.29	0.31	3.13E-04	hypothetical protein
<b>J4031_13715</b>	158.4	1.29	0.42	7.19E-03	hypothetical protein
<b>J4031_03140</b>	518.1	1.29	0.40	5.07E-03	AIPR family protein
<b>J4031_09890</b>	166.3	1.29	0.37	2.40E-03	DUF4369 domain-containing protein
<b>J4031_07820</b>	1438.1	1.30	0.23	7.16E-07	9-O-acetylcysteine aminotransferase
<b>J4031_01775</b>	99.0	1.30	0.23	5.20E-07	PriCT-2 domain-containing protein
<b>J4031_12010</b>	427.5	1.30	0.18	1.43E-11	acyltransferase
<b>J4031_12070</b>	1201.5	1.30	0.22	8.11E-08	WecB/TagA/CpsF family glycosyltransferase
<b>J4031_04460</b>	128.0	1.31	0.25	3.69E-06	hypothetical protein
<b>J4031_12025</b>	1147.7	1.31	0.16	3.02E-15	glycosyltransferase family 4 protein
<b>J4031_04640</b>	9.6	1.33	0.53	2.83E-02	hypothetical protein
<b>J4031_01760</b>	13.8	1.34	0.50	2.04E-02	hypothetical protein
<b>J4031_08055</b>	302.7	1.35	0.34	4.28E-04	sialate O-acetylcysteine aminotransferase
<b>J4031_11825</b>	1403.5	1.35	0.21	4.10E-09	glycosyltransferase family 2 protein

J4031_12015	413.1	1.35	0.17	3.60E-14	serine acetyltransferase
J4031_04600	296.3	1.36	0.39	2.29E-03	DUF4062 domain-containing protein
J4031_09255	738.1	1.36	0.40	2.80E-03	HAMP domain-containing protein
J4031_12295	253.8	1.36	0.41	3.99E-03	NAD(P)H-dependent oxidoreductase
J4031_02535	168.3	1.36	0.25	1.21E-06	sigma-70 family RNA polymerase sigma factor
J4031_10745	1191.1	1.37	0.31	9.99E-05	NAD(P)/FAD-dependent oxidoreductase
J4031_04285	611.0	1.37	0.56	3.43E-02	beta-glucosidase BglX
J4031_08905	1945.9	1.37	0.47	1.14E-02	cupin domain-containing protein
J4031_06875	681.6	1.38	0.39	2.29E-03	glycoside hydrolase family 97 protein
J4031_12050	729.1	1.38	0.15	3.72E-18	glycosyltransferase family 2 protein
J4031_06870	250.4	1.38	0.41	2.93E-03	L-fucose:H <sup>+</sup> symporter permease
J4031_07585	507.2	1.38	0.47	1.00E-02	glycoside hydrolase 43 family protein
J4031_02985	92.6	1.38	0.43	4.90E-03	beta-galactosidase
J4031_12565	5857.3	1.38	0.29	3.19E-05	methylmalonyl-CoA mutase
J4031_10565	492.4	1.39	0.33	1.93E-04	carboxylesterase/lipase family protein
J4031_00855	680.4	1.39	0.53	2.20E-02	copper homeostasis protein CutC
J4031_13780	251.1	1.40	0.19	2.36E-11	hypothetical protein
J4031_09900	375.3	1.40	0.25	8.51E-07	PD40 domain-containing protein
J4031_08915	66.2	1.40	0.47	8.98E-03	DNA-binding protein
J4031_04465	144.2	1.40	0.30	4.41E-05	NADAR family protein
J4031_09905	400.2	1.40	0.29	2.48E-05	PD40 domain-containing protein
J4031_14260	1033.0	1.40	0.44	5.45E-03	HAMP domain-containing protein
J4031_12020	965.7	1.40	0.14	1.64E-20	glycosyltransferase family 4 protein
J4031_12060	350.2	1.41	0.16	8.63E-16	acyltransferase
J4031_06830	480.0	1.42	0.38	1.21E-03	ABC transporter substrate-binding protein
J4031_04545	15.4	1.43	0.46	7.23E-03	hypothetical protein
J4031_10695	241.3	1.43	0.21	5.86E-10	thioredoxin family protein
J4031_05930	1562.5	1.43	0.28	8.61E-06	transporter substrate-binding domain-containing protein
J4031_08480	1177.7	1.44	0.50	1.23E-02	alpha-galactosidase
J4031_01360	1039.1	1.44	0.29	8.56E-06	DUF3256 family protein
J4031_09260	792.9	1.44	0.40	1.65E-03	HAMP domain-containing protein

