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.

Comparative genomics of *Butyrivibrio* and *Pseudobutyrivibrio* from the rumen

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

Doctor of Philosophy

in

Microbiology and Genetics

at Massey University, Manawatū, New Zealand.

Nikola Palevich

2016

Abstract

Determining the role of rumen microbes in plant polysaccharide breakdown is fundamental to understanding digestion, and maximising productivity, in ruminant animals. Rumen bacterial species belonging to the genera *Butyrivibrio* and *Pseudobutyrivibrio* are important degraders of plant hemicellulose, an abundant heterogeneous, branched polymer, involved in crosslinking cellulose microfibrils to lignin. To investigate their genes required for hemicellulose degradation, the genomes of 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains isolated from the plant-adherent microbiome of New Zealand bovine ruminants, were sequenced, and their CAZyme-encoding genes compared. Within the *Butyrivibrio* and *Pseudobutyrivibrio* pan-genomes, respectively, there were a total of 4,421 and 441 glycoside hydrolases, as well as 1,283 and 122 carbohydrate esterases with predicted activities involved in the degradation of the insoluble plant polysaccharides such as xylan and pectin. To examine species differences, the genes of the previously characterised bacterium *B. proteoclasticus* B316 were compared in detail with those from the newly sequenced *B. hungatei* MB2003. B316 was found to encode a much more developed polysaccharide-degrading repertoire and it was thus hypothesised that B316 would out-compete MB2003 when grown in co-culture on the insoluble hemicellulose substrate, xylan. To test this hypothesis, the two strains were grown on xylan and pectin, either alone in mono-cultures, or in direct competition in a co-culture. The results showed that MB2003 had little ability to utilise xylan or pectin alone, but was capable of significant growth when co-cultured with B316. This indicates a commensalistic interaction between these species, in which B316 initiates the primary attack on the insoluble substrate, while MB2003 has a secondary role, competing for the released soluble sugars. This work provides the first systematic phenotypic, comparative genomic and functional analysis of ruminal *Butyrivibrio* and *Pseudobutyrivibrio* species, which not only defines their conserved features involved in hemicellulose degradation, but is also beginning to differentiate their unique gene complements and growth characteristics that separate them as discrete species.

Acknowledgements

To my supervisors Dr Graeme Attwood, Dr Bill Kelly and Dr Jasna Rakonjac, thank you for taking me on first as a Masters and Doctorate student. Thank you for your encouragement, patience, wisdom and support throughout this journey and hope that you have enjoyed teaching me as I have enjoyed learning from all of you.

I wish to gratefully acknowledge AgResearch and the New Economy Research Fund (NERF) for funding this project and the research facilities that have allowed me to complete this study. I wish to thank Massey University and the Institute of Fundamental Sciences (IFS) for financial support. I would like to express my gratitude to the NZ Microbiological Society, also the IFS and NZ Society for Biochemistry and Molecular Biology for financial assistance to present this thesis work at the Australian Society for Microbiology annual meeting in 2015.

Thank you to the Rumen Microbiology team for all the assistance and support. I would like to extend my sincere gratitude to the following people for all their help: Dr Sinead Leahy, Dr Eric Altermann and Roger Moraga for your guidance with the bioinformatics and all things computational. Dr Peter Janssen, Dr Gemma Henderson, Dr Adrian Cookson and Dr Christina Moon for providing input and feedback throughout this project. Sarah Lewis and Bryan Treloar for help with VFA analysis. Dr Don Otter and Cornelis van Dijk for help with HPIC analysis. Siva Ganesh for guidance and support with the statistical analyses of the RNA-seq data. Joy Dick for help with referencing. I also wish to thank Dr Stuart Denman for collaborative efforts with the CAZyme analyses.

I have been blessed with a very loving and supportive family, for this I am the most grateful and thankful. My mum, Desanka Palevich, has always stressed the importance of education and I know that this respect for education has, in some unconscious way, shaped my values and made me the person that I am today. I am forever grateful to you for all of the sacrifices that you have made for me. My baba, Ruzica Radulovic, has been a source of constant and unconditional love for as long as I can remember. You have taught me about what it means to be both a “good person” and the “bigger person”, figuratively speaking. To my mother and father-in-law, Marilyn and Tom Cox, thank you for all the love and support over the past few years.

Finally, to my beautiful and loving wife Faith Palevich. You have encouraged me from the beginning and your support in harsh times is what got me to the end. For this you have my deepest gratitude.

Dedication

To the most beautiful woman in the world, Faith Palevich, I find it difficult to express into words my appreciation because it is so boundless. You are my best friend, my most enthusiastic cheerleader and an amazing wife. It has been an absolute privilege to work along side you over these past few years and you have undoubtedly matured me into the person I am today. You continually inspire me to work hard in all aspects of the game and to stand up for myself. I sincerely appreciate all of the sacrifices that you have made to make my career a priority in our lives. Without your love and support, I would be lost. You have shared this entire amazing journey with me and seen me through the ups and downs. So without question in my mind, I dedicate this dissertation to you.

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **iii**

Dedication ..... **v**

Table of Contents ..... **vii**

List of Figures ..... **xi**

List of Tables ..... **xvii**

Non-Standard Abbreviations ..... **xx**

Chapter 1 Introduction and Literature Review **1**

1.1 Introduction 2

1.2 The ruminant digestive system 4

1.3 Forage composition 6

1.4 Plant fibre degradation 12

1.5 Rumen microbial diversity and the fibre degrading rumen microbiome 18

1.6 Diversity of rumen *Butyrivibrio* and *Pseudobutyrivibrio* 22

1.7 *Butyrivibrio proteoclasticus* B316 25

1.8 Thesis aims 30

Chapter 2 Materials and Methods **31**

2.1 Materials 32

2.1.1 Bacterial strains 32

2.1.2 Buffers, reagents and solutions 32

2.1.3 Media components 37

2.1.4 Media 40

2.1.5 Enzymes, buffers and reagents 42

2.1.6 General laboratory equipment 44

2.1.7 Bioinformatic resources and software 44

2.2 Methods 47

2.2.1 Origin and selection of *Butyrivibrio* isolates 47

2.2.2 Bacterial culture revival and growth 47

2.2.3 Wet mounts and Gram staining 48

2.2.4 Long-term culture storage 48

2.2.5 Agarose gel-electrophoresis 48

2.2.6 DNA extractions 49
2.2.7 Nucleic acid extractions 51
2.2.8 Nucleic acid quantity and quality analysis 53
2.2.9 Sanger sequencing 54
2.2.10 Polymerase chain reaction (PCR) 55
2.2.11 Phylogenetic analysis of full-length 16*S* rRNA gene sequences 56
2.2.12 PCR amplification of enolase genes 58
2.2.13 Cloning *Butyrivibrio* marker genes 58
2.2.14 qPCR 60
2.2.15 Carbon source utilisation 62
2.2.16 Fermentation end product analysis 62
2.2.17 High-pressure ion chromatography (HPIC) 64
2.2.18 Electron microscopy 65
2.2.19 Cell motility assay 66
2.2.20 Pulsed-field gel electrophoresis (PFGE) 66
2.2.21 Whole-Genome Sequencing (WGS) 67
2.2.22 *Butyrivibrio* co-culture growth experiment 68
2.2.23 *In silico* analyses 70
2.2.23.1 *Butyrivibrio hungatei* MB2003 genome project 70
2.2.23.2 Comparative analysis of the pan-genome datasets 73
2.2.23.3 RNA sequencing analysis 78

Chapter 3 Selection and characterisation of candidate *Butyrivibrio* strains 85
3.1 Introduction 86
3.2 Selection and taxonomic assignment of *Butyrivibrio* strains 87
3.3 Morphological characterisation 90
3.4 Genome size estimation and *ApaI* restriction fragment length polymorphism (RFLP) assessed by PFGE 96
3.5 Cell motility 101
3.6 Carbon source utilisation by *Butyrivibrio* strains 106
3.7 Analysis of fermentation end products after growth on cellobiose 112
3.8 Discussion 114

Chapter 4 *Butyrivibrio* and *Pseudobutyrivibrio* pan-genome analyses..... 119
4.1 Introduction..... 120
4.2 Genomic features of *Butyrivibrio* and *Pseudobutyrivibrio* 122
4.3 Genome-based reconstruction of phylogeny 130
4.4 Synteny, amino acid and codon usage comparison 135
4.5 Pan-genome analysis 136
4.6 Glycobiomes of *Butyrivibrio* and *Pseudobutyrivibrio* 139
4.6.1 Carbohydrate-Binding Modules (CBMs) 143
4.6.2 Glycoside Hydrolases (GHs) 147
4.6.3 Carbohydrate Esterases (CEs) 160
4.6.4 Polysaccharide Lyases (PLs) 161
4.6.5 Glycosyl Transferases (GTs) 169
4.7 Genotypic variation associated with selected phenotypes 173
4.7.1 Cell motility and flagellum-encoding loci 173
4.7.2 Lactate production 180
4.7.3 Enolase genes 184
4.8 Discussion 189
4.8.1 Proteins and domains involved in xylan and pectin breakdown 190
4.8.2 Enolase and the glycolytic pathway 196
4.8.3 Flagellar operons and motility 198

Chapter 5 Xylan and pectin utilisation by *B. hungatei* MB2003 and *B. proteoclasticus* B316 201
5.1 Introduction..... 202
5.2 *Butyrivibrio hungatei* MB2003 202
5.2.1 *B. hungatei* MB2003 genome properties and comparison to *B. proteoclasticus* B316..... 203
5.3 Mono- and co-culture growth experiments..... 206
5.3.1 Quantitative PCR assays..... 206
5.3.2 *Butyrivibrio* co-culture growth on xylan and pectin..... 206
5.4 Transcriptome profiling of mono- and co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan or pectin..... 214
5.4.1 Global transcriptome dynamics 215
5.4.2 Transcriptome assembly and differential gene expression..... 223

5.4.3 Biological processes associated with differentially expressed genes 230
5.5 Carbohydrate degradation and metabolism 233
5.6 Oligosaccharide transport and assimilation 239
5.7 Discussion 243

Chapter 6 General discussion, conclusions and future directions 247
6.1 Introduction 248
6.2 Differentiation of the *Butyrivibrio* and *Pseudobutyrivibrio* genera 248
6.3 *Butyrivibrio* and *Pseudobutyrivibrio* genome features 250
6.4 Cell motility and operons encoding flagella genes 252
6.5 Plant fibre degradation by *Butyrivibrio* and *Pseudobutyrivibrio* 253
6.6 Proposed models for degradation of xylan and pectin by *Butyrivibrio* 258
6.6.1 Proposed model for GAX degradation 260
6.6.1.1 Extracellular GAX degradation 260
6.6.1.2 Transport of xylo-oligosaccharides and monosaccharides 264
6.6.1.3 Cytosolic degradation of xylo-oligosaccharides 266
6.6.2 Proposed model for XGA and RG-I degradation 267
6.6.2.1 Extracellular XGA and RG-I degradation 267
6.6.2.2 Transport of pectic-oligosaccharides and monosaccharides 271
6.6.2.3 Cytosolic degradation of pectic-oligosaccharides 272
6.7 Conclusions 273
6.8 Future perspectives 275

Appendices 279
Appendix I 280
Appendix II 287
Appendix III 311

References 387

List of Figures

Figure 1.1. Simplified diagram of the ruminant digestive tract. 5
Figure 1.2. Generalised structure of the primary plant cell wall 8
Figure 1.3. Simplified chemical structures of cellulose, hemicellulose and pectin 11
Figure 1.4. Enzymes involved in plant fibre-degradation 16
Figure 1.5. Bacterial diversity of the rumen microbial ecosystem 18
Figure 1.6. Phylogeny of Clostridial Cluster I and subcluster XIVa strains based on near full-length 16*S* rRNA gene sequences 26
Figure 1.7. Ultrastructure of *B. proteoclasticus* B316 cells 27
Figure 1.8. Genome atlas of *Butyrivibrio proteoclasticus* B316 29
Figure 2.1. Preparation pipeline used for prokaryotic strand-specific transcriptome analysis 54
Figure 2.2. Overview of *B. hungatei* MB2003 and *B. proteoclasticus* B316 co-culture growth experiment 69
Figure 2.3. Equations for calculation of ANI and AF 73
Figure 2.4. Overview of the workflow for RNA-seq *in silico* analysis 79
Figure 3.1. Phylogenetic tree of *Butyrivibrio* strains based on 16*S* rRNA full-length gene sequence data 89
Figure 3.2. Light micrographs of Gram-stained *B. fibrisolvens* strains grown on RM02 media containing cellobiose 93
Figure 3.3. Light micrographs of Gram-stained *Butyrivibrio* sp. 2 strains grown on RM02 media containing cellobiose 94
Figure 3.4. Light micrographs of Gram-stained *Butyrivibrio* sp. 3 strains grown on RM02 media containing cellobiose 95
Figure 3.5. PFGE profiles of uncut *B. fibrisolvens* genomic DNAs 98
Figure 3.6. PFGE profiles of uncut *Butyrivibrio* sp. 2 genomic DNAs 99
Figure 3.7. PFGE profiles of uncut *Butyrivibrio* sp. 3 genomic DNAs 100
Figure 3.8. Motility assays of control strains 101
Figure 3.9. Motility assays of *B. fibrisolvens* strains 102
Figure 3.10. Motility assays of *Butyrivibrio* sp. 2 strains 103
Figure 3.11. Motility assays of *Butyrivibrio* sp. 3 strains 104
Figure 3.12. End product analysis of *Butyrivibrio* strains grown on insoluble carbon sources 110

Figure 4.1. Pfam analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes 127
Figure 4.2. Prokaryotic COGs analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes..... 129
Figure 4.3. FGD of *Butyrivibrio* and *Pseudobutyrivibrio* genomes 131
Figure 4.4. Genome clustering comparison of Pfam and COG domains from *Butyrivibrio* and *Pseudobutyrivibrio*..... 133
Figure 4.5. ANI and AF analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes..... 134
Figure 4.6. Core- and pan-genomes of *Butyrivibrio* and *Pseudobutyrivibrio* defined using BLAST analysis..... 137
Figure 4.7. Flowerplot diagram of unique, group-specific and core gene families in the *Butyrivibrio* and *Pseudobutyrivibrio* genomes 138
Figure 4.8. CAZyme composition of *Butyrivibrio* and *Pseudobutyrivibrio* genomes and cluster groups 140
Figure 4.9. Comparative analysis of annotated *Butyrivibrio* and *Pseudobutyrivibrio* CAZymes 141
Figure 4.10. The distribution of each CAZyme class and family in *Butyrivibrio* and *Pseudobutyrivibrio* genomes..... 142
Figure 4.11. Heatmap of normalised relative abundances for CAZyme families determined for *Butyrivibrio* and *Pseudobutyrivibrio* genomes..... 145
Figure 4.12. Heatmap showing the abundances of CBM families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes..... 146
Figure 4.13. Heatmap showing the abundances of GH families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes..... 148
Figure 4.14. Predicted Pfam domains of GH5 endoglucanases..... 153
Figure 4.15. Predicted Pfam domains of GH51 α-L-arabinofuranosidases..... 155
Figure 4.16. Predicted Pfam domains of GH10 xylanases 159
Figure 4.17. Predicted Pfam domains of GH11 endo-xylanases 159
Figure 4.18. Heatmap showing the abundances of CE and PL families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes 162
Figure 4.19. Predicted Pfam domains of CE1 acetyl-xylan and ferulic acid esterases..... 165
Figure 4.20. Predicted Pfam domains of CE4 polysaccharide deacetylases..... 168

Figure 4.21. Heatmap showing the abundances of GT families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes 170
Figure 4.22. Predicted Pfam domains of GTs 172
Figure 4.23. Motility assays and flagellar biosynthesis operons of control strains 174
Figure 4.24. Motility assays and flagellar biosynthesis operons of *B. fibrisolvens* strains 175
Figure 4.25. Motility assays and flagellar biosynthesis operons of *Butyrivibrio* sp. 2 strains 177
Figure 4.26. Motility assays and flagellar biosynthesis operons of *Butyrivibrio* sp. 3 strains 179
Figure 4.27. Comparison of *ldh* Pfam domains in *Butyrivibrio* genome sequences 182
Figure 4.28. Phylogenetic tree of *ldhs* encoded in *Butyrivibrio* genomes 184
Figure 4.29. Enolases from *Butyrivibrio* genome sequences 186
Figure 4.30. Overview of possible carbohydrate metabolic pathways in *Butyrivibrio* leading to the formation of butyrate, formate, acetate and lactate from cellobiose 197
Figure 5.1. Transmission electron micrograph of a negatively stained *B. hungatei* MB2003 cell 203
Figure 5.2. Genome atlas of *B. hungatei* MB2003 205
Figure 5.3. qPCR of *B. proteoclasticus* B316 and *B. hungatei* MB2003 co-cultures grown on xylan and pectin 207
Figure 5.4. Mono- and co-culture growth of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin as determined by qPCR 208
Figure 5.5. pH measurements of mono- and co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin 210
Figure 5.6. Monosaccharides released by *B. hungatei* MB2003 and *B. proteoclasticus* B316 in mono- and co-cultures grown on xylan 212
Figure 5.7. Monosaccharides released by *B. hungatei* MB2003 and *B. proteoclasticus* B316 in mono- and co-cultures grown on pectin 213
Figure 5.8. NMDS ordination showing transcriptional similarity between *B. proteoclasticus* B316 samples 216
Figure 5.9. NMDS ordination showing transcriptional similarity between *B. hungatei* MB2003 samples 217

Figure 5.10. Permutation-based test of multivariate homogeneity of group dispersions and variances 219
Figure 5.11. Network analysis of the subset *B. proteoclasticus* B316 dataset 221
Figure 5.12. CIM analysis of genes significantly up-regulated in *B. proteoclasticus* B316 227
Figure 5.13. CIM analysis of genes significantly up-regulated in *B. hungatei* MB2003 228
Figure 5.14. Network analysis of genes significantly up-regulated in *B. proteoclasticus* B316 229
Figure 5.15. COG classifications of DEGs from *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown in mono- and co-culture on xylan..... 231
Figure 5.16. COG classifications of DEGs from *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown in mono- and co-culture on pectin..... 232
Figure 5.17. CAZyme encoding DEGs up-regulated in xylan-grown cultures 234
Figure 5.18. CAZyme encoding DEGs up-regulated in pectin-grown cultures..... 236
Figure 5.19. GTs up-regulated during mono-culture growth on pectin 238
Figure 5.20. Functional domains of DEGs encoding carbohydrate transport proteins and surrounding CAZymes identified in xylan- or pectin-grown cultures 242
Figure 6.1. Schematic diagrams of the structures of the main classes of xylan and pectin 259
Figure 6.2. Model for the degradation of GAX by *B. proteoclasticus* B316 and *B. hungatei* MB2003 262
Figure 6.3. Model for the degradation of XGA and RG-I by *B. proteoclasticus* B316 and *B. hungatei* MB2003 269
Figure A1.1. PFGE *Apa*I restriction profiles of *B. fibrisolvens* genomic DNAs 280
Figure A1.2. PFGE *Apa*I restriction profiles of *Butyrivibrio* sp. 2 genomic DNAs 281
Figure A1.3. PFGE *Apa*I restriction profiles of *Butyrivibrio* sp. 3 genomic DNAs 282
Figure A1.4. PFGE I-*Ceu*I restriction profiles of *B. fibrisolvens* genomic DNAs 283
Figure A1.5. PFGE I-*Ceu*I restriction profiles of *Butyrivibrio* sp. 2 genomic DNAs 284
Figure A1.6. PFGE I-*Ceu*I restriction profiles of *Butyrivibrio* sp. 3 genomic DNAs 285
Figure A2.1. TIGRfam analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes 291

Figure A2.2. KEGG pathway analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes 293
Figure A2.3. Genome clustering comparison of TIGRfam and KO domains from *Butyrivibrio* and *Pseudobutyrivibrio* 295
Figure A2.4. MUMmer plots of *B. fibrisolvens* draft genomes versus *B. proteoclasticus* B316 as the reference genome 296
Figure A2.5. MUMmer plots of *Butyrivibrio* sp. 2 draft genomes versus *B. proteoclasticus* B316 as the reference genome 297
Figure A2.6. MUMmer plots of *Butyrivibrio* sp. 3 draft genomes versus *B. proteoclasticus* B316 as the reference genome 298
Figure A2.7. Amino acid usage heatmap of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* based on their protein content 299
Figure A2.8. Codon usage heatmap of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* based on their protein content 300
Figure A2.9. Predicted Pfam domains of GH67 α-glucuronidases 308
Figure A2.10. Predicted Pfam domain of GH8 oligosaccharide reducing-end xylanases 309
Figure A2.11. PCR amplification of enolase genes from *Butyrivibrio* and *Pseudobutyrivibrio* strains 310
Figure A3.1. PCR amplification of *Butyrivibrio* strain DNAs using *Taq*Man primer/probe assay oligonucleotides 318
Figure A3.2. Standard curves derived from simplex real time PCR reactions 319
Figure A3.3. Distribution of all genes in *B. proteoclasticus* B316 based on *Q-value* from Rockhopper transcriptome assembly 334
Figure A3.4. Distribution of all genes in *B. hungatei* MB2003 based on *Q-value* from Rockhopper transcriptome assembly 335
Figure A3.5. CA plots showing transcriptional homogeneity in *B. proteoclasticus* B316 at different treatment groups 336
Figure A3.6. CA plots showing transcriptional homogeneity in *B. hungatei* MB2003 at different treatment groups 337
Figure A3.7. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. proteoclasticus* B316 338
Figure A3.8. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. proteoclasticus* B316 at $Q < 0.05$ 339

Figure A3.9. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. hungatei* MB2003 340
Figure A3.10. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. hungatei* MB2003 at $Q < 0.05$ 341
Figure A3.11. CIM analysis of the entire *B. proteoclasticus* B316 dataset 347
Figure A3.12. CIM analysis of the entire *B. hungatei* MB2003 dataset 348
Figure A3.13. CIM analysis of the subset *B. proteoclasticus* B316 dataset 349
Figure A3.14. CIM analysis of the subset *B. hungatei* MB2003 dataset 350
Figure A3.15. Ordinate plots comparing BH-adjusted ANOVA and BH-adjusted KW analyses 351

List of Tables

Table 1.1. Ruminant livestock numbers and product export value in New Zealand 2
Table 1.2. Composition of Type-I and Type-II primary and secondary plant cell walls 7
Table 1.3. Important fibre degrading enzymes involved in forage degradation in the rumen 13
Table 1.4. The major fibre degrading bacteria of the bovine rumen 20
Table 1.5. *B. proteoclasticus* B316 genome features and genes involved in polysaccharide breakdown 27
Table 2.1. Bacterial strains used in this thesis 33
Table 2.2. Anaerobic glycerol solution 32
Table 2.3. GenRFV 39
Table 2.4. Trace element solution 39
Table 2.5. Vitamin solution 40
Table 2.6. BY and RM02 media 41
Table 2.7. 16*S* rRNA gene primers 43
Table 2.8. *Taq*Man primers/probes 43
Table 2.9. Centrifuge specifications and suppliers 45
Table 2.10. Molecular weight size markers 45
Table 2.11. Bioinformatic resources and software used in this thesis 46
Table 2.12. Sampling times and isolate nomenclature 47
Table 2.13. Assembly of PCR reagents 55
Table 2.14. Thermal profile for PCR 56
Table 2.15. Assembly of reagents for LR-PCR 56
Table 2.16. Thermal profile for LR-PCR 56
Table 2.17. Reagents for PCR cloning 59
Table 2.18. Thermal profile of PCR for clone library 59
Table 2.19. Colony PCR reagents 60
Table 2.20. Thermal profile of the colony PCR reaction 60
Table 2.21. qPCR reagents 61
Table 2.22. qPCR thermal profile 62
Table 3.1. Taxonomic assignments and 16*S* rRNA gene sequence analyses of *Butyrivibrio* strains 88

Table 3.2. Growth and morphological characterisation of the *Butyrivibrio* strains 91
Table 3.3. Colony morphology of the *Butyrivibrio* strains 92
Table 3.4. Motility assay of *Butyrivibrio* strains 105
Table 3.5. Carbon source utilisation of the *Butyrivibrio* strains 107
Table 3.6. End product analysis of *Butyrivibrio* strains 113
Table 4.1. Additional strains for pan-genome analyses 121
Table 4.2. General features of *Butyrivibrio* and *Pseudobutyrivibrio* genomes 123
Table 4.3. Group specific CAZyme families 144
Table 4.4. Enolase gene occurrence in *Butyrivibrio* and *Pseudobutyrivibrio* genomes 185
Table 4.5. Summary of genes involved in the Methylglyoxyl Shunt pathway 188
Table 5.1. VFA analysis of mono- and co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin 211
Table 5.2. List of *B. proteoclasticus* B316 subset genes from Network analysis 222
Table 5.3. Transcriptome assembly and differential gene expression analysis for growth on xylan 224
Table 5.4. Transcriptome assembly and differential gene expression analysis for growth on pectin 225
Table A1.1. Summary of the genomic properties based on PFGE 286
Table A2.1. Genome annotation features of *Butyrivibrio* and *Pseudobutyrivibrio* genomes 287
Table A2.2. CAZy families analysed and their known activities 301
Table A3.1. Classification and general features of *B. hungatei* MB2003 311
Table A3.2. Project information 312
Table A3.3. Genome statistics of *B. hungatei* MB2003 312
Table A3.4. Assignment of MB2003 protein coding genes to COG functional categories 313
Table A3.5. Comparison of MB2003 and B316 protein coding gene percentages to COG functional categories 314
Table A3.6. List of top 20 candidate gene targets for the universal mono- and co-culture qPCR assays based on DBA 315
Table A3.7. List of top 20 candidate gene targets for the *B. proteoclasticus* B316 specific mono- and co-culture qPCR assays based on DBA 316

Table A3.8. List of top 20 candidate gene targets for the *B. hungatei* MB2003 specific mono- and co-culture qPCR assays based on DBA 317
Table A3.9. Quality of raw reads and FastQC results of trimmed reads 320
Table A3.10. Alignment details of trimmed reads to the *B. proteoclasticus* B316 genome 322
Table A3.11. Alignment details of trimmed reads to the *B. hungatei* MB2003 genome 324
Table A3.12. Transcript abundance for *B. proteoclasticus* B316 grown in mono- and co-culture on xylan and pectin 326
Table A3.13. Transcript abundance for *B. hungatei* MB2003 grown in mono- and co-culture on xylan and pectin 330
Table A3.14. List of *B. proteoclasticus* B316 subset genes from the network analysis 342
Table A3.15. List of genes significantly up-regulated during mono- and co-culture growth of *B. proteoclasticus* B316 on xylan 352
Table A3.16. List of genes significantly up-regulated during mono- and co-culture growth of *B. hungatei* MB2003 on xylan 353
Table A3.17. List of genes significantly up-regulated during mono- and co-culture growth of *B. proteoclasticus* B316 on pectin 362
Table A3.18. List of genes significantly up-regulated during mono- and co-culture growth of *B. hungatei* MB2003 on pectin 364

Non-Standard Abbreviations

AA	Auxiliary Activities
aa	Amino acid(s)
ABC	ATP-binding cassette
ACS	American Chemical Society
AF	Alignment fraction
ANI	Average nucleotide identity
ANOVA	One-way analysis of variance
BCVFA	Branched chain volatile fatty acids
BLAST	Basic Local Alignment Sequence Tool
bp	Base pair(s)
CA	Correspondence analysis
CAZY	Carbohydrate-Active enZYmes
CBM	Carbohydrate-Binding Module(s)
CBP	Carbohydrate Binding Protein(s)
cDNA	Complementary DNA
CDS	Coding Sequences
CE	Carbohydrate Esterase(s)
CIM	Clustered Image Map
COG	Clusters of Orthologous Groups of proteins
CUT	Carbohydrate Uptake Transporter
DBA	Differential BLAST analysis
dbCAN	DataBase for automated Carbohydrate-active enzyme ANnotation
DEG	Differentially expressed genes
DEPC H_2O	Diethylpyrocarbonate-treated water
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
DOE-JGI	United States Department of Energy's Joint Genome Institute
EC	enzyme Comission
EDTA	Ethylenediamine tetraacetic acid
EPS	Extracellular Polysaccharides
ER-IMG	Expert Review version of the IMG system
FGD	Functional genome distribution

FDR	False discovery rate
g	Gram(s)
Gb	Gigabase(s)
GB	Gigabyte(s)
GH	Glycoside Hydrolase(s)
GO	Gene Ontology
GT	Glycosyl Transferase(s)
h	Hour(s)
HMM	Hidden Markov models
HPIC	High-pressure ion chromatography
Hz	Hertz
IMG	Integrated Microbial Genomes
Interpro	Intergrative Protein Signature Database
JGI	Joint Genome Institute
Kb	Kilobase(s)
KB	Kilobyte(s)
KEGG	Kyoto Encyclopedia of Genes and Genomes
kg	Kilogram(s)
KO	KEGG Orthology
KW	Kruskal-Wallis
L	Litre(s)
LB	Luria-Bertani
MANOVA	Multivariate analysis of variance
Mb	Megabase(s)
MB	Megabyte(s)
µL	Microlitre(s)
M	Molar
MDS	Multidimensional scaling
mg	Milligram(s)
min	Minute(s)
mL	Millilitre(s)
mm	Millimeter(s)
mM	Millimolar
mRNA	Messenger RNA

mV	Millivolt(s)
ncRNA	Non-coding RNA
NDF	Neutral Detergent Fibre
NMDS	Non-metric multidimensional scaling
ng	Nanogram(s)
nm	Nanometer(s)
nt	Nucleotide(s)
OD	Optical density
ORF	Open reading frame(s)
OUT	Operational taxonomic unit(s)
PCoA	Principal component analysis
PCR	Polymerase chain reaction
Pfam	Protein families
PFGE	Pulsed-field gel electrophoresis
PL	Polysaccharide Lyase(s)
PP	Permease protein
PRIAM	PRofils pour l'Identification Automatique du Métabolisme
qPCR	Quantitative polymerase chain reaction
RE	Restriction endonuclease
RFLP	Restriction fragment length polymorphism
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RPKM	Reads per kilobase of gene per million reads mapped
SBP	Substrate-binding protein
SCVFA	Short chain volatile fatty acids
sec	Second(s)
SEM/TEM	Scanning/Transmission electron microscopy
SLH	S-layer homology
TE	Tris-EDTA buffer
TIGRfam	The Institute for Genomic Research's database of protein families
tRNA	Transfer RNA
UNG	Uracil N-Glycosylase
UPGMA	Unweighted Pair Group Method with Arithmetic means
V	Volts

v/v	Volume per volume
v/v/v	Volume per volume per volume
VFA	Volatile fatty acid(s)
w/v	Weight per volume

Chapter 1

Introduction and Literature Review

1.1 Introduction

Ruminants are considered the dominant terrestrial herbivores, and there is scientific interest in this group due to their diversity in species, wide geographic distribution across various climatic zones, and contribution to human nutrition (Haenlein, 2001). The earliest fossil records of ruminants date back 50 million years and indicate that ruminants have been found on all continents, except Antarctica (Vrba and Schaller, 2000). Ruminant animals belong to the order Artiodactyla within the suborder Ruminantia and are represented by 197 extant species, separated into six families; Tragulidae, Giraffidae, Antilocapridae, Moschidae, Cervidae, and Bovidae (Fernández and Vrba, 2005). The majority of ruminant species belong to the Bovidae and Cervidae families (Hackmann and Spain, 2010). The Bovidae family includes many of the well-known grazing animals such as cattle, sheep, goats, water buffaloes, oxen and yaks, while deer belong to the Cervidae family.

Cattle and sheep are the main ruminant species farmed in New Zealand (NZ, Table 1.1). In recent times dairy cattle numbers have increased and dairy products have become the major earner of NZ's export revenue. The value to the NZ economy of products derived from ruminant animals (including dairy, meat and wool) represented approximately $NZ 29 billion in 2015, and this accounts for more than 50% of NZ's total commodity exports. With the steady increase in the world's human population, the demand for meat and milk products is expected to increase (Alexandratos and Bruinsma, 2012; Tilman and Clark, 2014).

Table 1.1. Ruminant livestock numbers and product export value in New Zealand.

Ruminant species	Total stock numbers	Export value ($M)
Sheep	29,803,402	4,365.3
Cattle [A]	10,368,188	23,464.5
Deer	958,219	253.6
Goats	97,370	6.4
Total	**41,227,179**	**28,671.0** [B]

[A], Total cattle numbers include beef and dairy cattle (also including bobby calves). [B], Total value of principal exports also includes live animals, edible offal and other miscellaneous items. Export value is reported in $ million NZD ($M) and comprises of meat, dairy and hide products. All data is acquired from Statistics New Zealand (Statistics New Zealand, 2015).

In NZ ruminant livestock are predominantly fed pasture outdoors, as opposed to other parts of the world, where animals are housed for part of the year and fed with grains or other high energy feeds. More than half of the NZ landmass is used for farming and the national herd of dairy and beef cattle, combined with sheep, consumes more than a million tonnes of fresh pasture each day (Statistics New Zealand, 2015). The pasture consumed is predominantly a mixture of ryegrass and clover species that have a high fibre content (Holden *et al.*, 1994; Kusmartono *et al.*, 1997; Waghorn and Clark, 2004). Ruminants do not possess the enzymatic machinery required for degradation of fibrous plant material, hence they rely on the action of microbes in their reticulo-rumens to carry out this process. At the same time, the reticulo-rumen provides a fermentative incubator for a thriving microbial community. This symbiotic relationship between the host animal and rumen microbial community provides the microorganisms with an anaerobic environment and a semi-continuous supply of growth substrates. The anatomical and physiological adaptations of the ruminant digestive system enable them to benefit by using the microbial fermentation end products and converting an otherwise unusable food source (mainly in the form of cellulose and other recalcitrant carbohydrates such as xylan and pectin) into ruminant products (Van Soest, 1994).

Ruminant productivity is limited by the digestibility of the lignocellulosic component of plant fibre and the amount of plant material that can be eaten (Villas-Bôas *et al.*, 2002). There are four main factors that restrict lignocellulose degradation in the rumen (Cheng *et al.*, 1991); (i) host factors that mediate the availability of nutrients through mastication, salivation and digesta kinetics, (ii) the chemical composition and degradability of the insoluble component of the plant material, (iii) the type of feed and diet of the animal and (iv) the composition of the rumen microbial ecosystem and their polysaccharolytic capability.

Microbes initiate the plant fibre degradation process by adhering to plant material and producing both cell surface-associated and secreted fibrolytic enzymes. Examples of polysaccharide degrading enzymes (or polysaccharidases) include cellulases (predominantly endo- or exo-glucanases and β-glucosidases), hemicellulases (including xylanases, β-xylosidases, α-glucuronidases, α-L-arabinofuranosidases and acetyl xylan esterases), and pectate lyases. Polysaccharidases involved in forage degradation are detailed in Section 1.4 of this chapter. These enzymes have generated significant interest due to their potential applications in agriculture, such as the use of cellulases and xylanases as feed additives (Beauchemin *et al.*, 2003), and studies into their activities

have contributed, to our understanding of the fundamental mechanisms in the plant fibre degradation processes occurring in the rumen (Beg *et al.*, 2001; Subramaniyan and Prema, 2002; Kuhad *et al.*, 2011; Juturu and Wu, 2012, 2013). The rumen microbes that produce these enzymes also have potential use as probiotics for ruminants, in depolymerisation of lignocellulosic material for biogas and biofuel production (Shallom and Shoham, 2003; Howard *et al.*, 2004). Genetic engineering and directed evolution have been used to manipulate rumen microbes by targeting specific enzymatic applications, in order to lower production costs and improve enzymatic performance (Himmel *et al.*, 2010; Chandel and Singh, 2011; Kuhad *et al.*, 2011; Vazana *et al.*, 2012).

1.2 The ruminant digestive system

Rumen microbiology studies to date have mainly focused on cattle and sheep which are agriculturally important and make up a major proportion of the global ruminant population. The rapid evolution of ruminants can be attributed to their unique digestive system that integrates a symbiotic microbial population (Clauss *et al.*, 2010). Their unique digestive physiology and arrangement allows for microbial foregut fermentation to occur, a strategy that is thought to have given ruminants an evolutionary advantage over non-ruminants.

The digestion of plant material by ruminants is a complex process that occurs in several stages. Primary particle size reduction and fermentation occurs in the first two chambers of the four-compartment stomach, the rumen and the reticulum, by the combined action of chewing and fibre degradation by the resident microorganisms. Together the two are referred to as the "reticulo-rumen", or commonly just as the "rumen". The next compartment is the omasum which acts as a filter and site of water retention, followed by the abomasum which is the acid compartment of the stomach, and is similar to the stomach of monogastric animals in both structure and function (Hungate, 1975). Approximately 55-65% degradation of the plant material is accomplished by microbial digestion in the rumen and host enzymatic digestion in the abomasum, using polysaccharidases such as cellulases, xylanases and pectate lyases (Hungate, 1975; Waghorn *et al.*, 2007). The remaining 20-30% and 5-15% degradation occurs in the small and large intestine, respectively. A typical digestive tract of a cow is shown in Figure 1.1 from Carr (1999).

The rumen is a large organ that accounts for 12-25% of the ruminant live weight (Waghorn, 2002). The rumen is the largest of the fore-stomach compartments, accommodating 60-120 kg of digesta in cattle and 6-8 kg in sheep, with a daily digesta flow of 120-200 L in cattle and 10-30 L for sheep (Waghorn *et al.*, 2007). Rumen digesta is buffered by a large amount of saliva, 70-150 L/day in cows and 10-15 L/day in sheep, (Dado and Allen, 1995; Reynolds *et al.*, 2004; Kamra, 2005). The temperature is maintained at 39-40 ºC, except upon feeding when temperature rises marginally due to the heat generated from fermentation processes (Dale *et al.*, 1954). The rumen can be likened to an anaerobic, continuous-flow fermenter, which maintains an optimal pH between 5.8-6.8 with a gas phase consisting of ~0.5% O_2, ~65% CO_2 and ~35% CH_4 (Kolver and De Veth, 2002). Facultative anaerobes quickly metabolise the O_2 introduced by ingestion of plant material and the bicarbonate-based saliva buffers the acidic end-products of fermentation to maintain the pH (Williams *et al.*, 2001).

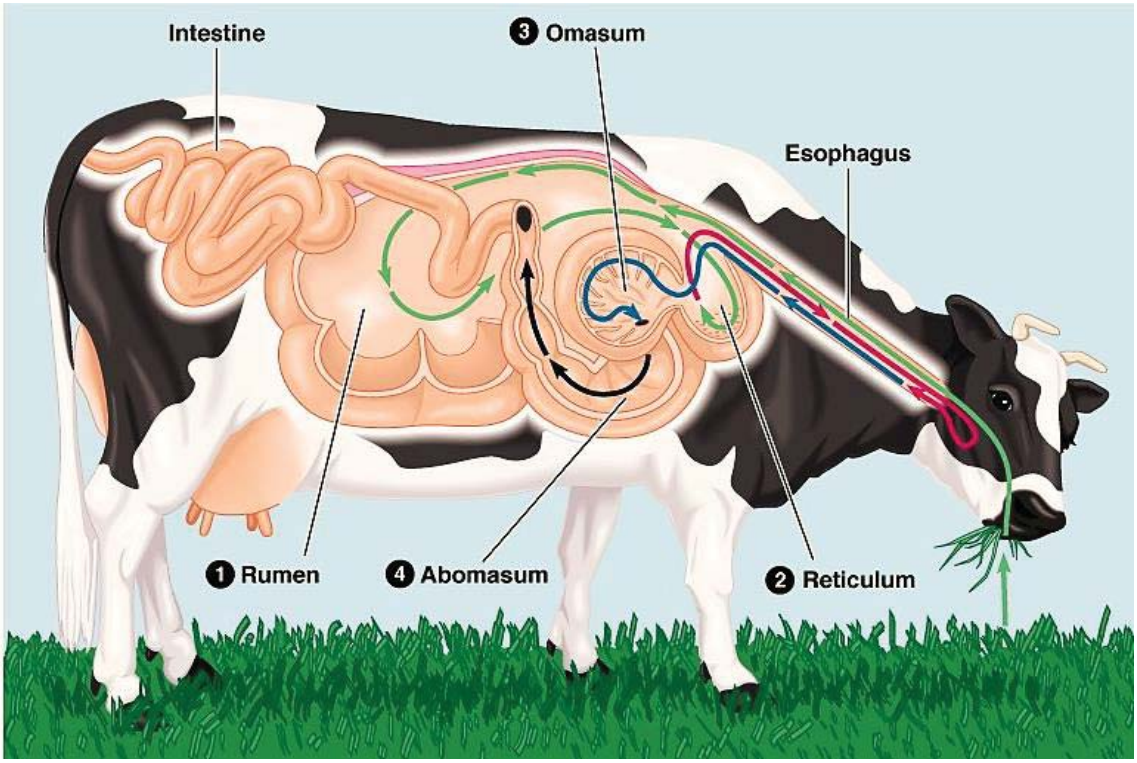


Figure 1.1. Simplified diagram of the ruminant digestive tract. The rumen is the largest compartment and primary site for microbially-mediated fermentation of ingested plant material. Digesta flows in a sequence from the rumen to the reticulum, omasum and abomasum. Figure adapted from Carr (1999) with permission.

1.3 Forage composition

Pastures in NZ are predominantly mixtures of perennial ryegrass (*Lolium perenne*) and variable amounts of white clover (*Trifolium repens*) (Woodfield, 1999; Waghorn and Clark, 2004). The chemical composition of pasture has been shown to alter the microbial composition in the rumen and affect pasture digestion efficiency; other factors such as season, climate, as well as the variety and developmental phase of the plant also affect the nutritional value of pasture (Waghorn *et al.*, 2007). The term 'plant fibre' is often used in ruminant nutrition to refer to the plant cell walls that provide the structural rigidity and tensile strength to the plant (Jung, 2011).

The cell wall constitutes approximately 80% of the plant material dry weight, and is composed of a complex and robust network of cellulose fibres (30-40%) surrounded by a matrix of non-cellulosic structural polysaccharides (20-40%) such as hemicellulose and pectin, and proteins (15-20%) (Cosgrove, 2005; Waghorn *et al.*, 2007). The remaining constituents include minerals (mainly potassium and calcium), lipids (waxes and cell membranes), phenolic compounds, lignin, nucleic acids, and secondary plant compounds (e.g. condensed tannins). Fibre dominates the plant material, hence its digestion is the major source of energy and driver of productivity in ruminant animals (Jung and Allen, 1995). Cellulose, hemicellulose and pectin are primarily degraded by rumen microbes adherent to the plant material, whereas the lignin component is not degraded to any great extent within the rumen (Balch *et al.*, 1954). The chemical composition of the cell wall varies greatly (Table 1.2), depending on the plant species, age and the tissue type.

Primary cell walls surround growing cells in young tissues, whereas expanded secondary cell walls confer mechanical strength after cell growth has ceased in mature tissues. The primary cell wall of flowering plants can be separated into two classes, which differ significantly in their structural features and chemical composition and contribute up to 20% of its dry weight (Wilkie, 1985; Rancour *et al.*, 2012). Type-I primary cell walls are made up of mainly cellulose microfibrils fixed in a matrix of xyloglucan (XyG)-rich hemicellulose, pectin, structural proteins, phenolic compounds and lignin, and are abundant in both dicotyledonous and noncommelinoid monocotyledonous plants as well as in gymnosperms. Type-II primary cell walls are found only in commelinoid mono-cotyledonous plants (i.e. grasses) and differ in that they are comprised of cellulose embedded in glucuronoarabinoxylans (GAX) with mixed linkage glucans (MLGs), and phenolic compounds. Type II primary cell walls differ from Type-I cell walls in that they

contain low levels of pectin and structural proteins (Carpita and Gibeaut, 1993; Carpita, 1996).

Secondary cell walls are a pronounced component of the plant fibre fractions and are arranged inside the primary cell walls (Grabber *et al.*, 2004). The secondary cell wall of grasses can also be separated into two classes, where both the Type-I and Type-II secondary cell walls are mainly composed of cellulose, GAX, and lignin, and contribute up to 50% of its dry weight (Brown and Saxena, 2000). As mentioned above, the two predominant plant types found in NZ pastures are perennial ryegrass and clover (Woodfield, 1999). Perennial ryegrass is composed of Type-II cell walls with high levels of GAX and ferulic acid crosslinks, whereas legumes (clover and lucerne) have Type-I cell walls that are high in pectin and xyloglucans (Varner and Lin, 1989). A schematic diagram of the primary plant cell wall is represented in Figure 1.2. Cellulose is composed of (1,4)-β-linked glucose units that spontaneously form insoluble microfibrils (approximately 3-5 nm wide) of single strands of up to 15,000 glucose monomers in length (Pizzi and Eaton, 1985). Individual microfibrils contain variable numbers of cellulose chains and may also associate with other microfibrils to form a larger single microfiber. The ability of cellulose to form strong bonds with a variety of non-cellulosic cell wall components and strong hydrogen bonding between cellulose molecules, accounts for the structural rigidity and tensile strength of the plant cell wall.

Table 1.2. Composition of Type-I and Type-II primary and secondary plant cell walls.

Cell wall component	**Primary wall** [A]		**Secondary wall** [A]	
	Type-I	**Type-II**	**Type-I**	**Type-II**
Cellulose	15-30 [B, C, D]	20-30 [B, E]	45-50 [B]	35-45 [B, F]
Hemicellulose				
Glucuronoarabinoxylans	5 [B]	20-40 [C]	20-30 [B, G]	40-50 [B, G]
Mixed linkage glucans	None	10-30 [C]	None	Minor
Xyloglucans	20-25	1-5 [B, C]	Minor	Minor
Mannans/glucomannans	5-10 [C]	Minor	3-5 [G]	Minor
Pectin	20-35 [C]	5 [B]	0.1 [B]	0.1 [B]
Structural proteins	10 [C, D]	1 [C]	Minor	Minor
Phenolic compounds	Minor	1-5 [B, C]	Minor	0.5-1.5 [B]
Lignin	Minor	Minor	7-10 [B]	20 [B]
Silica	None	None	Variable	5-15 [B]

[A], Values are the percentage of plant cell-wall dry weight measurements. The table was adapted from Vogel (2008) with permission. Estimates were obtained from a number of sources as detailed. [B] Ishii (1997), [C] O'Neill and York (2003), [D] Zablackis *et al.* (1995), [E] Mitchell *et al.* (2007), [F] Hatfield *et al.* (1999b) and [G] Ebringerová *et al.* (2005).

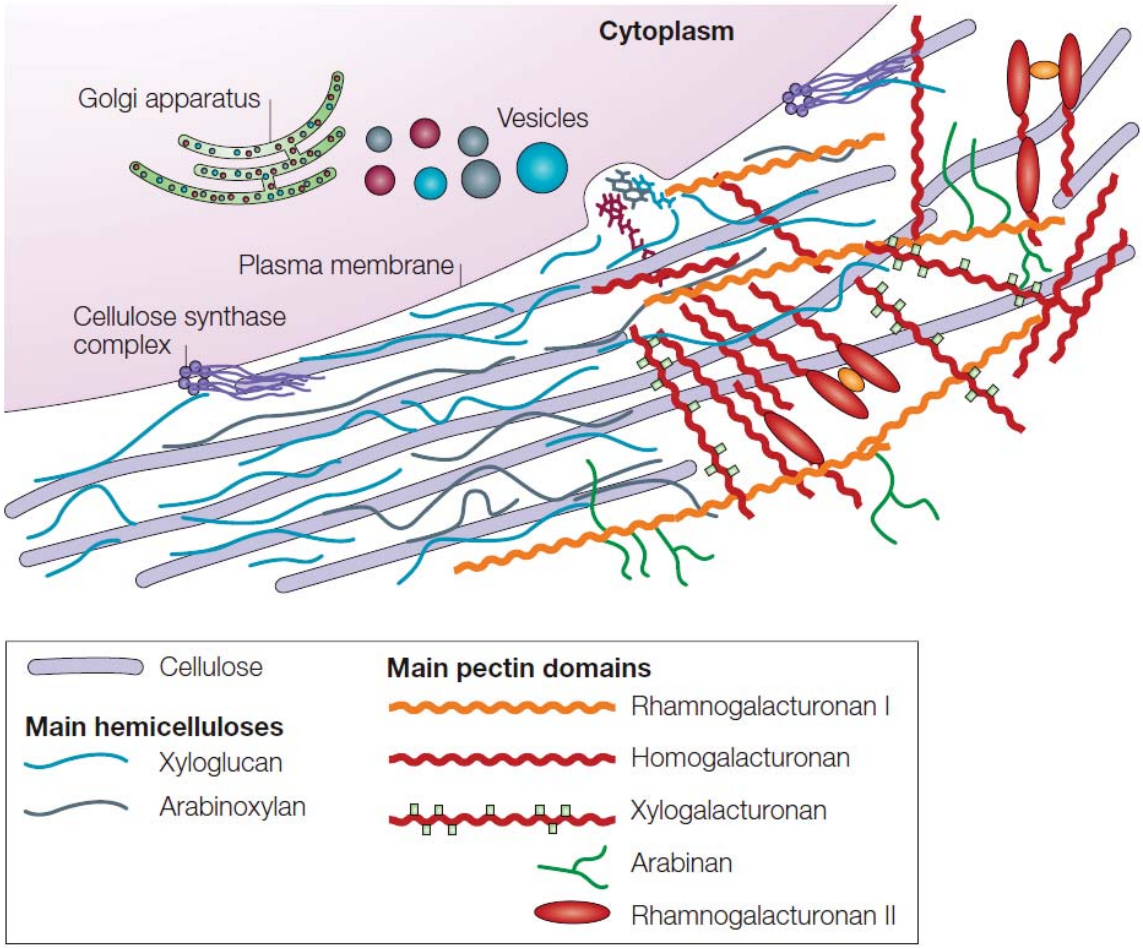


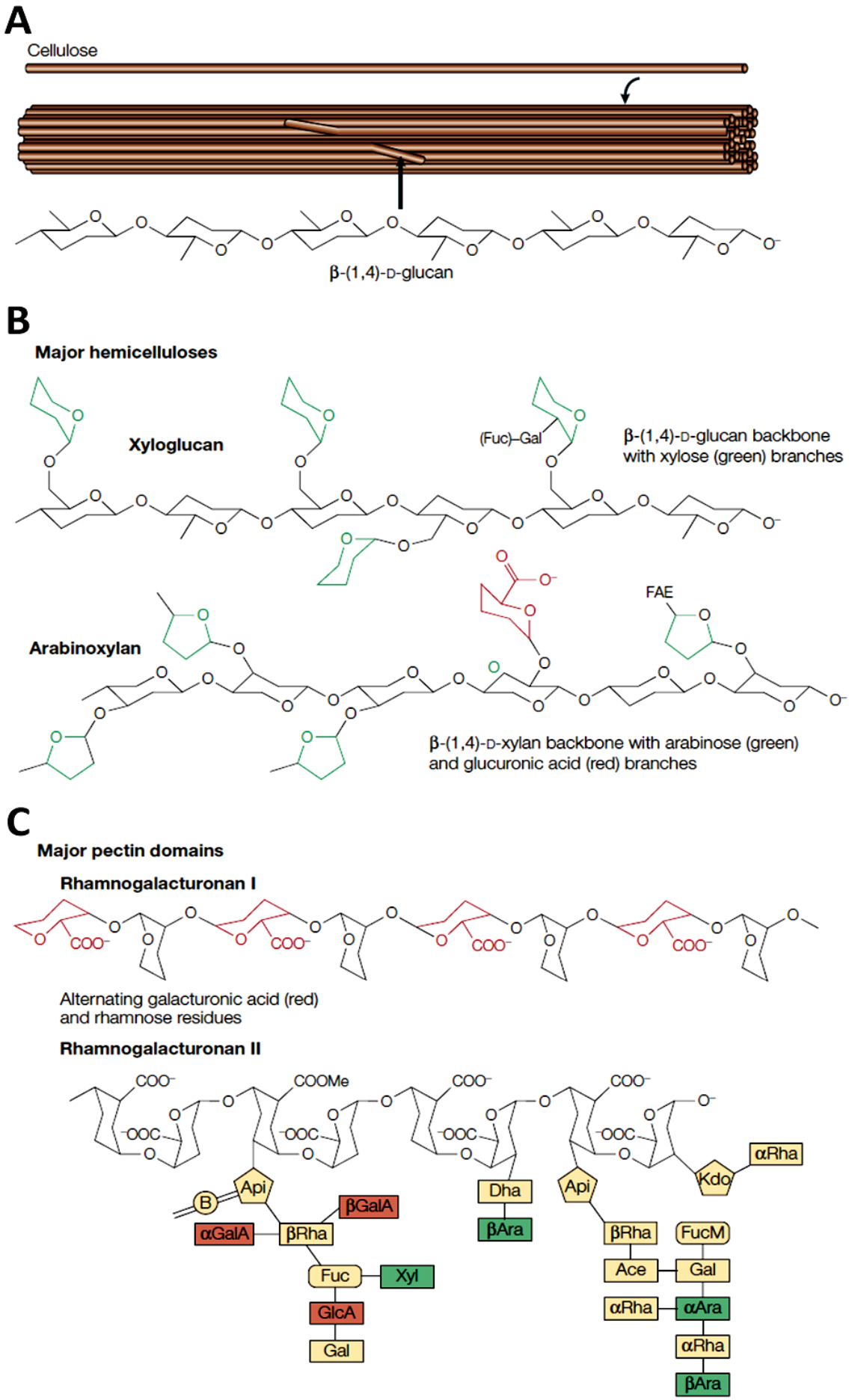
Figure 1.2. Generalised structure of the primary plant cell wall. The primary plant cell wall consists of three major components, cellulose microfibrils (purple rods), hemicelluloses and pectins. For simplicity, the hemicellulose–cellulose network (mainly xyloglucan (blue) and arabinoxylans (grey)) is shown on the left part of the cell wall scheme, and the pectins (rhamnogalacturonan I (orange), homogalacturonan (red), xylogalacturonan (red with green tracks), arabinan (green) and rhamnogalacturonan II (red circle)) are emphasized on the right. The box below the diagram depicts the various symbols used to signify the main plant cell wall structural polysaccharides. Covalent links (orange circles) bind pectin domains to xyloglucan and neutral pectin polysaccharides (green) can also bind cellulose. Figure adapted from Cosgrove (2005) with permission.