J4031_05920	1317.8	1.45	0.29	8.62E-06	PAS domain S-box protein
J4031_00455	18.8	1.45	0.48	9.23E-03	hypothetical protein
J4031_12665	1582.6	1.45	0.44	3.73E-03	response regulator
J4031_09775	100.2	1.46	0.24	4.25E-08	hypothetical protein
J4031_03135	217.8	1.46	0.44	3.72E-03	leucine-rich repeat domain-containing protein
J4031_05925	1215.7	1.46	0.27	2.36E-06	PAS domain-containing sensor histidine kinase
J4031_05935	1772.8	1.47	0.34	1.32E-04	transporter substrate-binding domain-containing protein
J4031_12140	106.4	1.47	0.54	1.73E-02	glycoside hydrolase family 130 protein
J4031_07615	5.5	1.47	0.57	2.45E-02	family 43 glycosylhydrolase
J4031_12740	720.7	1.48	0.40	1.08E-03	HAMP domain-containing histidine kinase
J4031_08910	65.6	1.49	0.50	9.30E-03	N-acetylmuramoyl-L-alanine amidase
J4031_00390	359.8	1.49	0.50	8.94E-03	HAMP domain-containing protein
J4031_08835	12.4	1.49	0.65	4.85E-02	DUF2264 domain-containing protein
J4031_05510	15.3	1.49	0.45	3.76E-03	hypothetical protein
J4031_12325	3321.4	1.49	0.42	1.66E-03	ROK family protein
J4031_08850	29.4	1.50	0.57	2.21E-02	glycoside hydrolase family 88 protein
J4031_13185	38.2	1.50	0.42	1.59E-03	flavodoxin family protein
J4031_12065	995.7	1.52	0.21	1.97E-11	glycosyltransferase
J4031_00395	453.1	1.52	0.53	1.23E-02	sensor histidine kinase
J4031_05880	1683.1	1.52	0.40	8.46E-04	alpha-glucuronidase
J4031_12835	124.6	1.53	0.46	3.88E-03	hypothetical protein
J4031_11035	22.1	1.53	0.59	2.37E-02	CotH kinase family protein
J4031_12560	5293.5	1.54	0.28	9.11E-07	methylmalonyl-CoA mutase small subunit
J4031_11995	1429.8	1.55	0.25	2.70E-08	glycosyltransferase family 2 protein
J4031_07200	1810.1	1.55	0.37	2.18E-04	mannose-1-phosphate guanylyltransferase
J4031_12760	3629.4	1.55	0.50	7.01E-03	FAD-binding oxidoreductase
J4031_13180	112.0	1.56	0.41	9.92E-04	alpha/beta hydrolase
J4031_09790	23.9	1.56	0.45	2.38E-03	cyclohexadienyl dehydratase
J4031_10040	662.8	1.56	0.37	2.35E-04	cob(I)yrinic acid a,c-diamide adenosyltransferase
J4031_01740	1252.2	1.56	0.39	4.08E-04	SpoIIE family protein phosphatase
J4031_07595	475.3	1.57	0.44	1.94E-03	acetylxylen esterase