Hemicellulose is semi-soluble and makes up 20-50% of the total plant cell wall. Unlike cellulose, it is a family of highly branched, chemically heterogeneous, polymers of pentoses, hexoses and sugar acids that form intricate networks to strengthen the plant cell wall (Linder *et al.*, 2003; Kabel *et al.*, 2007). The forms and relative quantities of hemicellulose vary significantly with plant variety, cell type, and developmental phase (Table 1.2). Hemicellulose interacts with neighbouring cellulose microfibrils, and other non-cellulosic cell wall polymers such as pectins, in a structurally dependent manner, according to the Sticky Network model (Atalla *et al.*, 1993; Cosgrove, 2000; Nakamura *et al.*, 2002). The two main hemicellulose types are GAX and xyloglucans (Aspinall,

1980; Pauly *et al.*, 1999; Liab *et al.*, 2000). GAX is the predominant hemicellulose in the primary cell walls of grasses, and MLGs are strongly integrated with cellulose and non-cellulosic polysaccharides in primary walls of elongating cells (Carpita and Gibeaut, 1993; Carpita, 1996; Kim *et al.*, 2000; Carpita *et al.*, 2001). The backbones of xyloglucans are composed of a repeating pattern of (1,4)-β-D-glucose monomers, whereas MLGs consist of unbranched glucose homo-polymers containing both (1,3)-β-D-linked and (1,4)-β-D-glucose monomers. GAX in grasses is comprised of a (1,4)-β-D-xylopyranose monomer backbone that can be substituted with a number of different groups depending on the plant type and stage of growth, such as (1,2)-α-linked and (1,3)-α-linked L-arabinofuranosyl groups, *O*-acetyl, and (1,2)-α-D-glucuronic acid or 4-*O*-methyl-D-glucuronic acid groups to varying degrees (Borneman *et al.*, 1990). The α-L-arabinofuranosyl side groups may also form diferulic crosslinks with phenolic polymers such as ferulic acid, by becoming esterified at the *O*-2 or *O*-5 positions (Kulkarni *et al.*, 1999).

Pectin is structurally the most intricate family of polysaccharides, because it is involved in the control of cell wall structure and expansion, cell-cell interactions and signalling, as well as plant defence mechanisms (Jarvis, 1984). Pectin is mainly located in the primary cell wall, featuring in two distinct types of polysaccharides: homogalacturonan (HG) and rhamnogalacturonan I (RG-I), with smaller amounts of xylogalacturonan (XGA), arabinan, arabinogalactan I, and rhamnogalacturonan II (RG-II) (Shibuya and Nakane, 1984; Saulnier and Thibault, 1999). It is believed that pectic polysaccharides are covalently cross-linked to one another via their backbones. Thus it is possible that both HG and RG-I are very tightly associated with other plant cell wall polysaccharides including hemicellulose and cellulose (Nakamura *et al.*, 2002).

Lignin is a crucial secondary cell wall component (5-20% of plant dry matter) as it plays an important role in maintaining cell structure integrity, water transport and protecting against plant pathogens. Lignin is not susceptible to enzymatic degradation by the rumen microbiota, and hence its presence in the cell wall is thought to be partly responsible for the recalcitrance of forages to breakdown in the rumen (Campbell and Sederoff, 1996). An important attribute of grass lignins is the incorporation of diferulic acids during development by means of ether bonds linking the ferulic acid hydroxyl groups (Scalbert *et al.*, 1985; Kondo *et al.*, 1990). The simplified chemical structures and major components of these molecules are shown in Figure 1.3.


A
Cellulose
β-(1,4)-D-glucan
B
Major hemicelluloses
Xyloglucan
(Fuc)–Gal
β-(1,4)-D-glucan backbone
with xylose (green) branches
Arabinoxylan
FAE
β-(1,4)-D-xylan backbone with arabinose (green)
and glucuronic acid (red) branches
C
Major pectin domains
Rhamnogalacturonan I
Alternating galacturonic acid (red)
and rhamnose residues
Rhamnogalacturonan II
COO⁻
COOMe
-OOC
B
Api
βGalA
αGalA
βRha
Fuc
Xyl
GlcA
Gal
Dha
βAra
Api
αRha
Kdo
βRha
FucM
Ace
Gal
αRha
αAra
αRha
βAra

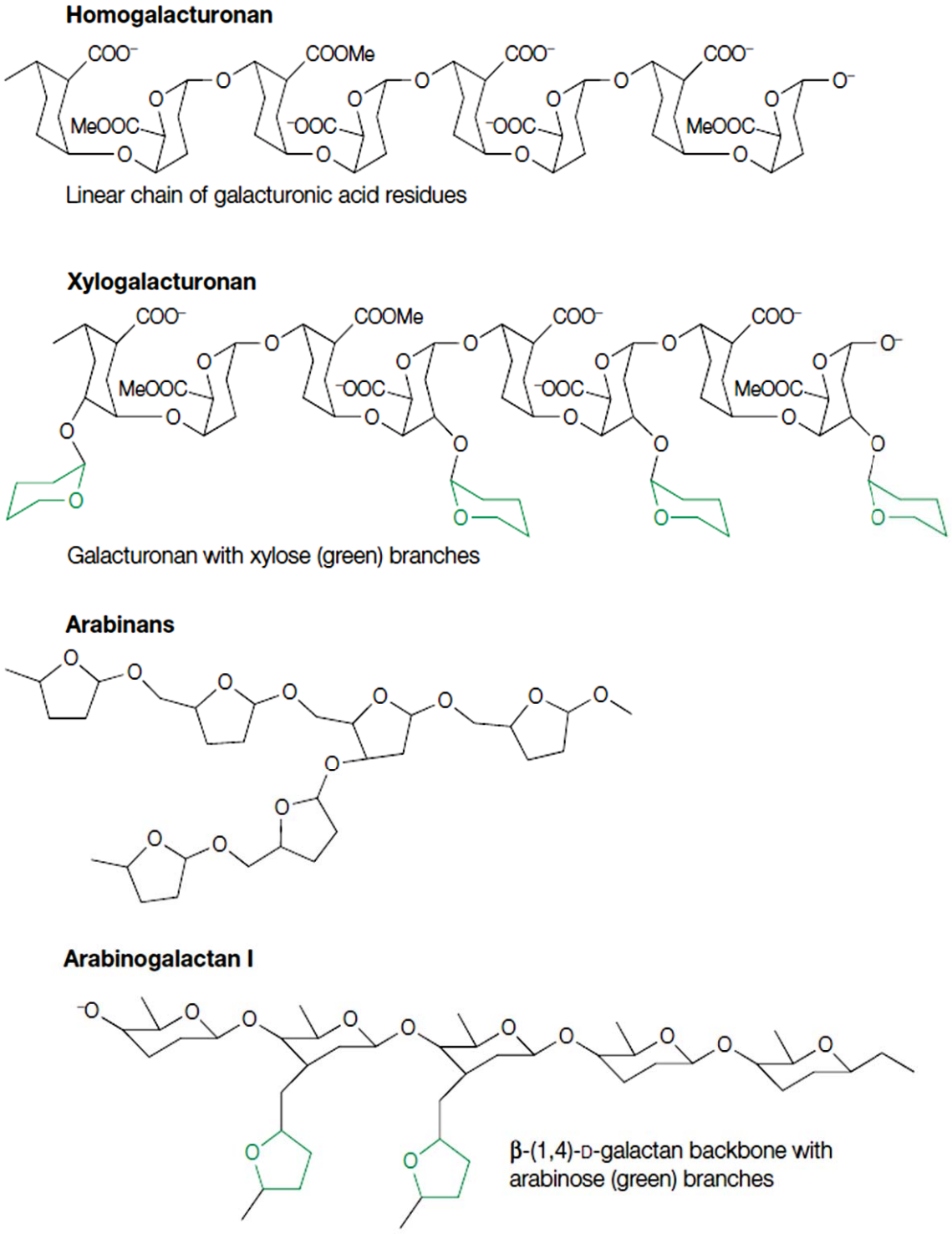


Figure 1.3. Simplified chemical structures of cellulose, hemicellulose and pectin. A, cellulose chain consisting of β-(1,4)-D-glucan monomers with intra-chain hydrogen bonding. B, major hemicelluloses: xyloglucan composed of β-(1,4)-D-glucan monomers with xylose (green) branches and arabinoxylan composed of β-(1,4)-D-xylan backbone with arabinose (green) and glucuronic acid (red) branches. C, major pectins: rhamnogalacturonan I with alternating galacturonic acid (red) and rhamnose residues, rhamnogalacturonan II, homogalacturonan consisting of a linear chain of galacturonic acid residues, xylogalacturonan composed of galacturonan with xylose (green) branches, arabinans and arabinogalactan I consisting of β-(1,4)-D-galactan backbone with arabinose (green) branches. Figure from Cosgrove (2005) with permission.

1.4 Plant fibre degradation

The complex structure and chemical composition of plant fibre material requires the combined action of several types of enzymes to break it down; (i) glycoside hydrolases (GH), which hydrolyse glycosidic bonds between carbohydrate monomers or between carbohydrate and non-carbohydrate moieties, (ii) carbohydrate esterases (CE), which hydrolyse the ester linkages of ferulic acid, acetate side chains or methoxy substituents and (iii) polysaccharide lyases (PL), which hydrolyse polysaccharide chains by β-elimination (Henrissat, 1991; Henrissat and Davies, 1997; Henrissat, 1998). Collectively, enzymes that degrade, modify, or create glycosidic bonds are known as Carbohydrate-Active enZYmes, or CAZymes (Cantarel *et al.*, 2009) and are organised in a dedicated database, called CAZy. As of March 2015, the CAZy database has 135 GH families, 16 CE families and 23 PL families.

These families are based on experimentally characterised proteins and are populated by proteins with significant sequence similarity from public databases. The CAZy classification reflects the structural features of these enzymes more so than their sole substrate specificity, and provides a useful framework for understanding enzyme mechanics, and improving the quality of annotations and functional predictions of genes from many genome projects. The types of fibre degrading enzymes important for forage degradation, their functions, EC number, CAZymes and Pfam protein families are summarised in Table 1.3. The chemical bonds hydrolysed by each class of enzyme are shown against the basic structural components of plant fibre in Figure 1.4.

For the complete hydrolysis of cellulose to glucose, three types of enzymes are required. Endo-β-1,4-glucanases (EC 3.2.1.4 and GH5, GH44, GH45, GH124) cleave 1,4-β-D-glycosidic bonds at internal positions along the cellulose chain generating cellodextrins of variable chain lengths or can cleave cellobiose residues. Exoglucanases such as cellodextrinases (EC 3.2.1.74 and GH9) and cellobiohydrolases (EC 3.2.1.150/176 and GH74, GH48) act in a progressive manner on the reducing or non-reducing ends of cellulose or xyloglucans to liberate glucose and cellobiose. Exoglucanases are also capable of degrading microcrystalline cellulose by stripping cellulose microfibrils from the microcrystalline structure (Teeri, 1997). The β-glucosidases (EC 3.2.1.21 and GH3, GH1) complete the degradation process by hydrolysing soluble cellobiose and cellodextrins from the reducing or non-reducing ends to glucose.

Table 1.3. Important fibre degrading enzymes involved in forage degradation in the rumen.

Main fibre-degrading enzymes	Function	Systematic Gene Name/Description	CAZyme family [A]	Pfam Protein Family [B]	EC Number [C]
Cellulases					
Endo/Exo-glucanases	Endohydrolysis of 1,4-β-D-glucosidic linkages in cellulose or cleave cellobiose residues.	Endo-1,4-β-glucanase/Cellulase	GH5, GH124	00150	3.2.1.4
	Hydrolysis of 1,4-linkages in 1,4-β-D-glucans, to remove successive glucose units.	Exo-1,4-β-glucanase/ cellodextrinase	GH9, GH44, GH45	00759	3.2.1.74
Cellobiohydrolases	Hydrolysis of 1,4-β-D-glucosidic linkages in cellulose releasing cellobiose.	Cellulose 1,4-β-cellobiosidase	GH48	02011	3.2.1.176
α-Glucosidases	Hydrolysis of terminal, non-reducing 1,4-α-D-glucose residues with release of *D*-glucose.	α-glucosidase	GH31, GH97	01055	3.2.1.20
β-Glucosidases	Hydrolysis of glucose from non-reducing end of cellobioses and cellooligodextrins.	β-glucosidase	GH1, GH3	00232, 00933	3.2.1.21
Hemicellulases					
Xylanases	Endohydrolysis of 1,4-β-D-xylosidic linkages in xylan backbone releasing xylooligomers.	Endo-1,4-β-xylanase/ Xylanase/chitin deacetylase	GH5, GH10, GH11	00150, 00331, 00457	3.2.1.8
	Hydrolysis of 1,4-β-D-xylose residues from the reducing end of oligosaccharides.	Reducing end xylose-releasing exo-oligoxylanase	GH8	01270	3.2.1.156
Mannanases	Hydrolysis of 1,4-β-D-mannosidic linkages in mannans releasing mannan oligomers.	Endo-1,4-β-glucanase/Cellulase	GH5	00150	3.2.1.4
		Endo-1,4-β-mannanase/β-mannanase	GH26	02156	3.2.1.78
	Hydrolysis of cellobiose from the reducing end of xyloglucans.	Oligoxyloglucan reducing-end cellobiohydrolase	GH74	02012	3.2.1.150
Lichenases	Endohydrolysis of 1,4-β-D-glucosidic linkages in β-D-glucans containing 1,3- and 1,4-bonds.	1,3-1,4-β-D-glucan 4-glucanohydrolase/ licheninase	GH8	01270	3.2.1.73
		Endo-1,3,4-β-glucanase	GH16	00722	3.2.1.6
β- Xylosidases	Hydrolysis of 1,4-β-D-xylans, to remove successive *D*-xylose residues from the non-reducing termini.	Xylan 1,4-β-xylosidase/β-xylosidase/O-Glycosyl hydrolase	GH3, GH30, GH39, GH43, GH120	00933, 02055, 01229, 04616, 07602	3.2.1.37
Endo-/Xylo-glucanases	Hydrolyze longer cellooligosaccharides, xylan and xyloglucans.	Endo-/Xylo-glucanases	GH44, GH45		3.2.1.4, 3.2.1.151
α-Glucuronidases	Cleave α-1,2 bond between glucuronic acid side chain substitutions releasing glucuronic acid.	α-D-glucuronidase Xylan α-1,2-glucuronosidase	GH4, GH67 GH115	02056, 07488 05270	3.2.1.139 3.2.1.131

α-L-Arabinofuranosidases	Hydrolysis of terminal non-reducing α-L-arabinofuranoside residues in α-L-arabinosides.	α-L-arabinofuranosidase/ xylosidase	GH43, GH51	04616, 06964	3.2.1.55
	Cleave arabinans at *O*-2 and *O*-3 positions on xylan back bone.	β-L-arabinofuranosidase	GH127	05270	3.2.1.185
1,4-β-Mannosidases	Hydrolysis of exo β-1,4-D bond of mannan oligomers releasing mannose.	β-mannosidase	GH1,GH2, GH5	00232, 00703, 00150	3.2.1.25
α-Mannosidases	Hydrolysis of terminal, non-reducing α-D-mannose residues in α-D-mannosides.	α-mannosidase	GH38	01074	3.2.1.24
Acetyl xylan esterases	Catalyses the hydrolysis of acetyl groups from polymeric xylan and acetylated xylose.	Acetyl xylan esterase/ Polysaccharide deacetylase	CE1, CE2, CE3, CE4, CE6, CE7, CE12	00756, 13472, 05448, 01522	3.1.1.72
Cinnamoyl/Feruloyl esterases	Cleaves hydroxycinnamic acids side chain substitutions releasing p-coumaric and ferulic acids.	Feruloyl esterase	CE1	00756	3.1.1.73
Pectinases					
α-Galactosidases	Hydrolysis of terminal, non-reducing α-D-galactose residues in α-D-galactosides, including galactose oligosaccharides, galactomannans and galactolipids.	α-D-galactosidase/Melibiase	GH27, GH97	02065	3.2.1.22
β-Galactosidases	Hydrolysis of terminal non-reducing β-D-galactose residues in β-D-galactosides.	β-galactosidase	GH35, GH42	01301, 02449	3.2.1.23
α-L-Fucosidases	Hydrolysis of α-1,6-linked fucose reducing-end *N*-acetylglucosamine from carbohydrate moieties of glycoproteins.	α-L-fucosidase	GH29	01120	3.2.1.51
	Hydrolysis of terminal non-reducing α-1,2-linked L-fucose residues from glycoproteins and oligosaccharides.	α-1,2-L-fucosidase	GH95		3.2.1.63
Pectate lyases	Eliminative cleavage of 1,4-α-D-galacturonan to give oligosaccharides.	1,4-α-D-galacturonan lyase/Pectate lyases	PL1	00544	4.2.2.2
	Cleave main chain and branches in pectic polysaccharides.	Pectate lyase	PL9, PL10	14200, 09492	
α-Amylases	Hydrolysis of 1,4-α-D-glucosidic linkages in polysaccharides containing three or more 1,4-α-linked *D*-glucose units.	α-amylase	GH13	00128	3.2.1.1
Pullulanases	Hydrolysis of 1,6-α-D-glucosidic linkages in pullulan and dextrins of amylopectin and glycogen.	Pullulanase			3.2.1.41

Polygalacturonases	Hydrolyse α-1,4 glycosidic linkages between galacturonate residues in polygalacturonic acid.	Polygalacturonase/Exo-polygalacturonase	GH28	00295	3.2.1.15, 3.2.1.67
Rhamnogalacturonases	Cleave α-1,2 linkages between galacturonic acid and rhamnose residues.	Rhamnogalacturonase			3.2.1.171
Rhamnohydrolases	Cleave single rhamnose units from the non-reducing end of rhamnogalacturonan.	Rhamnogalacturonan rhamnohydrolase			3.2.1.174
Rhamnogalacturonan galacturonohydrolases	Cleave single galacturonate units from the non-reducing end of rhamnogalacturonan.	Rhamnogalacturonan α-1,2-galacturono-hydrolase			3.2.1.173
Endo-1,4-β-galactanases	Specific hydrolyses 1,4-β-D-galactosidic linkages in type I arabinogalactans.	Arabinogalactan endo-1,4-β-galactosidase	GH53	07745	3.2.1.89
α-L-Rhamnosidases	Hydrolysis of terminal non-reducing α-L-rhamnose residues in α-L-rhamnosides.	α-L-rhamnosidase	GH78/ GH106	05592	3.2.1.40
β-Glucuronyl/ Rhamnogalacturonyl hydrolases	The enzyme is part of the degradation system for rhamnogalacturonan II.	d-4,5-unsaturated β-glucuronyl/ rhamnogalacturonyl hydrolase	GH88/ GH105	07470	3.2.1.-, 3.2.1.172
Pectin methyl esterases	Demethylation of pectin in the plant cell wall.	Pectin methylesterase	CE8	01095	3.1.1.11
Pectin acetyl esterases		Pectin acetylesterase/Pectin esterase	CE12/ CE13	13472, 03283	
Rhamnogalacturonan acetyl esterases	Hydrolytic cleavage of 2- or 3-*O*-acetyl groups of α-D-galacturonic acid in rhamnogalacturonan I.	Rhamnogalacturonan 2/3-*O*-acetyl-α-D-galacturonate *O*-acetylhydrolase	CE12	13472	3.1.1.86
Rhamnogalacturonan lyases	The enzyme is part of the degradation system for rhamnogalacturonan I.	α-L-rhamnopyranosyl-1,4-α-D-galacto-pyranosyluronate endolyase	PL11	14099	4.2.2.23
Others					
Methyl-glucuronoyl methylesterases	Cleave ester linkage between 4-*O*-methyl-D-glucuronic acid and alcohol group.	4-*O*-methyl-glucuronoyl methylesterase/Carboxylesterase	CE15	05870	3.1.1.1
Invertases	Inversion/hydrolysis of fructans in grasses containing polysaccharides such as inulin and sucrose.	Invertase, Endo/Exo-Inulinase, sucrose-fructosyltransferase	GH32	00251	3.2.1.26, 3.2.1.7, 3.2.1.80, 2.4.1.-
Phosphorylases	Phosphorolyze/cleave protein-bound β-1,2-glucans.	Cellobiose, cellodextrin, chitobiose phosphorylase	GH94	06165	2.4.1.20, 2.4.1.49, 2.4.1.-

Descriptions and classifications compiled from the [A], CAZy database (Cantarel *et al.*, 2009), [B], Pfam and *i*Pfam databases (Finn *et al.*, 2005; Finn *et al.*, 2013), and [C], Enzyme commission number from the PDB database (Rose *et al.*, 2013).

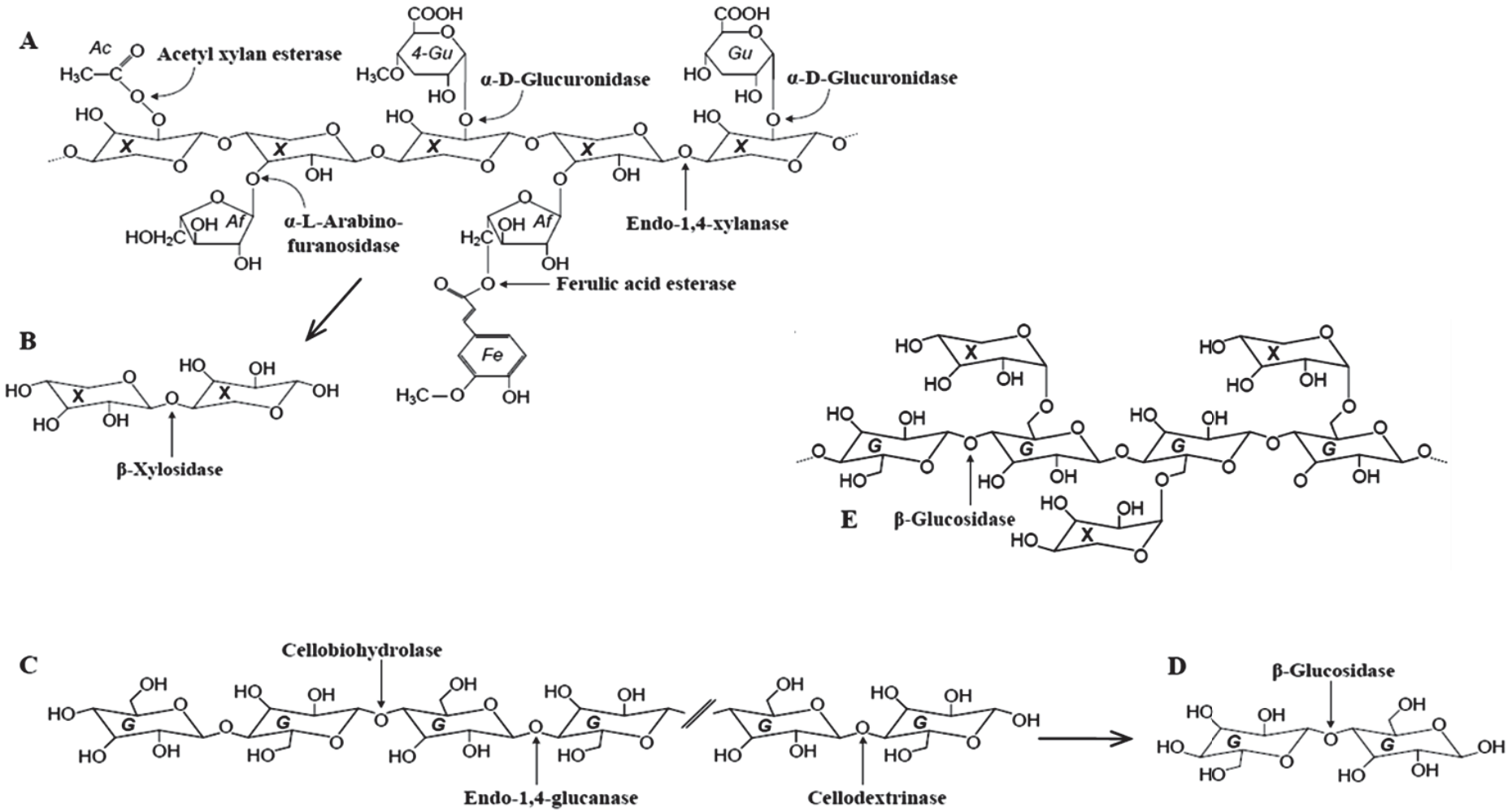


Figure 1.4. Enzymes involved in plant fibre-degradation. Sites of the enzymatic hydrolysis are shown for GAX (A), xylobiose (B), cellulose (C), cellobiose (D), and xyloglucan (E). Ac, *O*-acetyl; Af, α-L-arabinofuranoside; Fe, ferulic acid; G, glucose; 4-Gu, 4-*O*-methyl-D-glucuronic acid; Gu, glucuronic acid; X, xylopyranose. Figure adapted from Shallom and Shoham (2003) with permission.

Due to their complex nature and diverse chemical structures, hydrolysis of hemicelluloses requires a variety of enzymes. Endoxylanases, mainly endo-1,4-β-D-xylanases (EC 3.2.1.8/156 and GH5, GH10, GH11, GH8), cleave the internal linkages within the xylan backbone and from the reducing end of oligosaccharides, forming xylobiose and xylooligosaccharides (Shallom and Shoham, 2003). These end products are further successively degraded into xylobiose and/or xylose residues from the non-reducing termini by β-D-xylosidases, and namely 1,4-β-D-xylan-xylohydrolases (EC 3.2.1.37 and GH3, GH30, GH39, GH43, GH120). The diverse range of side groups attached to hemicellulose are hydrolysed from the xylan backbone by the activity of α-D-glucuronidases (EC 3.2.1.139/131 and GH67, GH115), α-L-arabinofuranosidases (EC 3.2.1.55 and GH43, GH51) (Sørensen *et al.*, 2003), acetyl-xylan esterases (EC 3.1.1.72/3 and CE1, CE2, CE3, CE4, CE6, CE7, CE12) ferulic acid esterases (EC 3.2.1.73 and CE1) and *p*-coumaric acid esterases (Kosugi *et al.*, 2002).

Pectin degradation is facilitated by three classes of enzymes that function to depolymerise the pectin backbone in the plant cell wall; lyases (acting by β-elimination), hydrolases, and esterases which hydrolyse acetyl, methyl, or ferulic acid substituents from the main chains of the 1,4-α-galacturonic acids. Pectate lyases (EC 4.2.2.2 and PL1, PL9, PL10) cleave 1,4-α-linkages between galacturonic acids and may preferentially attack regions of unmethylated polygalacturonate (pectate), although they can act on regions with low levels of methylation (Kluskens *et al.*, 2003). Rhamnogalacturonan lyases (EC 4.2.2.23 and PL11) catalyse cleavage of the polysaccharide rhamnogalacturonan I (RG-I) backbone in ramified regions of pectin (Ochiai *et al.*, 2009).

Hydrolases involved in pectin degradation mostly contain the GH28 catalytic domain and include polygalacturonases (EC 3.2.1.15), exo-polygalacturonases (EC 3.2.1.67), exo-poly-α-galacturonases (EC 3.2.1.82) and rhamnogalacturonases (EC 3.2.1.- and CE12) (Tardy *et al.*, 1997). Polygalacturonans are cleaved at random internal positions, and at the non-reducing chain ends in polygalacturonans. Rhamnogalacturonases hydrolyse bonds between D-galacturonic acid and L-rhamnose in rhamnogalacturonan I (Sakai *et al.*, 1993). Acetylesterases (EC 3.1.1.11 and CE12) remove C-2 and C-3 attached acetyl groups, and methylesterases (EC 3.1.1.11 and CE8) act on methoxylated groups releasing methanol. Pre-treatment of esterified pectins with methylesterase enhances acetylesterase activity, suggesting there is an optimal sequence of esterase activities for pectin degradation (Shevchik and Hugouvieux-Cotte-Pattat, 2003).

1.5 Rumen microbial diversity and the fibre degrading rumen microbiome

The rumen microbiome is composed of bacteria (10^{10}-10^{11} cells/mL of rumen contents), methanogenic archaea (10^{7}-10^{9} cells/mL), protozoa (10^{4}-10^{6} cells/mL), fungi (10^{3}-10^{6} cells/mL) and bacteriophage (10^{9}-10^{10} particles/mL) (Stewart *et al.*, 1997; Flint, 2004; Kamra, 2005; Sirohi *et al.*, 2012). The majority of rumen microbes are obligate anaerobes, however low numbers of facultative anaerobes are essential in maintaining anaerobic conditions in the rumen, by immediately metabolizing oxygen introduced as plant material is ingested (Hobson and Stewart, 1997). Culture-based studies of the rumen microbiota have identified a wide range of bacteria, fungi and protozoa that work in concert to degrade and ferment lignocellulosic biomass (Krause *et al.*, 2003; Kim *et al.*, 2011b). However, cultivation-independent molecular surveys investigating diversity of the rumen microbiome identified 19 bacterial phyla of rumen bacteria and revealed that only 11% of the total taxa have been cultured (Tajima *et al.*, 1999; Edwards *et al.*, 2004). Over 90% of all rumen bacteria (Figure 1.5) were categorised as the low G+C Gram-positive Firmicutes, represented mainly by the clostridial families, Lachnospiraceae and Ruminococcaceae, or the Gram-negative Bacteroidetes, dominated by Prevotellaceae (Tajima *et al.*, 1999; de Menezes *et al.*, 2011).

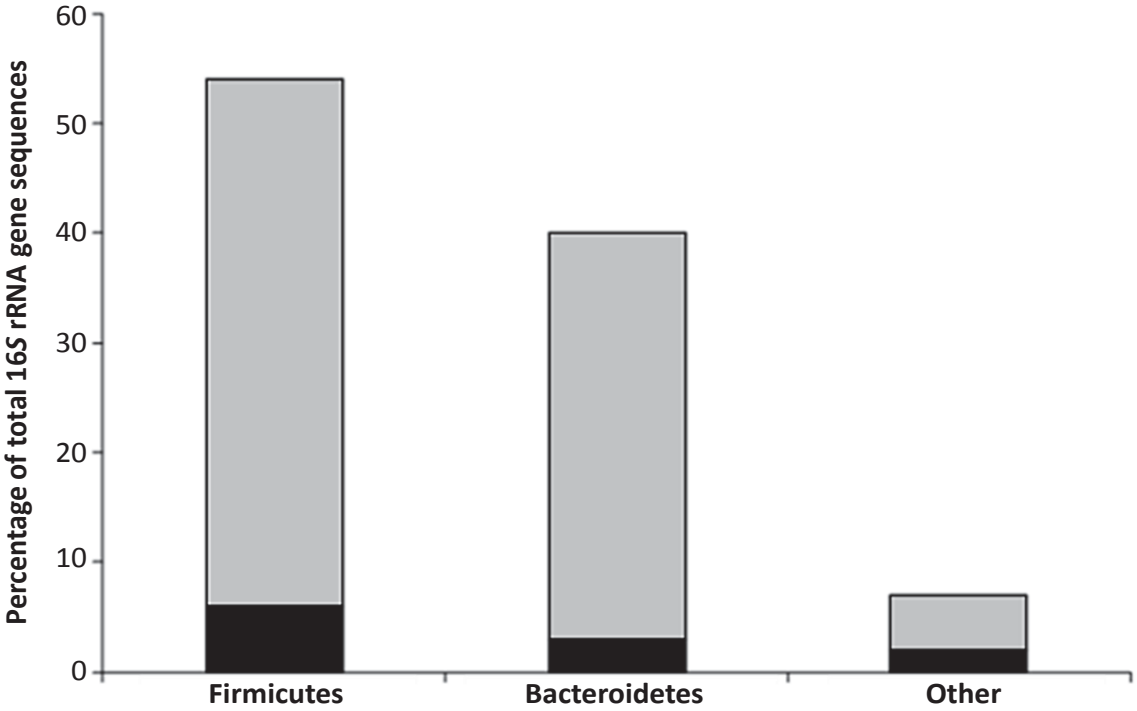


Figure 1.5. Bacterial diversity of the rumen microbial ecosystem. The percentage of 16*S* rRNA gene sequences represented by cultivated (black) and uncultivated (grey) bacteria is shown for the two predominant phyla, and minor or unknown phyla. Figure adapted from Edwards *et al*. (2008) with permission.

A widely used DNA-based approach to define taxonomic levels across bacteria is to assign their 16*S* rRNA gene sequences into bins based on similarity to other sequences in the community, referred to as operational taxonomic units (OTUs), where stringent sequence similarity cut-offs can be used to classify OTUs at species- and genus-level (Kenters *et al.*, 2011; Schloss and Westcott, 2011). To date, the accepted estimations of total bacterial species-level OTUs number between 400 to over 5000 (Edwards *et al.*, 2004; Brulc *et al.*, 2009; Kong *et al.*, 2010; Hess *et al.*, 2011; Kim *et al.*, 2011b). Therefore, the fundamental bacterial community differences of the rumen are observed at the genus- and species-levels (Table 1.4). This begs the exploration of a "core rumen microbiome", which is universally present across different animals, microbial fractions and diets (Creevey *et al.*, 2014). The core bovine rumen microbiome is comprised of 25-32 genera, where the predominant genus is *Prevotella*, but also include *Butyrivibrio*, *Fibrobacter*, *Oscillibacter, Pseudobutyrivibrio, Ruminococcus*, *Succiniclasticum* and *Treponema* (de Menezes *et al.*, 2011; Jami and Mizrahi, 2012; Wu *et al.*, 2012; Jami *et al.*, 2013).

The largest single study to date examined the microbial communities across 742 individual ruminants consisting of 32 different species, across a range of diets, and geographical regions, and determined that the core rumen bacterial microbiome is comprised of seven bacterial groups (Henderson *et al.*, 2015). The seven most abundant bacterial groups which comprised 67% of all bacterial sequence data were *Prevotella*, *Butyrivibrio*, and *Ruminococcus*, as well as unclassified Lachnospiraceae, Ruminococcaceae, Bacteroidales, and Clostridiales. With the exception of *Butyrivibrio*, the true extent of rumen bacterial diversity is currently unknown, as these genera and their functions are not sufficiently characterised by available cultures (Paillard *et al.*, 2007a).

Diet is the most significant determinant of rumen bacterial community composition within the sheep and cow rumen (Tajima *et al.*, 2000; Brulc *et al.*, 2009; Callaway *et al.*, 2010; Durso *et al.*, 2010; Pitta *et al.*, 2010; Kim *et al.*, 2011a). The rumen contents of cows consuming a forage-based diet, demonstrated a significant difference in bacterial community composition between the solid and liquid fractions (Fouts *et al.*, 2012). The solid fraction of rumen content was shown to be dominated by members of the order Clostridiales, predominantly *Butyrivibrio*, whereas the liquid fraction was predominantly from the order Bacteroidales, mainly *Prevotella* and *Tannerella*.

Table 1.4. The major fibre degrading bacteria of the bovine rumen.

Phylum	Species	Predominant Polysaccharide degraded [A]	Predominant niche [B]	Main fermentation products [C]	Reference
Firmicutes (Gram-positive)	*Butyrivibrio fibrisolvens*	X and P	Solid	A, B, F and L [D]	Bryant and Small (1956a)
	Butyrivibrio hungatei	X and P	Solid	A, B, F and L	Kopecny *et al.* (2003)
	Butyrivibrio proteoclasticus	X and P	Solid	A, B and F	Moon *et al.* (2008)
	Pseudobutyrivibrio xylanivorans	X and P	Solid	A, B, F and L	Kopecny *et al.* (2003)
	Lachnospira sp.	P	Solid	A, E, F, C	Bryant and Small (1956b)
	Ruminococcus flavefaciens	C and X	Solid	A, F and S	Ayers (1958)
	Ruminococcus albus	C and X	Solid	A and F	Patterson *et al.* (1975)
	Eubacterium cellulosolvens	C and X	Solid	B, L and F	Anderson and Blair (1996)
	Clostridium sp.	C and X	Solid	A, B, E and H	Kelly *et al.* (1987)
Bacteroidetes (Gram-negative)	*Prevotella albensis*	X and P	Liquid	A and S	Avguštin *et al.* (1997)
	Prevotella brevis	X and P	Liquid	A and S	Avguštin *et al.* (1997)
	Prevotella bryantii	X and P	Liquid	A and S	Avguštin *et al.* (1997)
	Prevotella ruminicola	X and P	Liquid/Solid	A, F and S	Avguštin *et al.* (1997)
Fibrobacteres (Gram-negative)	*Fibrobacter succinogenes*	C	Solid	F and S	Hungate (1950)
	Fibrobacter intestinalis	C	Solid	F and S	Montgomery *et al.* (1988)

[A], C, cellulose; P, pectin; X, xylan (Weimer, 1996). [B], refers to solid or liquid fractions of the rumen contents (Hungate, 1966). [C], A, acetate; B, butyrate; C, carbon dioxide; E, ethanol; F, formate; H, hydrogen; L, lactate; S, succinate (Weimer, 1998). [D], different *B. fibrisolvens* strains produce different fermentation products (Kopecny *et al.*, 2003).

Other similar studies have also demonstrated the presence of *Butyrivibrio* within the solid fraction of pasture-fed cows (Edwards *et al.*, 2004; Kong *et al.*, 2010; Pitta *et al.*, 2010). In a study that used tag amplicon pyrosequencing to characterise the rumen bacterial populations of 16 individual lactating cows fed a concentrate diet (Jami and Mizrahi, 2012), it was shown that *Butyrivibrio* was the second most abundant genus (relative abundance of 10.1%) after *Prevotella* (51%) within the core bacterial community shared by all animals. Furthermore the previously described study by Henderson *et al*., (2015), demonstrated that in rumen samples from bovines fed mixes of forage and concentrates, *Butyrivibrio* was found to be the most abundant genus in terms of the number of species present.

The adherence of microbes to plant material is important for efficient degradation of plant cell wall polysaccharides (McAllister *et al.*, 1994; Larue *et al.*, 2005), and it has been shown that bacteria attach to plant material within minutes of feed ingestion (Koike *et al.*, 2003a; Edwards *et al.*, 2008). Of the total microbial cells in the rumen, 70 to 80% of the total metabolically active rumen microbiota are fibre adherent bacteria (Edwards *et al.*, 2008). The rumen plant-adherent bacterial community is dominated by the Firmicutes (Kim *et al.*, 2011b), and these adherent bacteria also account for up to 80% of the total cellulase and xylanase activities (Graham and Balnave, 1995; Morgavi *et al.*, 2000; Wallace *et al.*, 2001), as well as 70% of amylase and 75% of the protease activities in the rumen (Minato *et al.*, 1966; Brock *et al.*, 1982). Free-living or planktonic bacteria associated with the liquid phase of the rumen, account for the remaining 20 to 30% of the total community microbes and are suspected to utilise products released into the rumen liquid phase from initial plant polysaccharide breakdown by the adherent rumen microbes (McAllister *et al.*, 1994; Bayer *et al.*, 2013). Therefore, it is thought that in the context of the rumen, the predominant degraders of lignocellulosic plant material are the fibre adherent bacteria (Flint, 2004; Foroozandeh *et al.*, 2009). The fibre-associated Gram-negative *Prevotella ruminicola* and *Prevotella bryantii* (Fondevila and Dehority, 1996; Purushe *et al.*, 2010), are generalists, able to degrade and utilise many different polysaccharides (except cellulose). Additional fibre-associated uncharacterised bacterial species, such as *Eubacterium ruminantium*, also exist which are likely to have important roles in rumen fibre degradation (Koike *et al.*, 2003b; Kim and Yu, 2012).

In 2011, a major sequencing project entitled "The Hungate 1000: a catalogue of reference genomes from the rumen microbiome" was initiated (Kelly *et al.*, 2014). The project aims to produce a reference set of rumen microbial genome sequences from

cultivated rumen bacteria and methanogenic archaea, together with representative cultures of rumen anaerobic fungi and ciliate protozoa, in order to capture the genomic diversity of organisms in the rumen that are of functional significance (Leahy *et al.*, 2013; Creevey *et al.*, 2014). As of March 2015, almost 350 genomes have been sequenced as part of this initiative and are publically available in draft or complete form. The information generated from the reference genome sequences will be useful for future interpretation of metagenomic data sets.

To date, the genome sequence analyses of multiple fibrolytic rumen bacteria showed that each organism appears to have a unique approach to plant polysaccharide breakdown. For example, the Gram-negative *Fibrobacter succinogenes* (Huang and Forsberg, 1990; Jun *et al.*, 2007; Suen *et al.*, 2011b; Ransom-Jones *et al.*, 2012) specialises as a cellulose degrader, by physical interaction and uptake of large polymers which are degraded in the periplasm by various endoglucanases (Brumm *et al.*, 2011; Suen *et al.*, 2011b); whereas the Gram-positive *Ruminococcus flavefaciens* and *Ruminococcus albus* (Rincon *et al.*, 2003; Miller *et al.*, 2009; Suen *et al.*, 2011a) both secrete a wide variety of enzymes, a number of which are assembled into an extracellular cellulosome, which facilitates plant polysaccharide degradation (Miller *et al.*, 2009). Species of the Gram-positive *Butyrivibrio* and *Pseudobutyrivibrio* genera are considered to be among the primary hemicellulose degraders of the rumen (Krause *et al.*, 2003; Kelly *et al.*, 2010; Morgavi *et al.*, 2013), and are abundant in the rumen of animals on high fibre diets (Attwood and Reilly, 1998; Paillard *et al.*, 2007b).

1.6 Diversity of rumen *Butyrivibrio* and *Pseudobutyrivibrio*

Members of the genus *Butyrivibrio* were first described as curved rod-shaped (or vibroids), non-spore-forming, obligate anaerobes that produced butyrate from fermentation of carbohydrates with a motile phenotype due to possession of one or more polar to sub-polar flagella (Bryant, 1986). *Butyrivibrio* were initially described as Gram-negative even though numerous strains have displayed an atypical Gram-positive cell wall ultrastructure (Cheng and Costerton, 1977; Dibbayawan *et al.*, 1985; Hespell *et al.*, 1993). *Butyrivibrio*, are a major component of the ruminal microflora and have been isolated from the gastrointestinal tracts and faeces of various ruminants, monogastric animals and humans (Bryant, 1986; Hespell, 1992b). *Butyrivibrio* contribute to digestion of fiber in both animals (Orpin *et al.*, 1985; Miron, 1991; Hespell, 1992b) and humans (Rumney *et al.*, 1995) due to their ability to degrade hemicellulose (Dehority, 1966;

Hespell, 1992a; Williams and Withers, 1992; Miron and Ben-Ghedalia, 1993a). Later they were shown to be involved in protein breakdown (Blackburn and Hobson, 1962), and biohydrogenation of fatty acids (Polan *et al.*, 1964; Kepler *et al.*, 1966).

The *Butyrivibrio* genus originally consisted of only one species, *Butyrivibrio fibrisolvens* which includes anaerobic monotrichous butyrate-producing curved rods obtained from animals (Bryant and Small, 1956a). In addition to the various phenotypic characterisations, previous studies have utilised DNA-DNA hybridisation, 16*S* rRNA gene sequencing and 16*S* rRNA-based hybridisation probes in an attempt to differentiate these organisms. Earlier studies on the collection of strains initially designated as *B. fibrisolvens*, within the *Clostridium* subphylum XIVa (Collins *et al.*, 1994), attempted to partition the *Butyrivibrio* genus based on phenotypic characteristics (Shane *et al.*, 1969; Hazlewood *et al.*, 1986), DNA-DNA hybridisation (Mannarelli, 1988; Mannarelli *et al.*, 1990), and extracellular polysaccharide analyses (Stack, 1988), but were largely unsuccessful due to heterogeneity of the strains. The defining characteristics of *Butyrivibrio* strains, such as their ability to metabolise similar carbohydrates, production of the same fermentation products, as well as cell wall structure and other phenotypic traits, raises fundamental questions regarding the phylogenetic intermixing and taxonomy of the *Clostridium* sub-cluster.

The DNA-DNA hybridization studies identified eight distinct *Butyrivibrio* genospecies and many other strains with no close *Butyrivibrio* relatives (Mannarelli, 1988; Mannarelli *et al.*, 1990). One of these groups contained the type strain *Butyrivibrio crossotus* (Moore *et al.*, 1976), which includes lophotrichous isolates obtained from human faeces, represented by the type strain *B. crossotus* NCFB 2416 (Willems *et al.*, 1996). Due to the substantial morphological (Bryant and Small, 1956a; Dibbayawan *et al.*, 1985), metabolic (Bryant and Robinson, 1962; Shane *et al.*, 1969; Roché *et al.*, 1973), and serological (Margherita and Hungate, 1963; Hazlewood *et al.*, 1986) diversity uncovered in *Butyrivibrio* strains, it is generally accepted that the genus *Butyrivibrio*, as previously defined, is phylogenetically heterogeneous and comprises many more distinct species groups (Bryant and Small, 1956a; Shane *et al.*, 1969; Bryant, 1986; Hespell, 1992b).

The taxonomy of *Butyrivibrio* was revised in 1996 with phylogenetic studies using 16*S* rRNA gene sequencing (Forster *et al.*, 1996; Willems *et al.*, 1996) and 16*S* rRNA-based hybridisation probes (Forster *et al.*, 1997). At least 13 separate *Butyrivibrio* genotypic groups were identified and a new genus, *Pseudobutyrivibrio,* was established

to accommodate the observed diversity amongst the newly discovered strains (Van Gylswyk *et al.*, 1996). The genotypic data recognised that the different *Butyrivibrio* strains corresponded to three distinct species groups based on their 16*S* rRNA gene sequences; two separate *B. fibrisolvens* and the *B. crossotus* lineages (Willems *et al.*, 1996). On the basis of its 16*S* rRNA phylogenetic profile, *B. crossotus* was thought to represent a separate and new genus. In a subsequent study based on 16*S* rRNA gene RFLP patterns (Kopecny *et al.*, 2003), *B. fibrisolvens* strains were divided into five taxa: *Butyrivibrio fibrisolvens*, *Butyrivibrio proteoclasticus*, *Butyrivibrio hungatei*, *Pseudobutyrivibrio ruminis* and *Pseudobutyrivibrio xylanivorans*.

The existing data does not allow for the *sensu stricto* phenotypic differentiation between *Butyrivibrio* and *Pseudobutyrivibrio*. At present the *Butyrivibrio* and *Pseudobutyrivibrio* genera are divided into six species that are distantly related and each may eventually be designated as a separate genus (Moon *et al.*, 2008). The rumen *Butyrivibrio* isolates are currently represented by *B. fibrisolvens*, *B. hungatei*, *B. proteoclasticus* and *B. crossotus*, and the *Pseudobutyrivibrio* isolates *P. xylanivorans* and *P. ruminis* (Figure 1.6), all of which belong to the genetically diverse Lachnospiraceae family, order Clostridiales (Clostridial sub-cluster XIVa).

Butyrivibrio and *Pseudobutyrivibrio* are considered the most efficient degraders of plant hemicelluloses amongst the rumen bacteria (Orpin *et al.*, 1985; Willems *et al.*, 1996; Forster *et al.*, 1997; Kopecny *et al.*, 2001; Kopecny *et al.*, 2003; Kim *et al.*, 2011b). Another important feature of *Butyrivibrio* species is their strong proteolytic activity (Attwood and Reilly, 1996). Species belonging to the genera *Butyrivibrio* and *Pseudobutyrivibrio* are thought to be the main butyrate producers in the rumen (Diez-Gonzalez *et al.*, 1999; Paillard *et al.*, 2007b). *Butyrivibrio* are capable of growing on a range of carbohydrates, from simple mono- or oligosaccharides to complex plant polysaccharides such as pectins, mannans, starch and hemicelluloses (Moon *et al.*, 2008). Furthermore, xylans of diverse chemical and physical properties from a range of forages are degraded by *Butyrivibrio* strains (Hespell and Cotta, 1995). Recent physiological studies have shown that several strains of *B. fibrisolvens* adhere to the solids fraction in rumen contents (Whitford *et al.*, 1998; Tajima *et al.*, 1999; Koike *et al.*, 2003b), and genomic analyses have proposed the proteins that mediate this adhesion. However the specific molecular mechanisms and the key fibre adhesins responsible for bacterial adhesion to the plant fibre have not been experimentally confirmed as yet.

1.7 *Butyrivibrio proteoclasticus* B316

Butyrivibrio proteoclasticus B316 was the first *Butyrivibrio* to be genome sequenced (Kelly *et al.*, 2010). This strain was isolated from rumen contents of a NZ cow grazing on ryegrass-clover pasture (Attwood *et al.*, 1996), and was originally named *Clostridium proteoclasticum* (Figure 1.6). However, further taxonomic studies indicated that its reclassification into the genus *Butyrivibrio* was more appropriate (Attwood *et al.*, 1996; Moon *et al.*, 2008). B316 has a Gram-positive cell wall structure even though cultures stain Gram-negative (Figure 1.7). The cells are slightly curved rods that form short chains or pairs and under certain conditions they can form long chains. The B316 cells have a single sub-polar flagellum, however they are not motile. *B. proteoclasticus* species are present, along with closely related *Butyrivibrio* strains, in pasture-fed ruminants. Estimates of *B. proteoclasticus* populations in the rumen bacterial community via competitive PCR or real time PCR are ~2% and ~9% respectively of total bacterial cell counts from whole rumen samples (Attwood and Reilly, 1998; Paillard *et al.*, 2007b). Consistently, approximately 3% *B. proteoclasticus* clones were detected in rumen bacterial 16*S* rRNA gene libraries (Edwards *et al.*, 2004).

Physiological analyses showed that B316 is highly xylanolytic and proteolytic and produces butyrate and formate as terminal catabolites (Attwood *et al.*, 1996). Besides xylan, B316 rapidly degrades other complex polysaccharides, such as pectin and starch, but not crystalline cellulose. B316 also utilises a range of soluble mono- and disaccharides, such as arabinose, cellobiose, galactose, glucose, rhamnose, sucrose, and xylose (Attwood *et al.*, 1996; Moon *et al.*, 2008). A metabolic footprinting study found that B316 was able to breakdown more fibre than its close relative, *B. fibrisolvens* AR10, producing tenfold more xylose, using almost tenfold more glucose, and producing substantial levels of rhamnose, α-mannoside, and β-arabinoside, indicating rapid degradation of xylan and pectin (Villas-Bôas *et al.*, 2006). This strain is also capable of converting linoleic acid into stearic acid and thus may have an important role in biohydrogenation of fatty acids in the rumen (Wallace *et al.*, 2006).