J4031_13200	97.3	1.58	0.38	3.00E-04	4Fe-4S dicluster domain-containing protein
J4031_07975	8.5	1.58	0.59	1.91E-02	P1 family peptidase
J4031_08095	1421.3	1.59	0.34	3.05E-05	GH92 family glycosyl hydrolase
J4031_04630	58.6	1.59	0.40	4.47E-04	TIR domain-containing protein
J4031_07340	1048.0	1.59	0.66	3.53E-02	ABC transporter substrate-binding protein
J4031_04625	40.2	1.59	0.37	1.82E-04	DNA/RNA non-specific endonuclease
J4031_04585	951.3	1.60	0.44	1.51E-03	hypothetical protein
J4031_07825	836.1	1.61	0.23	6.60E-11	NUDIX domain-containing protein
J4031_04605	709.1	1.61	0.44	1.51E-03	SEL1-like repeat protein
J4031_04620	244.5	1.61	0.40	4.38E-04	tetratricopeptide repeat protein
J4031_03585	97.7	1.61	0.46	2.23E-03	hypothetical protein
J4031_14090	495.2	1.62	0.49	3.67E-03	sulfide/dihydroorotate dehydrogenase-like
J4031_07425	13.3	1.63	0.56	1.07E-02	GH92 family glycosyl hydrolase
J4031_12765	1932.9	1.63	0.41	5.04E-04	carboxylesterase family protein
J4031_04650	471.2	1.63	0.54	8.81E-03	hypothetical protein
J4031_07490	831.2	1.64	0.47	2.39E-03	response regulator
J4031_04305	6.3	1.64	0.67	3.41E-02	anaerobic sulfatase maturase
J4031_04580	837.2	1.65	0.43	7.61E-04	hypothetical protein
J4031_08060	413.0	1.65	0.51	5.07E-03	MFS transporter
J4031_01105	998.4	1.65	0.44	9.81E-04	SusC/RagA family TonB-linked outer membrane protein
J4031_08755	5180.2	1.65	0.41	3.42E-04	TonB-dependent receptor
J4031_04610	285.0	1.66	0.39	1.75E-04	DUF4062 domain-containing protein
J4031_04635	126.5	1.67	0.34	1.70E-05	hypothetical protein
J4031_08090	694.7	1.68	0.32	4.18E-06	HAMP domain-containing histidine kinase
J4031_07140	299.8	1.70	0.49	2.33E-03	DUF4421 family protein
J4031_07145	488.2	1.70	0.44	6.98E-04	DUF4982 domain-containing protein
J4031_05490	83.8	1.71	0.31	1.30E-06	hypothetical protein
J4031_08050	261.3	1.72	0.42	2.77E-04	MFS transporter
J4031_06825	1135.2	1.72	0.37	4.02E-05	SpoIIE family protein phosphatase
J4031_12055	507.5	1.73	0.28	2.53E-08	glycosyl transferase
J4031_07995	5100.5	1.73	0.57	7.65E-03	RagB/SusD family nutrient uptake outer membrane protein

J4031_11430	66.2	1.73	0.32	1.99E-06	smalltalk protein
J4031_14095	594.1	1.73	0.51	2.70E-03	FAD-dependent oxidoreductase
J4031_07330	796.2	1.74	0.52	3.83E-03	glycosyl hydrolase 115 family protein
J4031_12780	1889.0	1.74	0.55	6.09E-03	RagB/SusD family nutrient uptake outer membrane protein
J4031_01580	7.6	1.74	0.75	4.52E-02	hypothetical protein
J4031_11690	937.4	1.75	0.49	1.88E-03	phosphatase PAP2 family protein
J4031_08070	62.5	1.75	0.59	1.02E-02	sigma-70 family RNA polymerase sigma factor
J4031_07720	316.7	1.76	0.68	2.42E-02	glycoside hydrolase family 3 C-terminal domain-containing protein
J4031_07620	17.4	1.77	0.67	2.10E-02	glycoside hydrolase family 43 protein
J4031_07980	1624.9	1.77	0.31	3.60E-07	histidine-type phosphatase
J4031_07370	34.5	1.78	0.51	2.48E-03	hypothetical protein
J4031_08900	885.3	1.78	0.57	6.01E-03	cytidylate kinase-like family protein
J4031_09265	423.1	1.78	0.31	2.28E-07	TlpA family protein disulfide reductase
J4031_07420	24.5	1.79	0.49	1.52E-03	TonB-dependent receptor
J4031_01785	120.8	1.79	0.37	2.00E-05	hypothetical protein
J4031_11050	49.5	1.79	0.65	1.69E-02	RagB/SusD family nutrient uptake outer membrane protein
J4031_12380	613.2	1.80	0.48	1.05E-03	TonB-dependent receptor
J4031_01555	12.2	1.80	0.60	9.02E-03	smalltalk protein
J4031_04590	1199.7	1.80	0.43	2.05E-04	tetratricopeptide repeat protein
J4031_04495	311.1	1.81	0.34	2.64E-06	PD-(D/E)XK nuclease family protein
J4031_08860	162.8	1.81	0.64	1.37E-02	TonB-dependent receptor
J4031_08000	10859.6	1.81	0.59	7.69E-03	TonB-dependent receptor
J4031_09270	371.9	1.82	0.32	3.20E-07	tetratricopeptide repeat protein
J4031_05535	6.2	1.82	0.67	1.78E-02	smalltalk protein
J4031_12775	1945.7	1.83	0.55	3.63E-03	DUF4957 domain-containing protein
J4031_14100	221.6	1.83	0.46	5.14E-04	MATE family efflux transporter
J4031_14265	767.2	1.83	0.47	5.32E-04	sensor histidine kinase
J4031_12785	4379.5	1.84	0.53	2.64E-03	TonB-dependent receptor
J4031_04395	441.3	1.86	0.59	5.90E-03	ROK family protein
J4031_01100	346.5	1.87	0.46	4.08E-04	RagB/SusD family nutrient uptake outer membrane protein
J4031_03475	1113.8	1.87	0.51	1.35E-03	cytidylate kinase-like family protein