The B316 genome was found to be 4.4 Mb in size with a total of 3,813 CDSs distributed across four replicons: the main chromosome (BPc1), two megaplasmids (pCY360 and pCY186) and a secondary chromosome or chromid (BPc2) (Table 1.5 and Figure 1.8). The megaplasmids were predicted to encode genes important for survival in the rumen, but no genes have been identified that contribute to the breakdown of lignocellulose (Kelly *et al.*, 2010; Yeoman *et al.*, 2011).

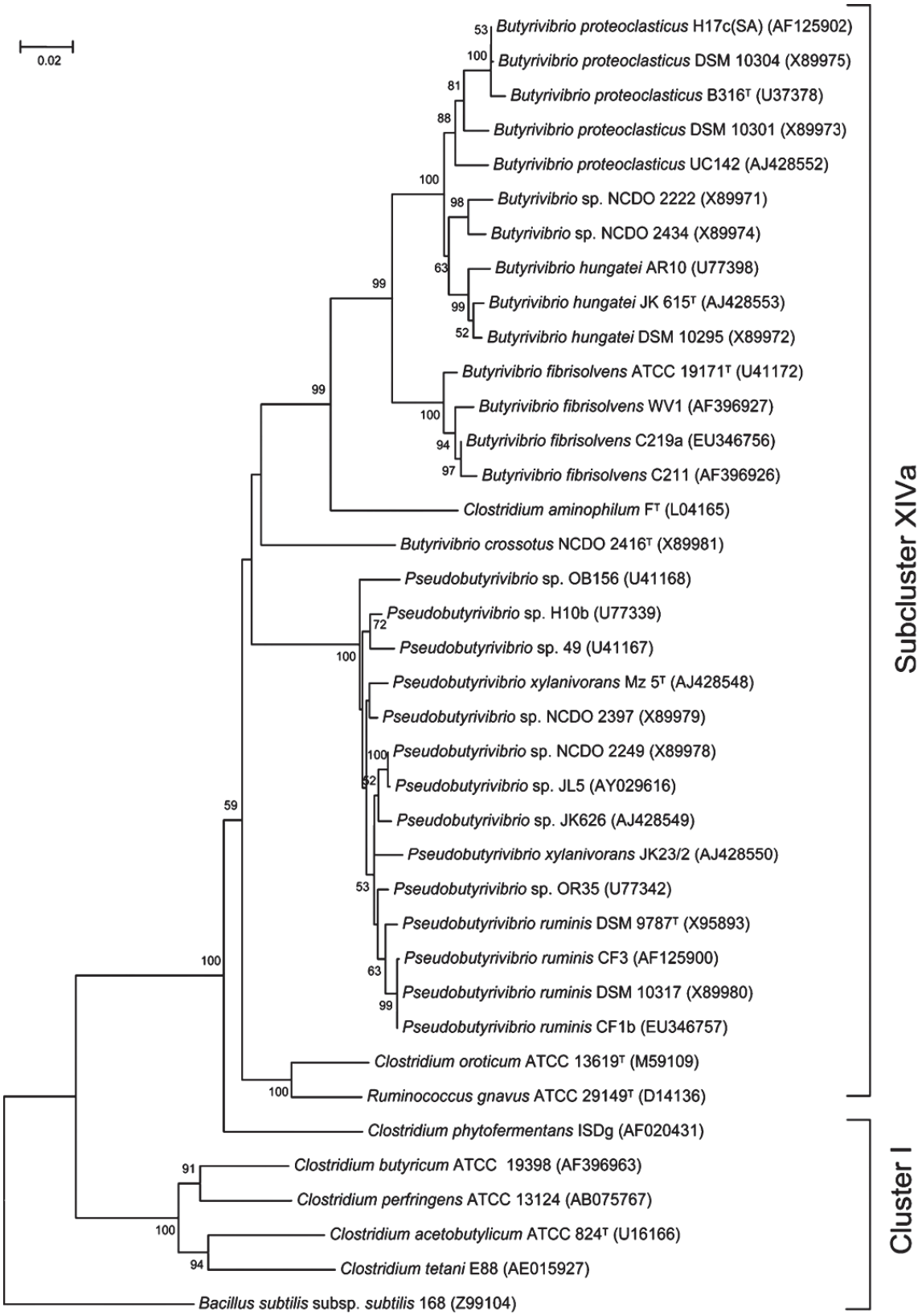


Figure 1.6. Phylogeny of Clostridial Cluster I and subcluster XIVa strains based on near full-length 16*S* rRNA gene sequences. *Bacillus subtilis* subsp. *subtilis* was used as the outgroup. Percentage bootstrap values were calculated from 10,000 replicates and are shown at nodes if >50 %. GenBank accession numbers of the 16*S* rRNA gene sequences are shown in parentheses. Bar, 0.02 nucleotide substitutions per site. ^{T}, indicates type strain. Figure is adapted from Moon *et al.* (2008) with permission.

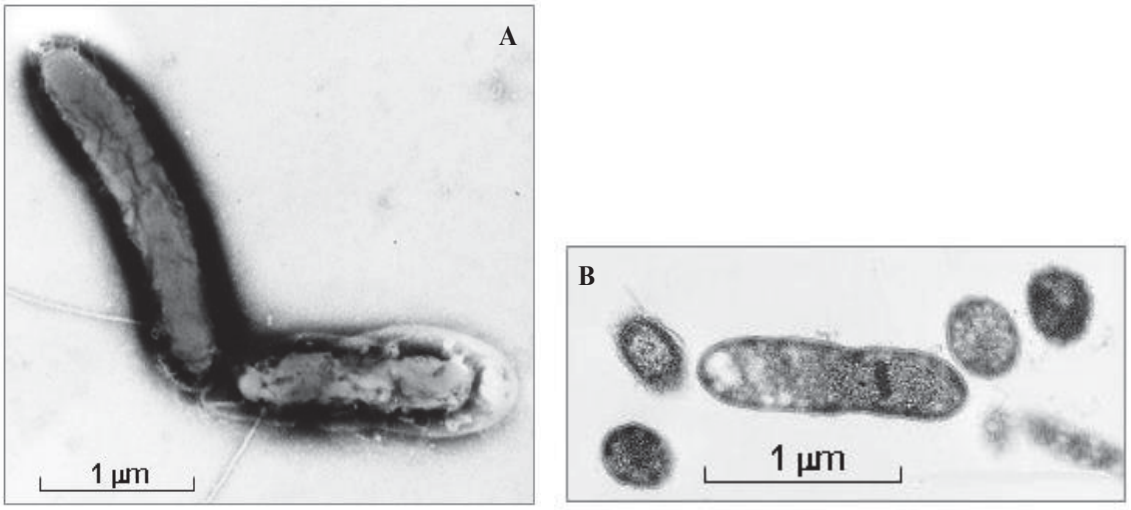


Figure 1.7. Ultrastructure of *B. proteoclasticus* B316 cells. Electron micrographs of negatively stained B316 cells showing the sub-polar flagellum (A), and thin-sectioned cells showing the Gram-positive cell wall structure (B). Figure from Attwood *et al.* (1996) with permission.

Table 1.5. *B. proteoclasticus* B316 genome features and genes involved in polysaccharide breakdown.

		Chromosome (BPc1)	Chromid (BPc2)	Megaplasmid (pCY360)	Megaplasmid (pCY186)	Total
General genome features	Size (bp)	3,554,804	302,358	361,399	186,325	4,404,886
	G+C%	40.21	40.04	38.95	38.1	39.33
	CDSs	2,939	251	425	198	3,813
	rRNAs	4	2	-	-	6
	tRNAs	47	2	-	1	50
Genes involved in polysaccharide breakdown	GHs [A]	98	15	-	-	113
	GTs [B]	120	0	0	1	121
	CEs [C]	19	-	-	-	19
	GH/CE*	4	-	-	-	4
	PLs [D]	1	1	-	-	2
	CBPs [E]	2	-	-	-	2
	Other CAPs [F]	83	6	-	-	89
	Transporter proteins	85	11	-	-	96

[A], glycoside hydrolases (GHs); [B], glycosyl transferase (GTs); [C], carbohydrate esterases (CEs); [D], polysaccharide lyases (PLs); [E], carbohydrate binding proteins (CBPs); [F], carbohydrate active proteins (CAPs). *, enzymes that contain multiple catalytic domains annotated as GH or CE. Table adapted from Kelly *et al.* (2010) with permission.

The chromid in B316 is unusual in that it is smaller than the megaplasmid pCY360 (Yeoman *et al.*, 2011). The B316 genome has 316 genes encoding enzymes that have a role in carbohydrate metabolism and transport. Almost 90% of these are located on the main chromosome, with the remainder found on the chromid (Table 1.5). B316 dedicates over 6% of its coding capacity to CAZymes (Kelly *et al.*, 2010) compared to an estimated 1-3% for prokaryote and eukaryote genomes in general (Davies *et al.*, 2005).

In silico analyses predicted that one third of these CAZymes were secreted and involved in the initial degradation of insoluble plant polysaccharides such as xylan and pectin, whereas complete degradation of xylo-oligomers is likely catalysed by various intracellular enzymes (Kelly *et al.*, 2010). The B316 genome encodes 113 GHs and 121 GTs, which are classified into 44 different CAZyme families. The GH families with the most representatives are the GH2, GH3, GH13, GH43, with 44 enzymes classified within these four families alone.

Recent studies investigating the proteome of B316 xylan-grown cells identified up-regulated secreted polysaccharidases containing single or combinations of catalytic activities: GH5 (endo-1, 4-β-glucanase, *cel5C*), GH10 (endo-1, 4-β-xylanase, *xyn10B*), GH43 (xylosidase/arabinofuranosidase, *xsa43J*), PL1 (pectate lyase, *pel1A*), CE8 and PL9 (pectin methylesterase, *pme8B*) (Dunne *et al.*, 2011; Bond *et al.*, 2012). Proteins with predicted functions in transport of oligosaccharides were significantly up-regulated, including 12 ATP-binding cassette (ABC) transporter associated proteins, substrate-binding proteins (SBPs). Based on these findings, a model was proposed that describes the initial xylan breakdown in the extracellular environment, followed by transport of hemicellulose-derived oligosaccharides into the cell by ABC transporter SBPs. The oligosaccharides are further degraded intracellularly by hemicellulases and the resulting monosaccharides metabolised (Dunne *et al.*, 2011; Bond *et al.*, 2012). A large portion of the B316 genes that encode lignocellulose degrading functions are organised into 36 polysaccharide utilisation loci (PULs; Figure 1.8, Circle 1). Genes that cluster within the B316 PULs account for approximately 7% of the total B316 genome, and include genes encoding ABC transport systems, transcriptional regulators such as AraC and LacI family proteins, environmental sensors including two-component histidine kinases and response regulator proteins, and genes involved in further oligosaccharide metabolism (Kelly *et al.*, 2010).

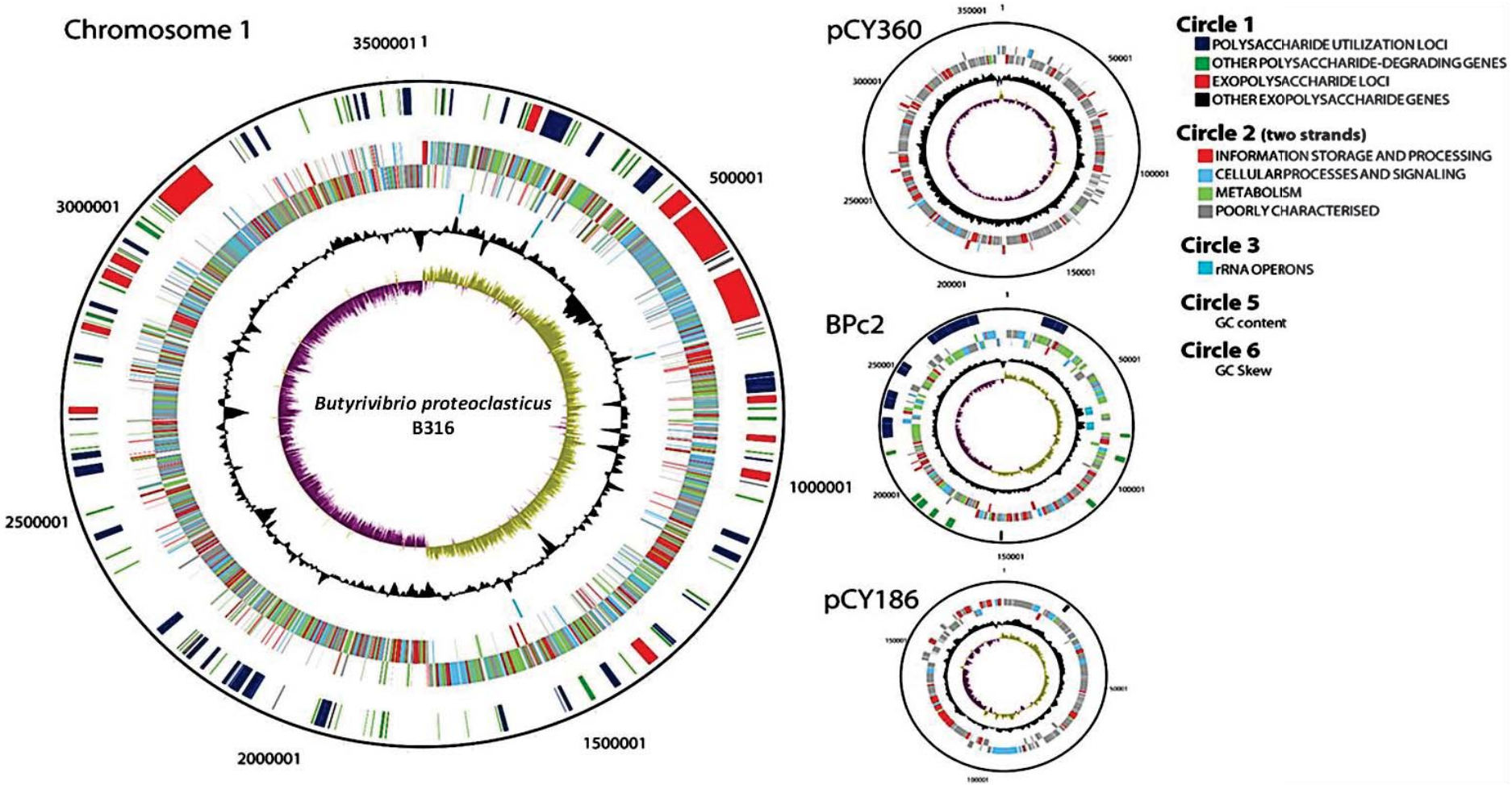


Figure 1.8. Genome atlas of *Butyrivibrio proteoclasticus* B316. The four replicons that make up the B316 genome and the location of the genes predicted to encode proteins involved in polysaccharide degradation are shown. The four rRNA operons on the main chromosome and the two on BPc2 are also shown. The colour coding of the genomic features in circle 2 represents the different COG categories. Figure adapted from Kelly *et al.* (2010) with permission.

1.8 Thesis aims

The aim of this project was to carry out a comprehensive comparative genomic and phenotypic analysis of the genus *Butyrivibrio* in order to understand how its members coexist, collaborate and/or compete in the same environmental niche. Representative *Butyrivibrio* strains derived from studies on the plant adherent fraction of rumen contents from NZ pasture-grazed dairy cows were selected (Noel, 2013) based on their phylogenetic relationships determined from 16*S* rRNA sequences and on their metabolic characteristics. These representative strains had their genomes sequenced as part of the Hungate 1000 project, and their draft genomes, along with genomes of recently cultured rumen *Butyrivibrio* strains (Palevich, 2011), were used to construct a *Butyrivibrio* pan-genome. Comparative genomics and bioinformatics of these microorganisms were used to identify the genes encoding enzymes responsible for the degradation of plant material. Because of their unusual genome arrangements, an investigation of *Butyrivibrio* plasmid, megaplasmid and chromid (secondary chromosome) composition was undertaken. Also, growth and gene expression studies were conducted to investigate the functions of genes involved in hemicellulose and pectin degradation. Specifically, the research questions of this PhD project were:

i) How does the composition and organisation of the genes encoding polysaccharide-degrading enzymes vary within the genomes of *Butyrivibrio* species?
ii) Is there conservation of genes that might imply particular importance in polysaccharide utilisation in this genus?
iii) Are there interspecies competitive strategies among *Butyrivibrio* strains?
iv) How are these different strategies encoded by the respective genomes?

The ultimate purpose of this study was to improve the knowledge of *Butyrivibrio* that inhabit the rumen. This is a significant group of rumen bacteria which play an important the role in plant fibre degradation in NZ ruminants. Determination and characterisation of their fibrolytic enzyme complements and strategies involved in fibre degradation is fundamental to understanding digestion, as well as maximising productivity in ruminant animals.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

The bacterial strains used in this work are listed in Table 2.1.

2.1.2 Buffers, reagents and solutions

Anaerobic glycerol solution (40% v/v)

Anaerobic glycerol solution is composed of 4.4 M glycerol (BDH, VWR International Ltd, Poole, UK) in water with the components of Table 2.2 added in the order listed. The solution was mixed, boiled for 5 min, cooled to room temperature under a continuous flow of O_2-free CO_2. Once cooled, L-cysteine-HCl (0.125 g) was added per 500 mL of the glycerol solution and transferred to serum bottles in 70 mL aliquots while being flushed with CO_2. The bottles were sealed with butyl rubber bungs and sealed with aluminium crimp caps before being autoclaved at 121 ºC for 20 min.

Table 2.2. Anaerobic glycerol solution.

Component	Volume or weight
dH_2O	130 mL
Salt solution A	85 mL
Glycerol	200 mL
Salt solution 2B	85 mL
Resazurin solution (1% w/v)	2 drops
$NaHCO_3$	2.5 g
L-cysteine-HCl	0.125 g

DNA-free water

Ultrapure water was collected from a Milli-Q Integral Water Purification System (EMD Millipore/Merck Millipore, Darmstadt, Germany) and sterilised by autoclaving, followed by microfiltration. The water was then irradiated with UV light (254 nm, 6 W) for 8 h.

Diethylpyrocarbonate (DEPC)-treated H2O

DEPC (0.1% v/v, Sigma-Aldrich, St. Louis, MO, USA) was added to dH_2O and autoclaved at 121 ºC for 20 min.

Table 2.1. Bacterial strains used in this thesis.

Species	Strain	Accession number *	Source/Reference
Butyrivibrio fibrisolvens	AB2020	NA	Noel (2013)
Butyrivibrio fibrisolvens	D1^T	DSM 3071	Bryant and Small (1956a)
Butyrivibrio fibrisolvens	FE2007	NA	Noel (2013)
Butyrivibrio fibrisolvens	MD2001	NA	Noel (2013)
Butyrivibrio fibrisolvens	ND3005	NA	Noel (2013)
Butyrivibrio fibrisolvens	WTE3004	NA	Noel (2013)
Butyrivibrio fibrisolvens	YRB2005	NA	Noel (2013)
Butyrivibrio hungatei	JK615^T	DSM 14810	Kopecny *et al.* (2003)
Butyrivibrio hungatei	MB2003	NA	Palevich (2011)
Butyrivibrio hungatei	NK4A153	NA	Noel (2013)
Butyrivibrio proteoclasticus	B316^T	DSM 14932	Attwood *et al.* (1996)
Butyrivibrio proteoclasticus	FD2007	NA	Noel (2013)
Butyrivibrio proteoclasticus	P6B7	NA	Noel (2013)
Butyrivibrio sp.	AC2005	NA	Noel (2013)
Butyrivibrio sp.	AD3002	NA	Noel (2013)
Butyrivibrio sp.	AE2005	NA	Noel (2013)
Butyrivibrio sp.	AE2015	NA	Noel (2013)
Butyrivibrio sp.	AE2032	NA	Noel (2013)
Butyrivibrio sp.	AE3003	NA	Noel (2013)
Butyrivibrio sp.	AE3004	NA	Noel (2013)
Butyrivibrio sp.	AE3006	NA	Noel (2013)
Butyrivibrio sp.	AE3009	NA	Noel (2013)
Butyrivibrio sp.	FC2001	NA	Noel (2013)
Butyrivibrio sp.	FCS006	NA	Noel (2013)
Butyrivibrio sp.	FCS014	NA	Noel (2013)
Butyrivibrio sp.	LB2008	NA	Noel (2013)
Butyrivibrio sp.	LC3010	NA	Noel (2013)
Butyrivibrio sp.	MB2005	NA	Noel (2013)
Butyrivibrio sp.	MC2013	NA	Noel (2013)
Butyrivibrio sp.	MC2021	NA	Noel (2013)
Butyrivibrio sp.	NC2002	NA	Noel (2013)
Butyrivibrio sp.	NC2007	NA	Noel (2013)
Butyrivibrio sp.	NC3005	NA	Noel (2013)
Butyrivibrio sp.	VCB2001	NA	Noel (2013)
Butyrivibrio sp.	VCB2006	NA	Noel (2013)
Butyrivibrio sp.	VCD2006	NA	Noel (2013)
Butyrivibrio sp.	WCD2001	NA	Noel (2013)
Butyrivibrio sp.	WCD3002	NA	Noel (2013)
Butyrivibrio sp.	WCE2006	NA	Noel (2013)
Butyrivibrio sp.	XBB1001	NA	Noel (2013)
Butyrivibrio sp.	XPD2002	NA	Noel (2013)
Butyrivibrio sp.	XPD2006	NA	Noel (2013)
Escherichia coli	DH5α	DSM 6897	Scheutz and Strockbine (2005)
Prevotella ruminicola	23	NA	Avguštin *et al.* (1994)
Pseudobutyrivibrio ruminis	AD2017	NA	Noel (2013)
Pseudobutyrivibrio ruminis	CF1b	NA	Noel (2013)
Pseudobutyrivibrio ruminis	HUN009	NA	Noel (2013)
Pseudobutyrivibrio sp.	LB2011	NA	Noel (2013)
Pseudobutyrivibrio sp.	MD2005	NA	Noel (2013)
Pseudobutyrivibrio xylanivorans	MA3014	NA	Palevich (2011)
Pseudobutyrivibrio xylanivorans	Mz5^T	DSM 14809	Kopecny *et al.* (2003)
Streptococcus bovis	2B	ATCC 33317	Iverson and Millis (1976)

* DSM: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany. NA, not associated with an accession number. ^{T}, indicates type strain.

dNTP solution

Deoxynucleotide triphosphates (dNTPs), dATP, dCTP, dGTP, and dTTP were supplied separately at 100 mM (Thermo Fisher Scientific Inc., Waltham, MA, USA), mixed in equimolar concentrations and diluted in dH_2O to make a 20 mM total dNTP stock solution. Stocks were stored at -20 ºC until required.

EC buffer

Ethylene diamine tetraacetic acid (EDTA) (100 mM), NaCl (1 M), *N*-Laurylsarcosine (35 mM, Sigma-Aldrich, St. Louis, MO, USA) and Trizma base (6 mM, Thermo Fisher Scientific Inc., Waltham, MA, USA) were prepared in dH_2O, adjusted to pH 7.6 with 1 M NaOH, then autoclaved at 121 ºC for 20 min.

EDTA-Sarkosyl solution

EDTA (0.5 M) and *N*-Laurylsarcosine (35 mM) were prepared in dH_2O and pH adjusted to pH 8.0 with 1 M NaOH, then autoclaved at 121 ºC for 20 min.

Mineral solution I (Salt solution 2B)

KH_2PO_4 (45 mM) and $KH_2PO_4.H_2O$ (60 mM), all supplied by VWR International Ltd, Poole, UK, were dissolved in dH_2O and autoclaved at 121 ºC for 20 min.

Mineral solution II

This solution contained $(NH_4)_2SO_4$ (45 mM), $CaCl_2.2H_2O$ (10 mM), $MgSO_4.7H_2O$ (10 mM), NaCl (200 mM) and KH_2PO_4 (45 mM). All ingredients (supplied by VWR International Ltd, Poole, UK) were dissolved in dH_2O and autoclaved at 121 ºC for 20 min.

Modified Karnovsky's fixative

Paraformaldehyde powder (VWR International Ltd, Poole, UK) was dissolved in 100 mL dH_2O and heated to 60 ºC with the drop wise addition of 1 M NaOH until the solution became transparent. Once cooled, Buffer Salts (2.51 g $Na_2HPO_4.12H_2O$ and 0.41 g KH_2PO_4) and Glutaraldehyde Solution (12 mL of 25% v/v glutaraldehyde in dH_2O) were added and the fixative was stored at -20 ºC.

Orange G loading dye (6× stock solution)

The Orange G loading dye contained 20 mM Tris-HCl (pH 7.6), 0.4% (w/v) 1-phenylazo-2-naphthol-6, 8-disulfonic acid disodium salt (Orange G sodium salt), 30% (v/v) glycerol, 60 mM EDTA (pH 8.0) and 0.1% (w/v) sodium dodecyl sulfate (SDS). Dye aliquots were stored short-term at 4 ºC or long-term at -20 ºC.

Phosphate-buffered saline (PBS) solution (10× stock solution)

To make 10× concentrated stock solution, $CaCl_2$ (0.01 g) and $MgCl_2$ (0.01 g) were dissolved in 100 mL dH_2O and added to a solution containing NaCl (8.00 g), KCl (0.20 g), Na_2HPO_4 (1.44 g) and KH_2PO_4 (0.24 g) dissolved in 1.0 L dH_2O (Sambrook *et al.*, 1989). The combined solution was adjusted to pH 7.4, mixed well and autoclaved at 121 ºC for 20 min (Neinhuis and Edelmann, 1996).

Paraformaldehyde (PFA) solution (4% w/v)

PFA powder (12 g, VWR International Ltd, Poole, UK) was dissolved in 195 mL dH_2O, heated to 60 ºC with the drop-wise addition of 1 M NaOH until the solution became transparent (Neinhuis and Edelmann, 1996). A volume of 99 mL of 3× PBS solution was added and once cooled, the pH was adjusted to 7.2, it was then filter sterilised and stored at -20 ºC.

Saline-EDTA solution

Saline-EDTA solution was prepared by adding EDTA powder (VWR International Ltd, Poole, UK) to a concentration of 10 mM, and NaCl to 150 mM to dH_2O. The pH was adjusted to 8.0 and solution was autoclaved at 121 ºC for 20 min.

Salt solution A

To make Salt solution A, NaCl (6.0 g), $(NH_4)_2SO_4$ (1.5 g), KH_2PO_4 (3.0 g), $CaCl_2.H_2O$ (0.79 g) and $MgSO_4.7H_2O$ (1.2 g) were dissolved in 1.0 L of dH_2O.

Salt solution 2B

To make salt solution B, either K_2HPO_4 (6.0 g) or $K_2HPO_4.H_2O$ (7.9 g) were added to 1.0 L of dH_2O.

SDS solution (20% w/v)

SDS (200 g) was dissolved in dH_2O, pH adjusted to 7.2 and volume adjusted to 1.0 L, followed by filter sterilization into a 1.0 L sterile Schott bottle.

Substrate solutions

Stock solutions (20% w/v) were made for 32 sugar substrates. The appropriate amount of sugars were dissolved in 20 mL of dH_2O in N_2-filled 100 mL Schott bottles and autoclaved at 121 ºC for 20 min. For insoluble polysaccharide substrates, 4 g of each substrate was added to 20 mL of dH_2O in N_2-filled 100 mL Schott bottles, mixed by heating with stirring and autoclaved at 121 ºC for 20 min. The D (+) enantiomers of all substrates were used and were as follows: amygdalin, arabinose, cellobiose, cellulose, dextrin, esculin, fructose, galactose, glucose, glycerol, glycogen, myo-inositol, inulin, lactose, maltose, mannitol, mannose, melezitose, melibiose, pectin, raffinose, rhamnose, ribose, rutin, salicin, sorbitol, starch, sucrose, trehalose, xylan, xylitol and xylose. The solutions were stored in the dark at room temperature until use (Leedle and Hespell, 1980).

Tris-acetate-EDTA (TAE) buffer (50× stock solution)

A solution containing glacial acetic acid (950 mM, VWR International Ltd, Poole, UK), EDTA (50 mM) and Trizma base (2 M, Thermo Fisher Scientific Inc., Waltham, MA, USA) was prepared in dH_2O, pH adjusted to 8.0 with 1 M NaOH and autoclaved at 121 ºC for 20 min. A working solution (1×) was made by diluting the stock solution 1:50 in dH_2O.

Tris-borate-EDTA (TBE) buffer (5× stock solution)

A solution containing boric acid (445 mM, VWR International Ltd, Poole, UK), EDTA (10 mM) and Trizma base (445 mM) was prepared in dH_2O and autoclaved at 121 ºC for 20 min. A working solution (0.5×) was made by diluting the stock solution 1:10 in dH_2O.

Tris-EDTA (TE) buffers

Three types of TE buffer were used: TE buffer 10/0.1 (10 mM Trizma base and 0.1 mM EDTA), TE buffer 10/1 (10 mM Trizma base and 1 mM EDTA) and TE buffer

10/100 (10 mM Trizma base and 0.1 M EDTA). The buffer components were dissolved in dH_2O, pH adjusted to 8.0 and autoclaved at 121 ºC for 20 min.

Tris-EDTA-sucrose (TES) buffer

EDTA (1 mM), sucrose (250 mM) and Trizma base (10 mM) were dissolved in dH_2O, pH adjusted to 7.5 and autoclaved at 121 ºC for 20 min.

Volatile fatty acid (VFA) solution

Butyric acid (114 mM), *iso*-butyric acid (28 mM) , *iso*-valeric acid (24 mM), *n*-valeric acid (24 mM) (all supplied by Sigma-Aldrich, St. Louis, MO, USA), D, L-2-methyl butyric acid (24 mM), propionic acid (202 mM) (both supplied by Merck, Darmstadt, Germany) and acetic acid (710 mM) were prepared in dH_2O, pH adjusted to 7.5 with 1 M NaOH and the mixture was stored at -20 ºC (Attwood *et al.*, 1998).

Wash solution for Pulse-Field Gel-Electrophoresis

NaCl (2 M) and Trizma base (20 mM) were dissolved in dH_2O, pH adjusted to 7.6 with 6 M HCl and autoclaved at 121 ºC for 20 min.

2.1.3 Media components

Rumen fluid collection

Rumen contents for preparation of rumen fluid were collected from a rumen-fistulated Holstein cow grazing on a rye-grass/clover hay (tag number 723) which had been fasting for 14-18 h, at AgResearch Ltd., Grasslands, NZ. The rumen contents were filtered through 2 layers of cheesecloth, the liquid was collected and centrifuged twice at 20,000 × *g* at 4 ºC for 20 min, with the pelleted microbial and plant material being discarded each time. The resulting rumen fluid supernatant was stored frozen at -20 ºC until required.

Clarified rumen fluid

Frozen rumen fluid was thawed and centrifuged at 20,000 × *g* at 4 ºC for 15 min. The decanted liquid was heated in a microwave until the point of boiling and then allowed to cool under a continuous flow of oxygen-free nitrogen gas (N_2) bubbled through the liquid, then autoclaved under O_2-free N_2 at 121 ºC for 15 min in serum vials closed with

a butyl rubber stopper and crimp-sealed. The autoclaved liquid was allowed to cool and with stirring, 1.63 g of $MgCl_2.6H_2O$ and 1.18 g of $CaCl_2.2H_2O$ per 100 mL of rumen fluid was added. A heavy precipitate formed which was removed by centrifugation at 20,000 × *g* at 4 ºC for 20 min. The resulting supernatant was collected and represents the Clarified Rumen Fluid.

No substrate-rumen fluid-vitamins (NoSubRFV)

NoSubRFV was prepared by adding 2 g of yeast extract to 100 mL of Clarified Rumen Fluid and bubbling with O_2-free N_2 for 15 min. The mixture was transferred to an O_2-free N_2-flushed, sterile, serum bottle through a 0.22 µm pore size filter (Millex, EMD/Millipore Merck, Darmstadt, Germany) using a 60 mL syringe and 21 gauge needle (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). Two mL of Vitamin Solution was added per 100 mL of mixture. To achieve final concentrations of 5% (v/v) rumen fluid, 4 mM each of Ca^{2+} and Mg^{2+}, and 1 g of yeast extract per litre, 0.5 mL of NoSubRFV was added to 9.5 mL of medium.

Rumen fluid-vitamin mix (GenRFV and 2× GenRFV)

GenRFV is composed of the components listed in Table 2.3 which were added to 100 mL of Clarified Rumen Fluid. The mixture was stirred and O_2-free N_2 gas was bubbled through the mixture for 15 min. The mixture was transferred to an O_2-free N_2-flushed, sterile, serum bottle through a 0.22 µm pore size filter using a syringe and needle. For each 100 mL of GenRFV, 2 mL of Vitamin Solution (see below) was added *via* syringe and needle. To achieve the final desired substrate and growth factor concentrations, 0.5 mL of GenRFV was added per 9.5 mL of medium. The final concentrations of each GenRFV component in the media were as follows: 5% (v/v) rumen fluid, 0.5 mM cellobiose, 1 mM each of glucose, xylose, and arabinose, 5 mM lactate, 4 mM each of Ca^{2+} and Mg^{2+}, and 1 g each of casamino acids, Bacto-Peptone, and yeast extract per L.

Selenite-tungstate solution

A solution containing NaOH (12.5 mM), $Na_2SeO_3 \cdot 5H_2O$ (0.01 mM) and $Na_2WO_4 \cdot 2H_2O$ (0.01 mM) was prepared in 1.0 L dH_2O and autoclaved at 121 ºC for 20 min.

Table 2.3. GenRFV.

Chemical	Volume per L	Final concentration [B]
D-glucose	0.36 g	0.02 M
D-cellobiose	0.34 g	0.01 M
D-xylose	0.30 g	0.02 M
L-arabinose	0.30 g	0.02 M
Na L-lactate syrup	0.88 mL	0.10 M
Casamino acids	2 g	2% (w/v)
Bacto-Peptone	2 g	2% (w/v)
Yeast extract	2 g	2% (w/v)
Rumen fluid	100 mL	N/A
dH_2O	950 mL	N/A
Vitamin 10 concentrate solution [A]	2 mL	N/A

[A], added to prepared anoxic mixture of all other components listed above. [B], 2× GenRFV was made by the same procedure as GenRFV, except that the 2× GenRFV contained two-fold final concentration of D-glucose (2.0 mM), D-cellobiose (1.0 mM), D-xylose (2.0 mM) and L-arabinose (2.0 mM).

Trace element solution (SL10)

The components listed in Table 2.4 were added to a final volume of 1.0 L dH_2O and the solution was mixed well. The nitrilotriacetic acid was dissolved in dH_2O and the pH adjusted to 6.5 using 1 M KOH before adding to the remaining solution. The mixture was autoclaved at 121 ºC for 20 min and stored at 4 ºC (Widdel *et al.*, 1983).

Table 2.4. Trace element solution.

Chemical	Final concentration
$KAl(SO_4)_2.12H_2O$	42 µM
$B(OH)_3$	162 µM
$CaCl_2.2H_2O$	680 µM
$CoSO_4.7H_2O$	640 µM
$CuSO_4.5H_2O$	40 µM
$FeSO_4.7H_2O$	360 µM
$MnSO_4.2H_2O$	2.24 mM
$MgSO_4.7H_2O$	12 mM
NaCl	17 mM
$NiCl_2.6H_2O$	100 µM
Nitrilotriacetic acid	8 mM
$Na_2MoO_4.2H_2O$	40 µM
$Na_2SeO_3.5H_2O$	1 µM
$ZnSO_4.7H_2O$	625 µM

Vitamin 10 concentrate solution

To make Vitamin 10 concentrate solution, dH_2O was boiled and allowed to cool to room temperature under O_2-free N_2. All the components listed in Table 2.5 (VWR International Ltd, Poole, UK) were added and O_2-free N_2 was allowed to bubble through the solution for a further 20 min before the solution was filter sterilised (0.22 μm pore size) into Hungate tubes in 10 mL aliquots. The tubes were wrapped in aluminium foil to protect against light and stored either at -20 ºC, or kept at 4 ºC as working stocks.

Table 2.5. Vitamin solution.

Chemical	Final concentration
D-(+)-biotin	8 μM
D-Ca-pantothenate	23 μM
Folic acid	4.5 μM
Lipoic acid	24 μM
Nicotinic acid	40 μM
p-Aminobenzoic acid	36 μM
Pyridoxine-HCl	50 μM
Riboflavin	13 μM
Thiamine-HCl dihydrate	15 μM
Vitamin B_{12}	75 nM

2.1.4 Media

Basal medium plus yeast extract (BY) and RM02 medium

BY and RM02 medium were prepared from the ingredients listed in Table 2.6 as follows: all components, except L-Cysteine-HCl and Vitamin Solution, were dissolved in 1.0 L of dH_2O, and boiled to remove dissolved O_2. The Vitamin Solution, NoSubRFV or GenRFV (or 2× GenRFV) containing the Clarified Rumen Fluid were injected into each Hungate tube just prior to use. Resazurin solution (1% w/v in dH_2O) was used to indicate that the solution was anoxic as it changed colour from red to colourless. The boiled, colourless solution was allowed to cool on ice to approximately room temperature, while gassing with O_2-free CO_2. L-Cysteine-HCl was added and the resulting solution was distributed in 9.5 mL aliquots into CO_2-flushed 10 mL Hungate tubes. The tubes were sealed with butyl rubber stoppers and autoclaved at 121 ºC for 20 min. Solid RM02 and BY media were prepared using the previously described methods for the liquid RM02 and BY media, except that bacteriological agar was added at 1.5% (w/v) to the media, and autoclaved at 121 ºC for 20 min prior to dispensing into Petri plates (Thermo Fisher

Scientific Inc., Waltham, MA, USA) inside a Coy Anaerobic Chamber (Coy Laboratory Products, Grass Lake, MI, USA). Once solidified and dried, the plates were used for streaking bacterial cultures for isolating single colonies. A 0.3% (w/v) soft agar form of the media were also prepared in Hungate tubes and used as agar stabs for motility assays.

Table 2.6. BY and RM02 media.

Medium component	Volume per L *	
	BY medium	RM02 medium
KCl	N/A	1.5 g
NaCl	1.0 g	N/A
KH_2PO_4	0.5 g	1.4 g
$(NH_4)_2SO_4$	0.25 g	0.6 g
$CaCl.2H_2O$	0.13 g	N/A
$MgSO_4.7H_2O$	0.20 g	N/A
K_2HPO_4	1.0 g	N/A
Clarified Rumen Fluid	300 mL	N/A
dH_2O	360 mL	950 mL
$NaHCO_3$	5.0 g	4.2 g
Resazurin (0.1% w/v)	10 drops	4 drops
L-cysteine-HCl	0.5 g	0.5 g
Trace Element Solution	1 mL	1 mL
Selenite-Tungstate Solution	N/A	1 mL
Bacto Yeast extract	1.0 g	N/A

* N/A refers to components that are not included in the media mixture.

Xylan and pectin-containing RM02 medium

Xylan (poly (β-D-xylopyranose [1→4])) and pectin (poly-D-galacturonic acid methyl ester) in RM02 medium was prepared using the method described above for the standard RM02 medium. To achieve the final desired substrate and growth factor concentrations, 1.75 mL of 20% (w/v) xylan or pectin was added to achieve the final substrate concentration of 0.5% (v/v), along with Vitamin Solution and 3.5 mL of NoSubRFV per 70 mL of medium in serum bottles. Xylan was sourced from oat spelts (CAS Number 9014-63-5, X0627, Sigma-Aldrich, St. Louis, MO, USA) and pectin was isolated from apple (CAS Number 9000-69-5, 76282, Sigma-Aldrich, St. Louis, MO, USA).

Luria-Bertani (LB) medium

LB liquid medium contains 1.0% (w/v) Tryptone, 0.5% (w/v) Yeast extract 1.0% (w/v) NaCl dissolved in dH_2O with pH adjusted to 7.0. LB solid medium was made by

the addition of 1.5% (w/v) bacteriological agar before autoclaving. The autoclaved bottles containing LB solution were allowed to cool to 55 ºC in a water bath before adding antibiotics as required and pouring the medium into Petri plates.

2.1.5 Enzymes, buffers and reagents

DNA Degradase

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

Lysozyme solution

Lysozyme (Roche Diagnostics, Risch-Rotkreuz, Switzerland) was prepared as a 25 mg/mL stock solution by dissolving 125 mg lysozyme in 5 mL TE (10/1) Buffer and stored at -20 ºC.

Oligonucleotides

All custom oligonucleotides were synthesised by Integrated DNA Technologies (IDT, Coralville, IA, USA) and supplied desalted and lysophilised at 50 nM concentration. Oligonucleotides were reconstituted in DNA-free water at a concentration of 100 µM and stored at -20 ºC until required. Stock solutions of oligonucleotides at 100 pmol/µL were made with the addition of DNA-free water and diluted to working concentrations of 10 pmol/µL when required. The oligonucleotides used to amplify 16*S* rRNA genes are detailed in Table 2.7. The primer design function within the *Gap4* program of the Staden package (Staden *et al.*, 1999) was used to design custom primers for gap closure of the *B. hungatei* MB2003 genome project. The details and sequences of the custom *Taq*Man primer/probe assays supplied *via* IDT for the *B. hungatei* MB2003 and *B. proteoclasticus* B316 co-culture experiment are described in Table 2.8.

Proteinase K

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

Table 2.7. 16*S* rRNA gene primers.

Primer *	Sequence	Reference
fD1	5'-GAGTTTGATCMTGGCTCAG-3'	Duthoit *et al.* (2003)
rD1	5'-AAGGAGGTGATCCARCCG-3'	Escalante *et al.* (2001)
1492r	5'-GGYTACCTTGTTACGACTT-3'	Lane (1991)
1100r	5'-GGGTTGCGCTCGTTG-3'	Lane (1991)
1382r	5'-CGGTGTGTRCAAGGCCC-3'	Dorsch and Stackebrandt (1992)
806r	5'-GGACTACVSGGGTATCTAAT-3'	Bergmann *et al.* (2011)
514r	5'-CCGCGGCKGCTGGCAC-3'	Lane (1991)
515f	5'-GTGCCAGCMGCCGCGGTAA-3'	Bergmann *et al.* (2011)

* f, forward/sense primer; r, reverse/antisense primer.

Table 2.8. *Taq*Man primers/probes.

Primer/Probe	Specificity	Gene annotation	Locus tag	Amplicon size (bp)	Sequence
buk_Forward	B316 and MB2003	Butyrate kinase, *buk*	bpr_I2323	90	5'-ACCTTGGAACCAACGATATGAG-3'
buk_Reverse					5'-GGTAAATGAAAGCGTCACGAAC-3'
buk_Probe					5'-TGTGTGATGAGGGCAACGAGAAGG-3'
CBD_Forward	B316	Cell wall binding domain-containing protein, *CBD*	bpr_I0264	86	5'-CGCTACAGAGGGAACTGAAATAG-3'
CBD_Reverse					5'-CCTCAAGGCTCTCATCAACTAC-3'
CBD_Probe					5'-CTGTGGAAGCATCAAGCGCAGC-3'
Arf51C_Forward	MB2003	α-L-arabinofuranosidase, *arf51C*	bhn_I1532	88	5'-GCAAAGGTATGTTTGGACTGTT-3'
Arf51C_Reverse					5'-AAACTACGGTTCTCAAGCATTTC-3'
Arf51C_Probe					5'-TGCATGAAGACCACCGTCAAGACC-3'

Restriction Endonucleases (REs)

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

Ribonuclease A (RNaseA)

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

2.1.6 General laboratory equipment

Samples were mixed with a PV1 vortex (Grant Instruments, Cambridge, UK). Heating and incubation operations were carried out in a recirculating water-bath (Grant Instruments, Cambridge, UK) or an Accublock digital dry bath (Labnet International, Edison, NJ, USA). Autoclaving was done in a high pressure steam steriliser ES-315 (Tomy, Tokyo, Japan). The pH meter used in this work was a PHM62 pH Meter (Radiometer, Copenhagen, Denmark). N_2, CO_2 and H_2 gas were supplied by BOC gas (Palmerston North, NZ). Work with anaerobic media plates was carried out in a Coy Anaerobic Chamber (Coy Laboratory Products, Grass Lake, MI, USA) containing a 95% CO_2: 5% H_2 atmosphere. The various centrifuges and gel migration standards used in this thesis are listed in Tables 2.9 and 2.10, respectively.

2.1.7 Bioinformatic resources and software

Bioinformatic resources and software used in this thesis are listed in Table 2.11.

Table 2.9. Centrifuge specifications and suppliers.

Brand	Vessels	Max RCF × *g* (rotor used)	Temp control	Supplier
Spectrafuge Mini Centrifuge	0.6 mL tubes	2,000	No	Labnet (Edison, NJ, USA)
Minispin personal microcentrifuge	1.5 mL and 2.0 mL tubes	16,000	No	Eppendorf (Hamburg, Germany)
MicroCL 17 Centrifuge	1.5 mL and 2.0 mL tubes	17,000	No	Thermo Fisher Scientific Inc., (Waltham, MA, USA)
Biofuge Fresco	1.5 mL tubes	16,000	Yes	Heraeus Holding GmbH (Hanau, Germany)
Sigma 3-18K	15 mL, 50 mL falcon tubes Hungate tubes 96 well plates	4,248 (11180/13190) 3,500 2,717 (11240)	Yes	Sigma-Aldrich (St. Louis, MO, USA)
Sorvall®Evolution RC	50 mL Oakridge tubes 250 mL Nalgene Tubes	48,000 (SS34) 27,500 (SLC1500)	Yes	Thermo Fisher Scientific Inc., (Waltham, MA, USA)

Table 2.10. Molecular weight size markers.

Standard	Range	Supplier
TrackIt 100 bp DNA ladder [A]	100 bp to 1.5 Kb	Thermo Fisher Scientific Inc., (Waltham, MA, USA)
1 Kb Plus DNA ladder [A]	100 bp to 12 Kb	Thermo Fisher Scientific Inc., (Waltham, MA, USA)
Lambda DNA/Hind III ladder [A]	100 bp to 24 Kb	Thermo Fisher Scientific Inc., (Waltham, MA, USA)
RNA ladder [A]	500 bp to 10 Kb	Thermo Fisher Scientific Inc., (Waltham, MA, USA)
Lambda DNA ladder [B]	48.5 Kb to 727.5 Kb	New England Biolabs (Ipswich, MA, USA)
Mid range marker I [B]	15 Kb to 242.5 Kb	New England Biolabs (Ipswich, MA, USA)
Low range marker [B]	2.03 Kb to 194 Kb	New England Biolabs (Ipswich, MA, USA)

Molecular weight size markers used for [A], agarose gel electrophoresis and [B], PFGE: pulse-field gel-electrophoresis.

Table 2.11. Bioinformatic resources and software used in this thesis.

Software	Application	Source	Reference(s)
Artemis	Genome sequence viewing and analysis	http://www.sanger.ac.uk	Rutherford *et al.* (2000)
BioNumerics 7.5	Comparative analysis of PFGE profiles	http://www.applied-maths.com	http://www.applied-maths.com
Bowtie 2	Alignment of short sequence reads	http://bowtie-bio.sourceforge.net	Langmead and Salzberg (2012)
ClustalW	Global sequence alignment	ftp://ftp.ebi.ac.uk	Thompson *et al.* (2003)
CMG-biotools	Comparative Microbial Genomics analysis	http://www.cbs.dtu.dk	Vesth *et al.* (2013)
dbCAN 3.0	HMM-based database for automated CAZyme annotation	http://csbl.bmb.uga.edu	Yin *et al.* (2012)
DNAPlotter	Circular and linear interactive genome visualization	http://www.sanger.ac.uk	Carver *et al.* (2009)
Galaxy	Web-based comparative genomics and bioinformatics platform	https://galaxyproject.org	Goecks *et al.* (2010)
GAMOLA	Automated genome sequence annotation	In-house software supplier	Altermann and Klaenhammer (2003)
Geneious 8.1	DNA sequence display and analysis	http://www.geneious.com	Kearse *et al.* (2012)
IMG/ER	Functional annotation and expert review of private and publicly available genomes	http://img.jgi.doe.gov	Markowitz *et al.* (2013)
MEGA6	Sequence alignment and phylogenetic analysis	http://www.megasoftware.net	Tamura *et al.* (2013)
MUMmer	DNA and protein sequence alignment	http://mummer.sourceforge.net	Delcher *et al.* (1999)
Newbler	*De novo* assembly of the 454 genome sequence	http://www.454.com	http://www.454.com/
PHRED	Base calling in long sequence reads	http://www.phrap.com	Ewing *et al.* (1998)
Staden package	DNA sequence assembly management	http://sourceforge.net	Staden *et al.* (1999)
Rockhopper	System for analysis of bacterial RNA-seq data	http://cs.wellesley.edu	McClure *et al.* (2013)
RStudio	Integrated development environment (IDE) platform for R	https://www.rstudio.com	Ihaka and Gentleman (1996)
Velvet	Manipulate *de Bruijn* graphs for genomic short read sequence assembly	https://www.ebi.ac.uk	Zerbino and Birney (2008)

2.2 Methods

2.2.1 Origin and selection of *Butyrivibrio* isolates

The initial candidate *Butyrivibrio* cultures for this study were chosen from a large isolation experiment that cultivated bacteria from the plant-adherent fraction of rumen contents obtained from fistulated Friesian dairy cows (production cattle) grazing *ad libitum* on a rye-grass/clover pasture (Noel, 2013). In this work, dilutions of rumen contents were incubated in a variety of media, resulting in 828 isolates identified by partial (V1-V3 region) 16*S* rRNA gene sequencing as being mono-cultures, of which 219 were assigned to the genus *Butyrivibrio* (Noel, 2013). Taxonomic analyses based on the V1-V3 regions and alignment with the SILVA database *via* ARB (Pruesse *et al.*, 2007), followed by MEGA6 (Tamura *et al.*, 2013) and Geneious (Kearse *et al.*, 2012) software packages, grouped the 219 *Butyrivibrio* isolates into 18 separate clusters. The 30 *Butyrivibrio* cultures selected for phenotypic and genotypic characterisation are representatives of each of the different cluster groups selected for this study. The isolate name designations correspond to the month and year of the sampling, and are outlined in Table 2.12.

Table 2.12. Sampling times and isolate nomenclature.

	Sampling date				
	May 2009	August 2009	November 2009	February 2010	May 2010
Isolate code *	M	A and YR	N, XP and XB	F, WC and WT	L and VC

* Codes refer to the first one or two letters of the isolate names, followed by a letter (A to E) referring to one of the five different animals sampled.

2.2.2 Bacterial culture revival and growth

All *Butyrivibrio* and *Pseudobutyrivibrio* strains were cultured anaerobically in RM02 or BY media at 39 ºC unless otherwise specified. Where possible, the number of serial sub-culturing was reduced to a minimum before re-freezing of stocks (-85 ºC), to avoid *in-vitro* culture-biased evolution (Papadopoulos *et al.*, 1999). For bacterial growth curves, fresh overnight cultures were used to inoculate (0.5 mL inoculum) Hungate tubes containing 9.5 mL of fresh, pre-warmed, RM02 medium, with 0.5% (w/v) cellobiose as the substrate. Uninoculated RM02 medium served as a blank. Optical density (OD) readings were taken using an Ultrospec 1100 Pro Spectrophotometer (GE Healthcare, Little Chalfont, UK) at 600 nm. OD_{600nm} readings were measured at regular 2 h intervals

until stationary phase, and the initial reading taken after inoculation was subtracted to calculate the change in OD. Mid-log phase growth on cellobiose substrate was consistently determined to occur at OD_{600nm} = 0.4, after a 12 h growth period (Palevich, 2011).