J4031_09865	913.1	1.88	0.62	7.92E-03	RagB/SusD family nutrient uptake outer membrane protein
J4031_07410	8.4	1.89	0.62	8.28E-03	SusE domain-containing protein
J4031_01545	743.1	1.90	0.52	1.48E-03	mechanosensitive ion channel family protein
J4031_07450	46.9	1.91	0.53	1.61E-03	cellulase family glycosylhydrolase
J4031_11025	72.1	1.91	0.48	4.43E-04	glycoside hydrolase family 3 C-terminal domain-containing protein
J4031_08840	7.1	1.91	0.70	1.65E-02	DUF4861 domain-containing protein
J4031_07445	59.1	1.92	0.60	4.86E-03	acetyl xylan esterase
J4031_07375	45.3	1.92	0.53	1.51E-03	lipocalin family protein
J4031_07715	472.7	1.92	0.68	1.37E-02	esterase
J4031_08855	112.4	1.93	0.61	5.80E-03	RagB/SusD family nutrient uptake outer membrane protein
J4031_08065	663.6	1.94	0.50	6.95E-04	glycoside hydrolase family 43 protein
J4031_06170	221.0	1.95	0.46	1.75E-04	methylase
J4031_02140	10547.5	1.96	0.50	6.40E-04	NADH peroxidase
J4031_09870	942.2	1.96	0.64	8.02E-03	hypothetical protein
J4031_07590	414.8	1.97	0.58	2.90E-03	prolyl oligopeptidase family serine peptidase
J4031_05500	35.8	1.97	0.39	5.44E-06	hypothetical protein
J4031_13985	139.3	1.98	0.56	2.08E-03	response regulator
J4031_11045	126.1	2.01	0.64	6.52E-03	hypothetical protein
J4031_07630	26.6	2.01	0.74	1.80E-02	RagB/SusD family nutrient uptake outer membrane protein
J4031_12385	317.4	2.01	0.50	4.34E-04	RagB/SusD family nutrient uptake outer membrane protein
J4031_05505	11.6	2.02	0.53	7.89E-04	hypothetical protein
J4031_06180	514.2	2.02	0.40	7.42E-06	EVE domain-containing protein
J4031_09860	2544.2	2.03	0.61	3.51E-03	TonB-dependent receptor
J4031_07510	157.0	2.03	0.73	1.49E-02	esterase
J4031_04310	34.2	2.07	0.55	9.64E-04	sulfatase-like hydrolase/transferase
J4031_02500	3.2	2.08	0.91	4.83E-02	hypothetical protein
J4031_05530	210.2	2.08	0.32	5.37E-09	DNA-binding protein
J4031_04655	283.8	2.08	0.55	9.46E-04	RyR domain protein
J4031_08865	46.9	2.08	0.58	1.56E-03	family 43 glycosylhydrolase
J4031_13190	31.3	2.09	0.43	1.47E-05	hypothetical protein
J4031_11055	222.1	2.12	0.63	3.13E-03	TonB-dependent receptor