2.2.3 Wet mounts and Gram staining

Wet mounts were made by placing ~25 µL of culture onto a glass microscope slide using a sterile 1 mL syringe and needle, followed by a coverslip laid over the culture. To Gram stain a bacterial culture, 100 µL of culture was spread on a glass microscope slide, left to air dry and then was heat fixed. 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) safranine for 30 sec. Slides were gently rinsed with tap water between each step. Microscopic verification of cultures after wet mount or Gram staining were examined at 100× magnification (with oil) under phase contrast or direct illumination using a Leica DM2500 microscope (Bio-Strategy, Auckland, NZ). The images produced were captured digitally using a Leica application suite software (Leica Microsystems, Wetzlar, Germany).

2.2.4 Long-term culture storage

Once culture purity was verified by microscopy and Gram-staining, sub-cultures were streaked and grown on BY agar plates and incubated under anaerobic conditions until single colonies appeared. Colony morphologies were recorded and single colonies from each culture were picked and sub-cultured twice sequentially into pre-warmed RM02 medium containing 0.5% (w/v) cellobiose, to ensure that colony morphologies were consistent. Culture purity and morphology were verified *via* wet mounts and Gram staining after each sub-culture. For long term storage, cultures were inoculated and grown overnight in 10 mL Hungate tubes containing 5 mL RM02 medium, 0.5 mL of culture inoculum and 0.5 mL of GenRFV. The cultures (6 mL) were supplemented with 3 mL of 40% glycerol solution (v/v) then stored at -85 ºC.

2.2.5 Agarose gel-electrophoresis

UltraPure™ Agarose (Invitrogen, Life Technologies, Carlsbad, CA, USA), (1% w/v) gels were made in 1× TAE buffer containing 1× final concentration of SYBR®

nucleic acid dye (Thermo Fisher Scientific Inc., Waltham, MA, USA). Gels were cast in stands provided by Bio-Rad (Hercules, CA, USA), submerged in 1× TAE buffer in either a Bio-Rad Sub-Cell® GT Agarose Gel Electrophoresis System (Bio-Rad, Hercules, CA, USA) or the 6MGel super 120 gel tank (BioKeystone, Vancouver, WA, USA). A 5 µL aliquot of sample was re-suspended with 1 µL of 6× Orange G loading dye and loaded into each well. An appropriate molecular weight-size marker (Table 2.10), was also loaded in the first lane. Gels were run at 90 V for 30 to 45 min and visualised using UV trans-illumination. Gels were photographed using a Nikon D700 camera with Kodak Gel Logic 200 Imaging System (Eastman Kodak, Rochester, NY, USA).

2.2.6 DNA extractions

Several methods for DNA extractions were used in this study. For all extractions, UltraPure™ buffer-saturated phenol (pH 7.4) from Thermo Fisher Scientific Inc. (Waltham, MA, USA), phenol: chloroform: isoamyl alcohol (25:24:1, v/v/v, pH 4.5, Ambion, Thermo Fisher Scientific Inc., Waltham, MA, USA) and chloroform from BDH (VWR International Ltd, Poole, UK) were used where specified.

Standard phenol-chloroform DNA extraction

The following phenol-chloroform method for genomic DNA extraction described by Saito and Miura (1963), was used for all molecular experiments unless otherwise specified. Briefly, 10 mL cultures were harvested at mid-exponential phase (OD_{600}= ~0.4) by centrifugation at 5,000 × *g* for 5 min. The cell pellet was resuspended in 500 µL Saline-EDTA solution and harvested by centrifugation at 5,000 × *g* for 5 min. The cell pellet was resuspended in 500 µL of Saline-EDTA Solution containing 1 mg/mL lysozyme and 20 µg/mL RNaseA. This solution was incubated for >1 hr at 37 ºC to allow the lysozyme to facilitate cell lysis, and the RNaseA to subsequently degrade RNA. SDS solution (27 µL), along with 11 µL of Proteinase K solution was added to give a final concentration of 1% (w/v) and 200 µg/mL respectively. The solution was incubated for 1.5 hr at 60 ºC to allow the Proteinase K to degrade cellular protein and inactivate nucleases. After incubation, 200 µL of TE Buffer (10/1) was added and one volume of room temperature, UltraPure™ buffer-saturated phenol was added to the DNA extraction. The sample was mixed by several inversions before precipitated protein and other organic material were fractionated by centrifugation at 13,000 × *g* for 5 min. The supernatant was carefully transferred to a fresh Eppendorf tube and mixed with an equal volume of phenol: chloform: isoamyl

alcohol mixture. The mixture was inverted several times before the phases were separated by centrifugation at 13,000 × *g* for 5 min. The supernatant was transferred to a fresh Eppendorf tube and this extraction step was repeated with an equal volume of chloroform to remove residual phenol.

The supernatant was transferred to a fresh Eppendorf tube and 0.1 volumes of 5 M ammonium acetate and 3 volumes of absolute ethanol were added and mixed. The mixture was left for 16-24 hr at -20 ºC then centrifuged at 16,000 × *g* for 30 min at 4 ºC. The supernatant was carefully removed using a pipette to avoid disturbing the DNA pellet. The pellet was washed in 200 µL of 70% (v/v) ethanol and centrifuged again at 16,000 × *g* for 15 min at 4 ºC. The supernatant was carefully removed using a pipette and allowed to dry at 37 ºC for 30 min. The DNA pellet was then dissolved in an appropriate volume of either sterile dH_2O, TE (10/1) or 10 mM Tris-HCl, depending on the subsequent application.

Genomic DNA extraction for PFGE

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

(PMSF) to eliminate residual Proteinase K activity, and incubated on the flask shaker for 2 h at 37 ºC. Plugs were washed once in TE (10/100) and stored at 4 ºC in the same solution (Yeoman, 2009).

Genomic DNA extraction for Whole-Genome Sequencing (WGS)

Butyrivibrio and *Pseudobutyrivibrio* strains were revived and grown using 2× GenRFV for 24 h in 400 mL cultures, and Gram staining was used to check culture morphology and purity. Genomic DNA (gDNA) was extracted using the Qiagen Genomic-Tip 500/G Maxi kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, that includes DNAse treatment with Proteinase K (Qiagen, Hilden, Germany) and RNAse treatment with RNase A (Qiagen, Hilden, Germany).

2.2.7 Nucleic acid extractions

Genomic RNA and DNA was extracted from *Butyrivibrio* co-cultures using a modified version of a liquid N_2 and grinding method from Lueders *et al*., (2004).

RNA handling practises

To prevent RNase contamination during RNA experiments, equipment, surfaces and pipettes were wiped with Ambion RNaseZap (Thermo Fisher Scientific Inc., Waltham, MA, USA) and only RNase-free pipette tips and plastic-ware were used. All solutions and buffers were made using RNAse-free, DEPC-treated H_2O and all equipment for RNA extractions were made RNAse-free by heating at 180 ºC for 2 h and autoclaved at 121 ºC for 20 min.

Modified phenol-chloroform nucleic acid isolation with bead beating

The following protocol was for 10 g of frozen culture sample harvested as described below. The frozen sample was weighed into a 50 mL Falcon tube containing 10 mL of RNA Bacterial Protect Reagent (Qiagen, Hilden, Germany), and left to thaw on ice for 2 h. The thawed sample was centrifuged at 15,000 × *g* for 10 min at 4 ºC and the supernatant discarded. The cell pellet was resuspended in 1 mL of Extraction buffer A (200 mM NaCl, 20 mM EDTA), 420 µL of 20% (w/v) SDS and 1 mL of a mixture of phenol: chloroform: isoamyl alcohol solution. The resuspended mixture was transferred to 2 mL screw cap vials (Sarstedt, Nümbrecht, Germany) containing 500 mg of 0.1 mm diameter zirconia/silica beads (BioSpec, Bartlesville, OK, USA), and cells were disrupted by bead-

beating samples twice using a Mini-Beadbeater-96 (BioSpec, Bartlesville, OK, USA) for 4 min at full speed. Samples were stored at 4 ºC overnight followed by three rounds of bead-beating, centrifuged at 10,000 × *g* for 5 min and the top layer was transferred to a fresh 1.5 mL Eppendorf tube. An equal volume of isopropanol and 0.1 volume of 3 M sodium acetate (pH 5.5) were added, gently mixed and stored at -20 ºC overnight. The RNA/DNA was harvested by centrifugation at 10,000 × *g* for 30 min at 4 ºC, the supernatant was discarded and residual isopropanol was carefully removed using a pipette. The RNA/DNA pellets were washed twice with 500 µL ice-cold 70% (v/v) ethanol and centrifuged at 10,000 × *g* for 10 min at 4 ºC after each wash (Lueders *et al.*, 2004). The washed pellets were air-dried for 20 min and re-suspended in 100 µL of nuclease-free water (Thermo Fisher Scientific Inc., Waltham, MA, USA). The RNA/DNA samples were stored at -85 °C and used for qPCR analyses.

DNAse treatment

Following RNA/DNA extraction, the residual double and single stranded DNAs were removed from the RNA samples using the Baseline-ZERO DNase (Epicentre Technologies, Madison, WI, USA) treatment kit according to the manufacturer's instructions. To concentrate the purified RNA, samples were precipitated by addition of 1:10 volumes of 5 M ammonium acetate and 2.5 volume of 100% ethanol, mixed and incubated at -20 ºC for 30 min. Samples were centrifuged at 16,000 × *g* for 15 min at 4 ºC and the supernatant discarded. The pellets were washed with 500 µL of ice-cold 70% (v/v) ethanol and centrifuged again as before after each wash. The washed pellets were air-dried for 20 min and re-suspended in 100 µL of nuclease free water (Thermo Fisher Scientific Inc., Waltham, MA, USA). RNA was stored at -85 °C until required. DNA elimination was verified by a lack of product from PCR amplifications of the 16*S* rRNA gene.

RNA purification and quantification

Following DNAse treatment, the MEGAClear kit (Thermo Fisher Scientific Inc., Waltham, MA, USA) was used to purify mRNA following the manufacturer's instructions, which also removes tRNAs and 5*S* rRNA. In addition, the elution procedure was repeated with a second pre-heated 50 µL aliquot of Elution solution. RNA yield and quality were measured using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with the RNA 6000 Nano Chip kit following the manufacturer's instructions.

RNA integrity was also determined using agarose gel electrophoresis under RNase free conditions to detect distinct 16*S* and 23*S* ribosomal RNA (rRNA) bands. RNA samples were stored at -85 ºC prior to sending for RNA sequencing at the Beijing Genomics Institute (BGI, China).

cDNA library preparation for RNA-seq

cDNA was synthesised from the DNase-treated RNA samples at the BGI (Beijing, China) and an overview of the experimental pipeline for transcriptome preparation is outlined in Figure 2.1. Briefly, pre-treatment of the RNA sample involved thawing samples on ice, centrifugation and mixing prior to quality testing. Concentration of RNA was determined using NanoDrop® ND-1000 for $OD_{260/280nm}$ and $OD_{260/230nm}$ and electrophoresis profile analysis in the the RNA 6000 Nano Chip Kit for capillary electrophoresis (2100 Bioanalyzer; Agilent Technologies, Santa Clara, CA, USA), with the quality score determined by RIN analysis of for 23*S*/16*S* rRNA peaks. After total RNA concentration and quality was determined, the remaining cDNA synthesis and sequencing were done by BGI. Firstly, fragmentation buffer was added for disrupting mRNA to short fragments. These fragments and random hexamer-primers were then used to synthesise the first cDNA strand. The second cDNA strand was synthesised using buffer, dATPs, dGTPs, dCTPs, dUTPs, RNase H and DNA polymerase I, respectively, after removing dNTPs (including dTTP). Short fragments (100 to 500 nucleotides) were purified using QIAquick® PCR purification kit (Qiagen, Hilden, Germany) and dissolved with Elution Buffer (EB) prior to addition of poly(A). Fragments were ligated to sequencing adapters and the Uracil N-Glycosylase (UNG) enzyme was used to degrade the second strand cDNA (containing dUTP). The product was purified using the MiniElute PCR purification kit (Qiagen, Hilden, Germany) prior to PCR amplification. The generated libraries were then sequenced using Illumina HiSeq 2000 technology (BGI, Beijing, China).

2.2.8 Nucleic acid quantity and quality analysis

Nucleic acid quantity was measured using a fluorometer (Qubit® 2.0; Life Technologies, Carlsbad, CA, USA) as per the manufacturer's instructions. DNA was quantified using the Quant-iT dsDNA Broad-range (BR) Assay kit (2 to 1000 ng) and RNA was quantified using the Quant-iT RNA Assay kit (5 to 100 ng).

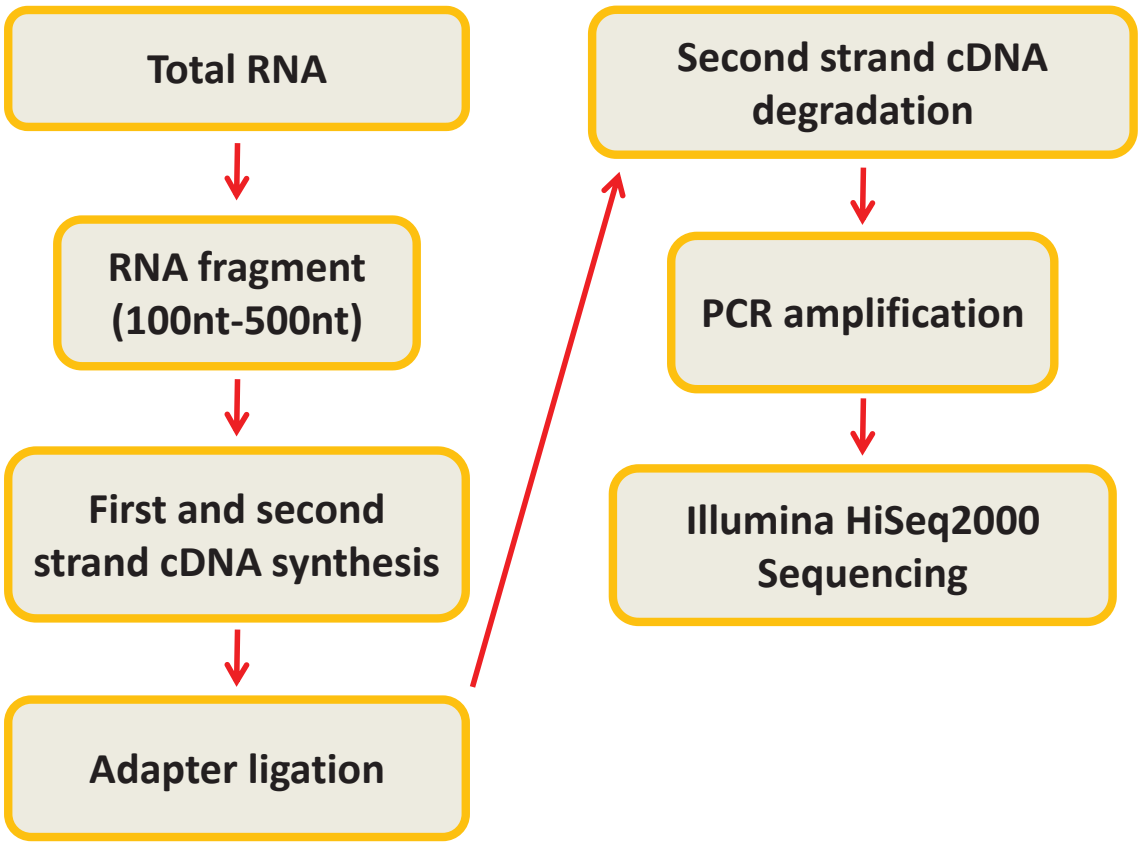


Figure 2.1. Preparation pipeline used for prokaryotic strand-specific transcriptome analysis.

Nucleic acid quality was also measured using the NanoDrop ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), where DEPC-treated H_2O was used as a blank. The nucleic acid concentration and quality were recorded using absorbance at 260 nm and 230 nm wavelengths respectively. Nucleic acid yield was also quantified by running 3 µL of sample on a 2% (w/v) agarose gel (using the 1 Kb Plus DNA or Lambda DNA/Hind III molecular weight-size markers) and checking nucleic acid integrity.

2.2.9 Sanger sequencing

Sanger sequencing of PCR products was conducted by Massey Genome Service (Massey University, Palmerston North, NZ). This service included fluorescent labelling of PCR products using BigDye™ Terminator (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). Data was received as ABI trace files, which is the file format produced by Applied BioSystems

sequencing machines. ABI files, containing the sequence and associated probabilities of the four nucleotide bases, were viewed using Trev (Ewing *et al.*, 1998; Bonfield *et al.*, 2002), a DNA trace visualization and editing tool from the Staden package and analysed using Geneious 8.1 (Kearse *et al.*, 2012).

2.2.10 Polymerase chain reaction (PCR)

PCR and Long Range PCR (LR-PCR) were performed in 50 µL unless specified otherwise. All PCR reagents were manufactured by Invitrogen and Platinum *Taq* or Platinum High Fidelity (HiFi) *Taq* DNA polymerases were used. The standard PCR and LR-PCR reaction mixes and the order in which they are added, are listed in Tables 2.13 and 2.15 respectively. The amount of dH_2O added to each PCR reaction was dependent on the concentration of the template DNA added and was therefore adjusted to bring the total volume of each reaction to 50 µL. All reaction components were added (except the template DNA) into a 0.2 mL thin-walled PCR tube (Eppendorf, Germany), vortexed for 5 sec and quickly centrifuged to return the contents to the bottom of the tubes. Template DNA was added last and PCR amplifications were performed on Mastercycler® proS (Eppendorf, Hamburg, Germany) thermal cycler using conditions listed in Tables 2.14 and 2.16 for PCR and LR-PCR respectively. The PCRs were carried out in triplicate for each sample and the products visualised by agarose gel electrophoresis. DNA fragments resulting from PCR amplifications were purified using QIAquick® PCR Purification Kits (Qiagen, Hilden, Germany) as per manufacturer's instructions.

Table 2.13. Assembly of PCR reagents.

PCR reagents *	Amount per reaction (µL)	Final concentration
DNA-free water	37.3	N/A
10× PCR Buffer, minus Mg	5	1×
10 mM dNTP mixture (each)	4	0.2 mM
50 mM $MgCl_2$	1.5	1.5 mM
Primer Mix (20 µM each 5' & 3')	0.5 each	0.2 µM
Template DNA (diluted)	1	(as required)
Platinum *Taq* DNA Polymerase	0.2	1.0 U
Total Volume	50	

***** Reagents were added in the order listed.

Table 2.14. Thermal profile for PCR.

PCR step	Temperature (ºC)	Time
1 Initial denaturation	94	2 min
2 Denaturation	94	30 sec
3 Annealing	56	30 sec
4 Extension	72	2 min*
5 Repeat steps 2 to 4 total 35×		
6 Final extension	72	10 min

***** Recommended by manufacturer, 1 min per Kb of DNA template.

Table 2.15. Assembly of reagents for LR-PCR.

LR PCR constituents *	Amount per reaction (µL)	Final concentration
DNA-free water	37.8	N/A
10× HiFi PCR Buffer	5	1×
20 mM dNTP mixture (each)	2	0.2 mM
50 mM $MgSO_4$	2	2.0 mM
Primer Mix (20 µM each 5' & 3')	1 each	0.4 µM
Template DNA (diluted)	1	(as required)
HiFi Platinum *Taq* (5 units/µL)	0.2	1.0 U
Total Volume	50	

***** Components were added in the order listed.

Table 2.16. Thermal profile for LR-PCR.

PCR step	Temperature (ºC)	Time
1 Initial denaturation	93	3 min
2 Denaturation	93	15 sec
3 Annealing	56	30 sec
4 Extension	68	2 min*
5 Repeat steps 2 to 4 total 35×		
6 Final extension	68	10 min

***** Recommended by manufacturer, 1 min per Kb of DNA template.

2.2.11 Phylogenetic analysis of full-length 16*S* rRNA gene sequences

PCR amplification of the 16*S* rRNA gene

PCR amplifications were performed using the Eppendorf Mastercycler® proS (MediRay, NZ) thermal cycler, using a standard reaction mix (Table 2.13) with the annealing temperature adjusted to 50 ºC for 1 min. The primer pair of fD1 (5'-GAGTTTGATC-MTGGCTCAG-3') and rD1 (5'-AAGGAGGTGATCCARCCG-3') was used to amplify the V1-V3 region of the 16*S* rRNA gene for subsequent Sanger sequencing. All strains produced single, clear PCR product bands when analysed by

agarose gel electrophoresis, with concentrations of approximately 60-100 ng/µL determined by NanoDrop. Full-length 16*S* rRNA marker gene sequences (>1,400 bp) obtained by PCR amplification, were sequenced by automated Sanger sequencing using 16*S* rRNA gene-specific primers (Table 2.7). The eight primers: fD1, rD1, 1492r, 1382r, 1100r, 806r, 514f and 514r (Lane, 1991; Dorsch and Stackebrandt, 1992; Escalante *et al.*, 2001; Duthoit *et al.*, 2003; Bergmann *et al.*, 2011), were used to sequence the almost full length 16*S* rRNA marker gene.

16*S* rRNA gene sequencing and analysis

Full-length 16*S* rRNA gene sequences were received as ABI files and were viewed, edited, assembled, manually processed and primer sequences removed using the Staden (Staden *et al.*, 1999) and the Geneious software packages (Kearse *et al.*, 2012). The almost full-length isolate sequences were assessed for quality (no greater than 1% ambiguous base calls and >1,400 bp long sequences), and the consensus sequence for each isolate was converted to FASTA file format to be used in further analyses. Assembled sequences were assessed and checked for the presence of chimeric PCR products by individual confirmation using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) algorithm to compare the nucleotide 16*S* rRNA gene sequences to their closest cultured relatives in the "16*S* ribosomal RNA sequences (Bacteria and Archaea)" NCBInr non-redundant database (http://blast.ncbi.nlm.nih.gov) and using the Megablast algorithm (Pruitt *et al.*, 2005), and the Ribosomal Database Project (RDP) (Maidak *et al.*, 2001).

Phylogenetic analyses

Global alignment of the almost full-length 16*S* rRNA gene nucleotide sequences was performed using ClustalW (Thompson *et al.*, 2003), and phylogenetic analyses were performed using Molecular Evolutionary Genetics Analysis (MEGA6) software version 6.0 (Tamura *et al.*, 2013). Phylogenetic trees were constructed using the Neighbor-Joining method (Saitou and Nei, 1987) with distances calculated using the Kimura 2-parameter method (Kimura, 1980) and pair-wise deletions of gaps. Phylogeny based on 16*S* rRNA gene sequence was inferred using the Maximum Likelihood method (Akaike, 1998). Bootstrap analysis with 10,000 replicates was used to assess the statistical strength of the branch positions (Kimura, 1980). The 16*S* rRNA gene sequence from

Methanobrevibacter ruminantium M1 (Genbank accession number CP001719) was used as an out-group for the tree.

2.2.12 PCR amplification of enolase genes

Genomic DNAs from *Butyrivibrio* strains were screened for the presence of enolase genes by PCR amplification using specifically designed primers from *B. crossotus* DSM 2876 (forward 5'-AATGGACCTAYGCAGATGC-3' and reverse 5'-ATCTGGTTRAG-CTTWATAAG-3') as described by Kelly *et al*., (2010). PCRs were performed in 50 µL volumes using the standard PCR protocol and reaction mix (Table 2.13), with the annealing temperature adjusted to 50 ºC for 30 sec. PCR products were analysed by agarose gel electrophoresis and concentrations were determined by NanoDrop and Qubit® dsDNA BR assay kit (Thermo Fisher Scientific Inc., Waltham, MA, USA).

2.2.13 Cloning *Butyrivibrio* marker genes

DNA isolated from B316 and MB2003 was used as a template in triplicate PCR reactions (Tables 2.17 and 2.18) using the specific primers designed for B316 and MB2003 marker genes (Table 2.8). The amplicons were visualised by agarose gel electrophoresis, pooled, combined with 20% (v/v) of 6× Orange G loading dye and added to a single well of a 0.8% (w/v) agarose gel containing 2× SYBR® DNA stain. The PCR product was visualised on a transilluminator and excised from the gel with a sterile scalpel. DNA was extracted from the gel slice using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) following the manufacturer's instructions.

Purified PCR products (4 µL) were ligated into the pCR 2.1 TOPO cloning vector using the TOPO TA cloning kit (Thermo Fisher Scientific Inc., Waltham, MA, USA), following the manufacturer's instructions and then chemically transformed into One Shot TOP10 competent cells (Thermo Fisher Scientific Inc., Waltham, MA, USA) following the manufacturer's instructions. Transformed cells were plated onto pre-warmed LB agar plates containing 50 µg/mL of ampicillin and 40 µL of 40 mg/mL X-gal spread onto each plate. The plates were incubated overnight at 37 ºC and the white, insert-containing colonies were purified by streaking onto a LB agar plate containing ampicillin only (no X-gal).

Table 2.17. Reagents for PCR cloning.

Reagent	Amount per reaction (µL)	Final concentration
DNA-free water	18.9	N/A
10× PCR buffer containing Mg^{++}	2.5	1×
50 mM $MgCl_2$	1	2 mM
10 mM dNTP mixture (each)	0.5	0.2 mM
Sense primer F * (10 µM)	0.5	0.2 µM
Antisense primer R * (10 µM)	0.5	0.2 µM
Taq DNA polymerase (5 U/µL)	0.1	0.02 U
Template DNA	1	(as required)
Total volume	25	N/A

* Primers used to generate PCR products are detailed in Table 2.8.

Table 2.18. Thermal profile of PCR for clone library.

PCR step	Temperature (°C)	Time
1 Initial denaturation	94	2 min
2 Denaturation	93	15 sec
3 Annealing	56	30 sec
4 Extension	72	2 min *
5 Repeat steps 2 to 4 total 30×		
6 Final extension	72	10 min

* Recommended by manufacturer, 1 min per Kb of DNA template.

Colony PCRs

Putative insert-containing *E. coli* colonies (toothpick scrapes) were resuspended directly into the PCR reagents. Colony PCR amplifications were performed in an Eppendorf Mastercycler® proS thermal cycler, using the reaction mix listed in Table 2.19. The colony plus PCR mixture was vortexed for 5 sec and quickly spun to return the contents to the bottom of the tubes and the reactions were cycled as in Table 2.20. The custom primer pair, GEM2987F and TOP168R (Thermo Fisher Scientific Inc., Waltham, MA, USA) was used to amplify the cloned fragment. The primer pair, M13F (GTAAAACGACGGCCAG) and M13R (CAGGAAACAGCTAT-GAC) (Thermo Fisher Scientific Inc., Waltham, MA, USA) supplied by manufacturer was used for sequencing the amplified insert. The PCR products were separated by agarose gel electrophoresis and visualised by staining using the SYBR® stain. The PCR products of expected sizes (80-100 bp) were Sanger sequenced.