<b>J4031_07625</b>	21.7	2.13	0.76	1.51E-02	hypothetical protein
<b>J4031_01565</b>	115.6	2.14	0.58	1.18E-03	WYL domain-containing protein
<b>J4031_11040</b>	59.8	2.14	0.64	3.47E-03	family 16 glycosylhydrolase
<b>J4031_07355</b>	251.8	2.17	0.61	2.05E-03	endo-1,4-beta-xylanase
<b>J4031_04595</b>	883.0	2.17	0.51	1.95E-04	DUF4062 domain-containing protein
<b>J4031_13195</b>	22.3	2.17	0.51	2.03E-04	antibiotic biosynthesis monooxygenase
<b>J4031_14125</b>	419.0	2.19	0.58	9.21E-04	hypothetical protein
<b>J4031_08845</b>	9.1	2.20	0.69	5.24E-03	heparinase II/III family protein
<b>J4031_12810</b>	20.5	2.21	0.80	1.59E-02	TonB-dependent receptor
<b>J4031_07275</b>	157.4	2.23	0.62	1.56E-03	family 43 glycosylhydrolase
<b>J4031_04500</b>	311.9	2.25	0.48	4.26E-05	UvrD-helicase domain-containing protein
<b>J4031_07270</b>	123.5	2.25	0.70	5.07E-03	family 43 glycosylhydrolase
<b>J4031_01575</b>	222.0	2.26	0.55	3.41E-04	hypothetical protein
<b>J4031_00745</b>	58.2	2.26	0.64	1.93E-03	DUF262 domain-containing protein
<b>J4031_01320</b>	1989.5	2.27	0.63	1.67E-03	anaerobic C4-dicarboxylate transporter
<b>J4031_13350</b>	2902.2	2.27	0.55	2.85E-04	response regulator
<b>J4031_11030</b>	28.7	2.29	0.56	3.09E-04	glycoside hydrolase family 5 protein
<b>J4031_04400</b>	458.8	2.30	0.60	7.91E-04	glycoside hydrolase family 2
<b>J4031_14130</b>	472.9	2.30	0.63	1.36E-03	leucine-rich repeat domain-containing protein
<b>J4031_06175</b>	280.2	2.31	0.47	1.73E-05	hypothetical protein
<b>J4031_06185</b>	187.7	2.31	0.47	1.12E-05	5-methylcytosine-specific restriction endonuclease system specificity protein McrC
<b>J4031_01570</b>	6.8	2.32	0.70	3.88E-03	hypothetical protein
<b>J4031_05200</b>	70.6	2.33	0.57	3.38E-04	hypothetical protein
<b>J4031_06145</b>	777.7	2.34	0.50	4.02E-05	leucine-rich repeat protein
<b>J4031_07635</b>	69.4	2.38	0.70	2.84E-03	TonB-dependent receptor
<b>J4031_07515</b>	458.2	2.39	0.75	5.34E-03	glycoside hydrolase family 3 C-terminal domain-containing protein
<b>J4031_11790</b>	4082.4	2.39	0.60	4.92E-04	HAMP domain-containing histidine kinase
<b>J4031_07415</b>	8.6	2.46	0.66	1.08E-03	RagB/SusD family nutrient uptake outer membrane protein
<b>J4031_01755</b>	14.6	2.47	0.49	7.93E-06	hypothetical protein
<b>J4031_05020</b>	2880.1	2.48	0.66	1.01E-03	hypothetical protein
<b>J4031_00120</b>	177.1	2.51	0.61	3.21E-04	DUF4492 domain-containing protein

J4031_06325	726.9	2.51	0.60	2.47E-04	hypothetical protein
J4031_07255	61.1	2.52	0.62	3.33E-04	hypothetical protein
J4031_03110	543.3	2.53	0.63	3.79E-04	hypothetical protein
J4031_07440	79.7	2.54	0.45	3.43E-07	TonB-dependent receptor
J4031_07435	39.8	2.56	0.45	2.93E-07	RagB/SusD family nutrient uptake outer membrane protein
J4031_05210	262.6	2.59	0.57	6.88E-05	OmpA family protein
J4031_04990	73.5	2.59	0.69	1.07E-03	cytochrome c biogenesis protein CcsA
J4031_01595	1737.0	2.62	0.60	1.31E-04	cytidylate kinase-like family protein
J4031_11685	854.5	2.65	0.67	4.59E-04	NAD-dependent dihydropyrimidine dehydrogenase subunit PreA
J4031_04985	80.8	2.66	0.69	6.75E-04	cytochrome c biogenesis protein ResB
J4031_08225	1725.7	2.67	0.66	3.74E-04	rubredoxin
J4031_05970	239.5	2.67	0.61	1.25E-04	hypothetical protein
J4031_01560	144.0	2.69	0.46	2.03E-07	DNA-binding protein
J4031_04855	1750.7	2.70	0.63	1.58E-04	MBL fold metallo-hydrolase
J4031_12815	12.8	2.70	0.79	2.78E-03	RagB/SusD family nutrient uptake outer membrane protein
J4031_07265	167.9	2.70	0.63	1.58E-04	glycosyl hydrolase family 31
J4031_07130	4.1	2.71	1.01	1.91E-02	peptidase M15
J4031_06155	86.9	2.76	0.52	2.94E-06	hypothetical protein
J4031_05205	630.6	2.77	0.64	1.41E-04	hypothetical protein
J4031_01055	386.7	2.80	0.75	1.12E-03	anaerobic C4-dicarboxylate transporter
J4031_01550	451.7	2.83	0.55	5.37E-06	hypothetical protein
J4031_00710	123.5	2.87	0.73	5.90E-04	hypothetical protein
J4031_02490	162.4	2.88	0.71	3.67E-04	hypothetical protein
J4031_07260	123.9	2.88	0.75	6.82E-04	hypothetical protein
J4031_00750	19.8	2.91	0.68	1.58E-04	hypothetical protein
J4031_04975	430.6	2.93	0.71	2.97E-04	NapC/NirT family cytochrome c
J4031_05585	319.5	2.94	0.67	1.25E-04	hypothetical protein
J4031_13345	23244.6	2.95	0.77	7.72E-04	TonB-dependent receptor
J4031_11675	326.2	3.00	0.74	3.42E-04	metallophosphoesterase
J4031_06150	109.8	3.05	0.57	2.31E-06	hypothetical protein
J4031_04665	1013.4	3.07	0.50	2.59E-08	VWA domain-containing protein

J4031_06160	160.1	3.07	0.54	3.10E-07	hypothetical protein
J4031_00115	1215.7	3.08	0.92	3.51E-03	cytochrome ubiquinol oxidase subunit I
J4031_01585	219.1	3.09	0.81	8.83E-04	STAS domain-containing protein
J4031_13340	13049.6	3.12	0.76	2.87E-04	RagB/SusD family nutrient uptake outer membrane protein
J4031_10620	1350.3	3.16	0.76	2.68E-04	alanine dehydrogenase
J4031_11680	231.9	3.16	0.73	1.46E-04	hypothetical protein
J4031_00715	222.2	3.21	0.81	5.28E-04	hypothetical protein
J4031_07520	56764.2	3.24	0.77	2.35E-04	hypothetical protein
J4031_04695	118.2	3.32	1.01	4.10E-03	hypothetical protein
J4031_05580	440.7	3.34	0.71	3.24E-05	hypothetical protein
J4031_04690	146.1	3.36	0.98	2.67E-03	two pore domain potassium channel family protein
J4031_00725	271.2	3.36	0.93	1.61E-03	PD-(D/E)XK nuclease family protein
J4031_04980	1267.4	3.36	0.72	3.50E-05	ammonia-forming cytochrome c nitrite reductase
J4031_09615	153.1	3.38	0.82	2.96E-04	RagB/SusD family nutrient uptake outer membrane protein
J4031_07460	1264.3	3.45	1.17	1.02E-02	TonB-dependent receptor
J4031_08825	1296.6	3.46	0.97	1.66E-03	TonB-dependent receptor
J4031_00110	812.2	3.49	0.98	1.81E-03	cytochrome d ubiquinol oxidase subunit II
J4031_09610	133.3	3.50	0.75	4.36E-05	DUF4960 domain-containing protein
J4031_02495	224.5	3.51	0.72	1.73E-05	hypothetical protein
J4031_09605	70.1	3.51	0.81	1.34E-04	glycoside hydrolase family 32 protein
J4031_07465	1014.6	3.58	1.15	6.70E-03	RagB/SusD family nutrient uptake outer membrane protein
J4031_05350	1087.4	3.58	0.72	1.15E-05	(Fe-S)-binding protein
J4031_04680	878.1	3.67	1.04	2.11E-03	SUMF1/EgtB/PvdO family nonheme iron enzyme
J4031_09620	458.6	3.67	0.77	2.75E-05	TonB-dependent receptor
J4031_12755	3470.3	3.68	0.71	3.94E-06	L-lactate permease
J4031_00720	89.2	3.68	1.01	1.42E-03	hypothetical protein
J4031_01045	791.9	3.69	0.96	7.79E-04	aspartate ammonia-lyase
J4031_08830	1217.3	3.74	0.96	6.55E-04	RagB/SusD family nutrient uptake outer membrane protein
J4031_04675	830.6	3.74	1.02	1.34E-03	hypothetical protein
J4031_04685	220.0	3.76	1.11	3.08E-03	hypothetical protein
J4031_04670	2198.5	3.77	0.94	4.22E-04	hypothetical protein

<b>J4031_00730</b>	160.4	3.78	0.96	5.04E-04	phospholipase
<b>J4031_05575</b>	283.7	3.79	0.60	1.43E-08	hypothetical protein
<b>J4031_05355</b>	1272.3	3.84	0.79	1.73E-05	lactate utilisation protein
<b>J4031_09595</b>	2215.7	3.93	0.97	3.74E-04	DUF4980 domain-containing protein
<b>J4031_09585</b>	1155.4	4.08	0.99	3.00E-04	carbohydrate kinase
<b>J4031_01050</b>	86.0	4.11	1.16	2.02E-03	hypothetical protein
<b>J4031_05360</b>	665.0	4.11	0.80	4.81E-06	LUD domain-containing protein
<b>J4031_09635</b>	1046.6	4.12	1.05	5.37E-04	DUF4960 domain-containing protein
<b>J4031_09590</b>	1233.0	4.25	1.02	2.57E-04	MFS transporter
<b>J4031_09630</b>	2552.4	4.38	1.02	1.68E-04	DUF4975 domain-containing protein
<b>J4031_09645</b>	9664.1	4.48	1.02	1.15E-04	TonB-dependent receptor
<b>J4031_09640</b>	2467.3	4.54	1.03	1.03E-04	RagB/SusD family nutrient uptake outer membrane protein

## 7.11 Differentially abundant KHP1 proteins

**Table 7.10. Significantly differentially abundant KHP1 proteins between cobalamin treatments.**

Protein ID	PGAP annotation	Student's T-test q-value <sup>1</sup>	Log <sub>2</sub> fold change (+B <sub>12</sub> /-B <sub>12</sub> )	T-test statistic (+B <sub>12</sub> /-B <sub>12</sub> )
QVJ81054.1	sirohhydrochlorin cobaltochelataase	8.00E-04	-6.72	-10.93
QVJ81053.1	TonB-dependent receptor	0.00E+00	-5.26	-20.92
QVJ82096.1	O-acetylhomoserine aminocarboxypropyltransferase/cysteine synthase	3.14E-03	-3.31	-7.44
QVJ81056.1	precorrin-3B C methyltransferase	1.00E-03	-3.27	-13.17
QVJ79760.1	DUF4465 containing protein	1.33E-03	-2.83	-14.03
QVJ81055.1	TonB-dependent receptor	0.00E+00	-2.20	-18.03
QVJ81057.1	cobyric acid synthase	6.67E-04	-2.20	-10.28
QVJ82080.1	response regulator transcription factor	4.27E-03	-1.20	-7.08
QVJ81973.1	5'/3'-nucleotidase SurE	4.05E-02	-1.20	-3.98
QVJ82095.1	cysteine synthase A	4.25E-03	-1.05	-7.06
QVJ80504.1	methylmalonyl-CoA mutase	1.00E-03	1.65	9.66
QVJ80503.1	methylmalonyl-CoA mutase small subunit	8.89E-04	1.71	9.54
QVJ80162.1	glycine cleavage system protein GcvH	8.17E-03	1.75	5.91

<sup>1</sup> Q-values were calculated using the false discovery rate (FDR) (Benjamini & Hochberg, 1995).