Table 2.19. Colony PCR reagents.

Reagent	Amount per reaction (µL)	Final concentration
DNA-free water	31.5	N/A
10× PCR buffer containing Mg^{2+}	5	1×
50 mM $MgCl_2$	1	1 mM
2 mM dNTP mixture (each)	10	0.4 mM
Sense primer GEM2987F [A] (10 µM)	0.5	0.1 µM
Antisense primer TOP168R [B] (10 µM)	0.5	0.1 µM
Taq DNA polymerase (5 U/µL)	0.5	0.05 U
Template DNA	0-1	(as required)
Total volume	50	N/A

[A], GEM2987F, CCCAGTCACGACGTTGTAAAACG (Thermo Fisher Scientific Inc., Waltham, MA, USA) and [B], TOP168R, ATGTTGTGTGGAATTGTGAGCGG (Thermo Fisher Scientific Inc., Waltham, MA, USA).

Table 2.20. Thermal profile of the colony PCR reaction.

PCR step	Temperature (ºC)	Time
1 Initial denaturation	94	1 min
2 Denaturation	94	15 sec
3 Annealing	56	30 sec
4 Extension	72	1 min *
5 Repeat steps 2 to 4 total 30×		
6 Final extension	72	7 min

*Recommended by manufacturer, 1 min per Kb of DNA template.

2.2.14 qPCR

Abundances of the marker genes from *B. hungatei* MB2003 and *B. proteoclasticus* B316 samples were quantified using a Rotor-Gene 6000 real-time rotary analyser (Corbett Life Science, Qiagen, Hilden, Germany) and amplicon detection by KAPA SYBR® FAST qPCR Kit Master Mix (2×) Universal (KAPA Biosystems, Wilmington, MA, USA). Control DNA (plasmids containing specific marker genes) were quantified by fluorimetry using the Qubit® dsDNA BR assay kit, and diluted 10-fold in series to produce 5 standards ranging from 10^3 to 10^{11} copies per µL DNA in each reaction. Samples were analysed in triplicate and a control reaction lacking template was performed for each sample used in the qPCR. Reactions were set up in a Gene-Disc 100 (Corbett Life Science, Qiagen, Hilden, Germany) and sealed with permanent adhesive film (Corbett Life Science, Qiagen, Hilden, Germany). The plasmid DNAs containing the PCR-amplified marker genes were used to generate the standard curves and the final concentration of primers used was 100 nM. The cycle threshold (C_T) values were

converted to present real-time PCR data as absolute or relative expression levels, following transformation of the data using a standard curve (Chen *et al.*, 2005). The data is presented as copy numbers per µL DNA in each reaction, where the dilution factors have been accounted for and normalised to present overall calculated concentration per sample.

Reactions (20 µL) were prepared as shown in Table 2.21, with the extracted RNA/DNA (1 µL) used as the template. DNA amplification was performed in 5 different dilutions of the template (undiluted, 1:10, 1:100, 1:1,000 and 1:10,000), to give 10 replicates for each sample. qPCR experiments utilised 3 biological replicates per strain and all qPCR reactions were performed in triplicate. Specifically designed *Taq*Man primer/probe assay (100 nM/µL$_F$) sense and antisense oligonucleotides (Table 2.8) designed *via* IDT (Integrated DNA Technologies, Coralville, IA, USA), were used to generate amplicons (80-100 bp each) within target genes in B316 and MB2003 respectively. The target specificity for each gene primer/probe assay were examined by reciprocal BLAST analysis against the individual genomes. The accepted cross-amplification threshold of ≤ 1-2 mismatches/misalignments in each primer and/or probe assay was achieved. The PCR amplification parameters are detailed in Table 2.22. The absence of non-specific signals was evaluated by applying melting curves for each qPCR between 72 ºC and 95 ºC. The Rotor-Gene 6000 series software (Corbett Life Science, Qiagen, Hilden, Germany) was used to analyse qPCR data, and copy numbers of target genes were calculated based on the dilutions made and DNA yield per g of template.

Table 2.21. qPCR reagents.

Reagent	**Amount per reaction (µL)**	**Final concentration**
PCR-grade water	Up to 20	N/A
KAPA SYBR® FAST qPCR Master Mix (2×) Universal *****	10	1×
*Taq*Man Sense primer (10 µM)	1.0	100 nM
*Taq*Man Antisense primer (10 µM)	1.0	100 nM
*Taq*Man Probe	1.0	100 nM
Template DNA	0-1	≤20 ng
Total volume	20	N/A

***** Final $MgCl_2$ concentration at 1× is 2.5 mM.

Table 2.22. qPCR thermal profile.

PCR step	Temperature (ºC)	Time	Cycles
1 Enzyme activation	95	10 min	Hold
2 Denaturation	95	10 sec	40
3 Annealing	56	30 sec	
4 Extension	72	30 sec [A]	
5 Melt curve	72-95	5 sec [B]	

[A], Recommended by manufacturer, 1 min per Kb of DNA template. [B], Melt curve temperature was 1 ºC per step, wait for 90 sec of pre-melt conditioning on first step and 5 sec for each step there afterward.

2.2.15 Carbon source utilisation

To test bacterial carbon source utilisation, fresh overnight *Butyrivibrio* cultures were inoculated into Hungate tubes containing RM02 medium broth with the separate addition of each of the 32 carbon sources, including seven polysaccharides (glycogen, pectin, inulin, cellulose, dextrin, starch and xylan) at 0.5% (w/v) final concentration. Cultures were inoculated in triplicate and grown under anaerobic conditions overnight at 39 ºC. OD_{600nm} readings were measured initially after inoculation, and after 24 h and 72 h of incubation for soluble substrates (Moon *et al.*, 2008).

For the soluble substrates, ΔOD_{600} readings of 0.5-1.0 were scored as ++, 0.2-0.5 scored as +, and 0-0.2 scored as -. Positive controls were individual strains grown in RM02 with GenRFV containing D-glucose, D-cellobiose, D-xylose and L-arabinose. Two types of negative controls were included; inoculation controls in which individual strains were inoculated into RM02 without carbon source added, and uninoculated Hungate tubes containing RM02 with substrate. Uninoculated Hungate tubes containing only RM02 without substrate were used as spectrophotometer blanks. Growth on polysaccharide substrates was assessed by measurement of volatile fatty acids (VFA) production.

2.2.16 Fermentation end product analysis

The insoluble substrates analysed were pectin, inulin, cellulose, dextrin, starch and xylan. Additional strains used as controls were *B. hungatei* MB2003, *B. proteoclasticus* B316 and *P. xylanivorans* MA3014. Also uninoculated media supplemented with 0.5% cellobiose (w/v) and cultures grown with cellobiose (0.5% w/v) as the sole substrate were also used as controls. The *Butyrivibrio* cells for use as inocula, were cultured anaerobically at 39 ºC in Hungate tubes containing pre-warmed 9.5 mL of RM02 medium with 0.5% w/v cellobiose, and passaged three times. Cultures were then inoculated in triplicate (0.5% w/v inoculum) into Hungate tubes containing 9.5 mL of RM02 medium

and 0.5% (w/v) insoluble substrate, and grown anaerobically at 39 ºC with gentle horizontal shaking for 48 h.

The VFA production profiles of *Butyrivibrio* strains were determined from two replicate broth cultures of each *Butyrivibrio* strain. End products are expressed as the average of the total VFA concentrations (mM) of two duplicate cultures with background inoculum and insoluble substrate VFAs in media controls subtracted. After growth, 1.8 mL aliquots of bacterial cultures were centrifuged for 10 min at 21,000 × *g* at 4 ºC. A 900 µL aliquot of the supernatant was transferred to a 1.5 mL Eppendorf tube and the protein precipitated with the addition of 100 µL 20% (v/v) orthophosphoric acid containing the internal VFA standard, 2-ethylbutyric acid (Sigma-Aldrich, St. Louis, MO, USA).

Samples were frozen overnight at -20 ºC, thawed and re-centrifuged under the same conditions to remove any remaining proteins and solids. A 0.8 mL aliquot of the clear supernatant was transferred to a 1.5 mL crimp top GC vial for immediate analysis or stored at -20 ºC until required. Samples were analysed using a Hewlett Packard/Agilent 6890 GC gas chromatograph series (GMI, Ramsey, MN, USA) with flame ionisation detection. A 30 mm x 0.53 mm I.D. x 25 µm film HP-PLOT molesieve capillary column (Agilent Technologies, Santa Clara, CA, USA) was used for separation with a column helium flow rate of 5.5 mL/min. A 1.0 µL aliquot of sample was injected into the column inlet (cool on-column inlet) and the inlet temperature was set to track oven temperature, while the detector was at 240 ºC. The column temperature was programmed to start at 85 ºC and increase by 10 ºC/min up to 180 ºC (held for 5 min), giving a run time of 14.5 min per sample. This protocol was used to quantify acetic, butyric and propionic acids and the branched chain acids (BCVFAs) isobutyric and isovaleric acids, which were used as indicators of growth and insoluble polymer substrate utilisation of the *Butyrivibrio* strains. The BCVFAs and propionic acids were present only in trace amounts and were not further analysed.

Derivatisation protocol for analysis of lactate, formate and succinate

An additional down-scaled method (Richardson *et al.*, 1989) was used to derivatise formic, lactic and succinic acids, which could not be detected using the VFA method described above and are important fermentation products used to characterise phenotypically *Butyrivibrio* strains. A 200 µL aliquot of acidified *Butyrivibrio* culture fluid supernatant containing the internal standard 2-ethylbutyric acid (Sigma-Aldrich, St. Louis, MO, USA) was mixed vigorously with 100 µL of 37% (v/v) HCl ACS reagent

(Sigma-Aldrich, St. Louis, MO, USA) and 800 µL diethyl ether. The mixture was left to settle for 1 min and the top aqueous layer containing ether was transferred into a 2 mL Eppendorf tube. This extraction process was repeated by adding a further 800 µL of diethyl ether to the aqueous phase and the two extracts were then combined. A 800 µL aliquot of extract was derivatised with 100 µL of *N*-methyl-*N*-*t*-butyldimethylsilyltrifluoroacetamide (MTBSTFA) (Sigma-Aldrich, St. Louis, MO, USA) (Richardson *et al.*, 1989), and transferred into a crimp-top GC vial. The samples were heated at 80 ºC for 20 min, and left at room temperature for approximately 48 h to allow for the full derivatisation of lactate.

Samples were analysed using a Shimadzu GC-2010 Plus Gas Chromatograph with helium ionisation detector and a 30 mm x 0.53 mm I.D. x 50 µm film MXT-Msieve5A PLOT capillary column (Restek Corporation, Bellefonte, PA, USA) was used for separation. A split injection of 1 µL sample was made at a ratio of 20:1, with a column helium flow rate of 21.36 mL/min. Injector and detector temperatures were both set at 240 ºC. The column temperature was initially held at 50 ºC for 2 min and then increased at 5 ºC/min to 130 ºC, followed by 15 ºC/min to 240 ºC (held for 4.7 min). This resulted in a run time of 30 min per sample.

2.2.17 High-pressure ion chromatography (HPIC)

The monosaccharides released during *Butyrivibrio* mono-culture and co-culture growth were determined using a HPIC method. Culture samples from the eight time intervals (0, 2, 4, 6, 8, 10, 12 and 16 h) were used for analysis. To minimise the effects of co-elution of interfering compounds, samples of culture (1 mL) were first prepared by centrifugation at 13,000 × *g* followed by a filtration through a 0.22 µm filter. All samples and standards were prepared in triplicate and diluted with de-carbonised water to 1 mg/mL. The de-carbonised water, obtained through a Milli Q-system and ultrasonic cleaning using an ultrasonic bath for at least 40 min at 400 kHz, was carried out to prevent any carbon influencing the column during the HPIC run. The HPIC system used was the DIONEX ICS-5000 (Thermo Fisher Scientific Inc., Waltham, MA, USA). Processed samples were loaded into 1.5 mL glass flat-bottomed HPIC screw vials (Thermo Fisher Scientific Inc., Waltham, MA, USA) and closed with screw caps with silicone/blue PTFE pre-slit septa (Thermo Fisher Scientific Inc., Waltham, MA, USA). A CarboPac PA20 3 × 30 mm column (Thermo Fisher Scientific Inc., Waltham, MA, USA), was operated at a temperature of 30 ºC and at a flow rate of 0.5 mL/min. The samples were injected in a

volume of 10 µL and the detector used was an ED50 electrochemical detector. NaOH (10 mM) was used as the mobile phase during the run and 200 mM NaOH for regeneration of the column and co-elution of interfering compounds from previous injection.

Chemicals used for standards were: arabinose (Lot. 112K-0591, Sigma-Aldrich, St. Louis, MO, USA), fucose (Lot. 058H-0962, Sigma-Aldrich, St. Louis, MO, USA), galactose (Lot. 121F-0014, Sigma-Aldrich, St. Louis, MO, USA), glucose (AgR No. 25988, VWR International Ltd, Poole, UK), rhamnose (AgR No. 11783, VWR International Ltd, Poole, UK) and xylose (AgR No. 11762, VWR International Ltd, Poole, UK). The standards used to determine retention times were diluted with de-carbonised water to a concentration of 1 mg/mL for test runs. New standards are made up after a comparison with the test samples according to a 4 point-multi-calibration standard sequence with concentrations corresponding to the found test sample concentrations. Each standard was run individually to ensure retention times were the same as those observed in previous HPIC runs. A mixed sample of all six monosaccharides was then used to generate calibration curves for each sample batch. The concentrations (mg/mL) of monosaccharides released in the samples were then calculated using standard curves and converted to the total amount of monosaccharides released (in mM). HPIC results were analysed using the Chromeleon software (Thermo Fisher Scientific Inc., Waltham, MA, USA).

2.2.18 Electron microscopy

Scanning (SEM) and Transmission (TEM) electron microscopy were carried out at the Manawatu Microscopy and Imaging Centre (MMIC), Massey University, Palmerston North, NZ. Bacterial cultures for SEM and TEM were grown in 10 mL BY medium supplemented with 0.5% w/v cellobiose containing a single 10 mm neutral detergent fibre (NDF) strand (Maynard *et al.*, 1979; Van Soest *et al.*, 1991) in triplicate. The NDF fraction was produced and prepared for experimental use as outlined by Maynard *et al.* (1979). Biological samples for SEM were fixed in modified Karnovsky's fixative for two days at room temperature, then washed five times for 10-15 min each, in PBS. Samples were dehydrated through a series of ethanol solutions: 25% (v/v), 50% (v/v), 75% (v/v), and 95% (v/v) for 10-15 min each and 2× in 100% ethanol for approximately 1 h. The samples were then Critical Point (CP) dried using liquid CO_2 as the CP fluid. The dried samples were mounted onto an aluminium specimen support stub using double-sided

adhesive tape or conductive silver paint, sputter coated with gold, and observed *via* SEM (Neinhuis and Edelmann, 1996).

Triplicate bacterial samples for TEM were grown in 10 mL RM02 and BY media supplemented with 0.5% (w/v) cellobiose and incubated at 37 ºC for 24 h. A drop of bacterial sample was placed onto a piece of parafilm followed by placement of a Formvar grid (shiny side down) on top of the drop of the sample. After 4 min, the liquid was drained from the grid with the edge of a piece of filter paper. The grid was placed on a drop of 2% (v/v) phosphotungstic acid stain (or 2% w/v uranyl acetate) in water for 4 min and drained with filter paper. The grid was allowed to dry, and observed under TEM (Bozzola and Russell, 1999).

2.2.19 Cell motility assay

To examine motility of *Butyrivibrio* strains, a Motility Agar Stab test was carried out using 5 mL of 0.3% (w/v) agar in RM02 with cellobiose as the carbon source in a 10 mL Hungate tube. Freshly grown *Butyrivibrio* cultures were stab-inoculated into the agar using a straightened inoculating loop, about two thirds of the way into the agar, ensuring to make the stab as straight as possible. The cultures were incubated at 37 ºC for 24 to 48 h and observed under a light source. Non-motile strains displayed visible growth confined to the area immediately surrounding the initial inoculation stab and were indicated with a -. Motile bacterial cells migrate through the agar to produce a diffuse or cloudy growth pattern as evidenced by turbidity located a distance away from the inoculation stab and were indicated with a + (Tittsler and Sandholzer, 1936). The strains used as negative controls include *B. hungatei* MB2003, *B. proteoclasticus* B316, and *Streptococcus bovis* 2B, while *P. xylanivorans* MA3014 was used as a positive control.

2.2.20 Pulsed-field gel electrophoresis (PFGE)

Restriction endonuclease digestion of PFGE gel plugs

Genomic DNA extracted as described previously (Genomic DNA extraction for PFGE) and gel plugs were cut to ~2.5 mm using a sterile scalpel, carefully transferred to a sterile Eppendorf tube containing 1 mL of TE 10/0.1 and allowed to equilibrate for 1 h. The TE 10/0.1 buffer was removed, the plug immersed in an excess of the appropriate RE Buffer (and BSA where necessary), and the plug was allowed to equilibrate in this solution for 1 h. The RE Buffer was removed and a fresh 100 µL of the same buffer was

added along with 0.2 µL of the appropriate 1 U RE. The plug was incubated in this solution at the optimal temperature of the RE for 16-24 h. The RE buffer was removed and 500 µL of TE 10/100 buffer was added and the plug kept at 4 ºC until PFGE was performed (Yeoman, 2009).

PFGE

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

2.2.21 Whole-Genome Sequencing (WGS)

For WGS of 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains, bacterial cultures were grown and genomic DNA was extracted using a Qiagen Genomic-Tip 500/G Maxi kit as described previously. To confirm isolate identity, gDNA from each isolate was used for PCR amplification of the full-length 16*S* rRNA genes using primers 27f and 1492r (Weisburg *et al.*, 1991) and the PCR products were sequenced. An additional RNAse treatment and RNA removal by double-RNase digestion (DNA Genotek, Ottawa, ON, Canada), was carried out for all gDNA samples. The *Butyrivibrio* and *Pseudobutyrivibrio* gDNAs were submitted to the U.S. Department of Energy's Joint Genome Institute (DOE

JGI) for sequencing as part of the Community Sequencing Project (CSP) "The Hungate 1000" (Proposal ID 612).

The genome sequencing methods used differed among the 46 genomes sequenced. *B. proteoclasticus* B316 (Kelly *et al.*, 2010), was sequenced using Sanger and 454 Pyrosequencing technology, *B. hungatei* MB2003 and *P. xylanivorans* MA3014 (Palevich, 2011), were sequenced using 454 Pyrosequencing in combination with Illumina MiSeq technology and all the remaining genomes were sequenced using Illumina HiSeq 2000 technology. Ten genomes (*Butyrivibrio* sp. WCE2006 and *Butyrivibrio* sp. NC2002 from *Butyrivibrio* sp. 2 group, *Butyrivibrio* sp. AE2005, *Butyrivibrio* sp. AE3003, *Butyrivibrio* sp. LB2008, *B. proteoclasticus* P6B7, *Butyrivibrio* sp. MC2021 and *Butyrivibrio* sp. AE2032 from *Butyrivibrio* sp. 3 group, *Pseudobutyrivibrio* sp. LB2011 and *Pseudobutyrivibrio* sp. MD2005 from *Pseudobutyrivibrio* group) were also sequenced using Illumina HiSeq 2500 technology, and three genomes, (*Butyrivibrio* sp. AE3004, *Butyrivibrio* sp. FCS014 and *Pseudobutyrivibrio ruminis* HUN009), were also sequenced using PacBio RS technology (Glenn, 2011). The JGI provided draft assembled genome sequences and automated analysis to their current specification standard for each of the submitted strains.

2.2.22 *Butyrivibrio* co-culture growth experiment

An outline of the experimental design is shown in Figure 2.2. *B. hungatei* MB2003 and *B. proteoclasticus* B316 were cultured anaerobically at 39 ºC in pre-warmed RM02 medium supplemented with 0.5% (w/v) cellobiose. Culture purity was verified *via* wet mounts and Gram staining. OD_{600nm} readings were taken at 2 h intervals until stationary phase using an Ultrospec 1100 Pro spectrophotometer (GE Healthcare, UK) to determine growth curves of the two bacteria. Mid-log phase on cellobiose substrate consistently occured at OD_{600nm} = ~0.4, after 12 h growth for both bacteria.

Freshly grown cultures were transferred into Serum bottles containing 70 mL of pre-warmed RM02 medium with 0.5% w/v cellobiose, then sub-cultured into 500 mL Schott bottles containing 350 mL RM02 medium and 0.5% w/v cellobiose. Once the mono-cultures reached mid-log phase, cell counts of each culture were carried using a Petroff-Hausser chamber (Cat. No. 3900, Hausser Scientific, Horsham, PA, USA) according to the manufacturer's instructions. Also, 2 mL of culture was collected for DNA extractions for qPCR analysis to confirm cell counts.

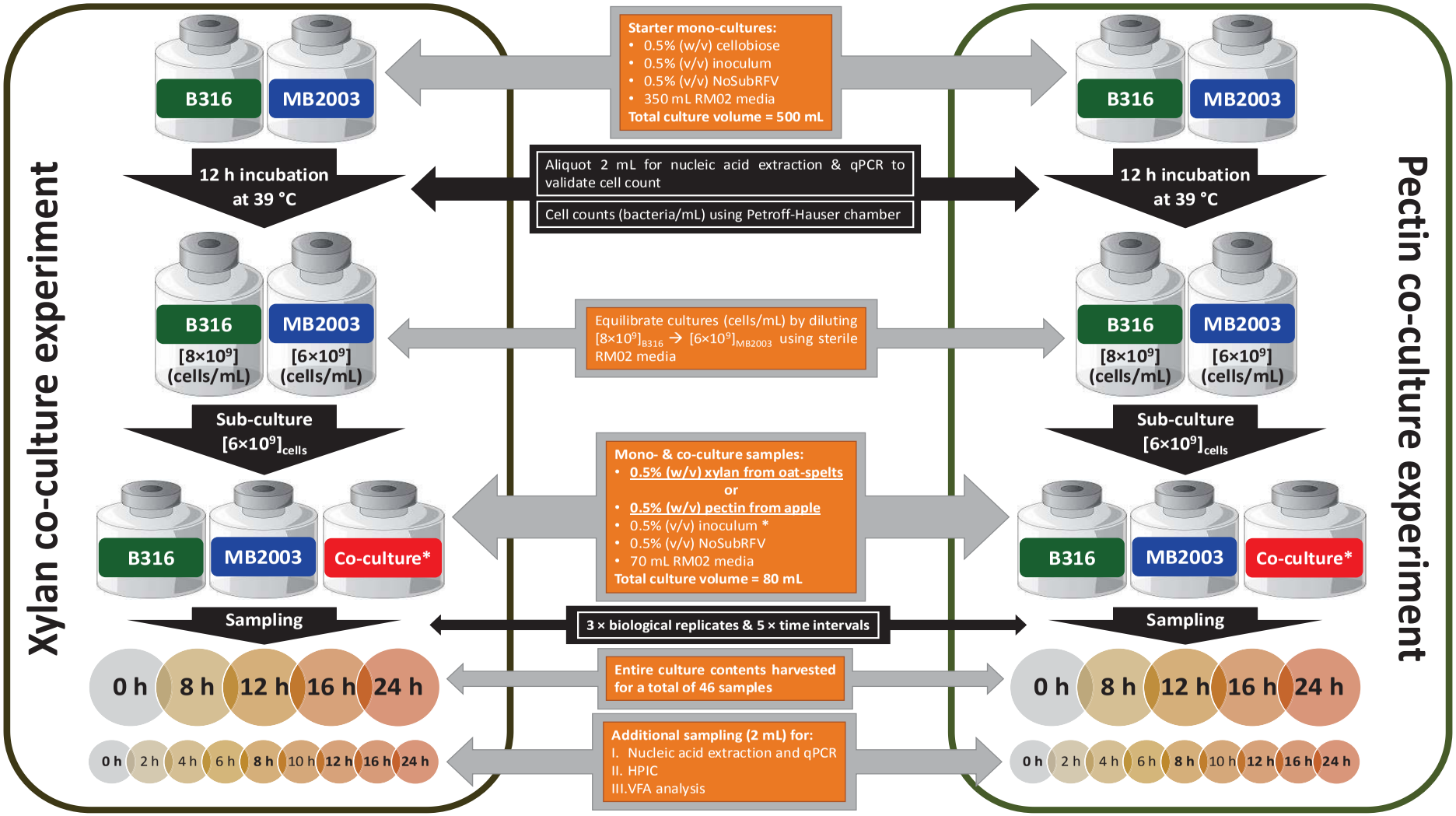


Figure 2.2. Overview of *B. hungatei* MB2003 and *B. proteoclasticus* B316 co-culture growth experiment. * 0.5% (v/v) co-culture inoculum comprised of 0.25% (v/v) *B. hungatei* MB2003 and 0.25% (v/v) *B. proteoclasticus* B316 inocula.

To account for individual differences in cell densities of the two bacteria as previously determined by cell counts, mono-cultures of MB2003 and B316 were brought to the same cell density/mL medium by the addition of sterile RM02 media. Serum bottles containing 70 mL pre-warmed RM02 media, supplemented with either 0.5% (w/v) of the insoluble substrate xylan, from oat-spelts (xylan experiment) or pectin, from apple (pectin experiment) as the main carbohydrate source, were inoculated with 0.5% (v/v) inoculum of the cell density-equilibrated cultures. For the mono-cultures, B316 or MB2003 were inoculated separately into the media and for the co-culture samples, 0.25% (v/v) inocula of each equilibrated culture was added. Inoculations were done in 15 serum bottles per organism (or co-culture) per substrate, allowing triplicate serum bottles to be harvested at 5 time points (0, 8, 12, 16 and 24 h).

On collection of each sample, the pH was measured and recorded immediately upon opening the serum bottles. Additional sampling of 2 mL per culture was done at nine time intervals (0, 2, 4, 6, 8, 10, 12, 16 and 24 h) and stored at -85 ºC for downstream qPCR, HPIC and VFA analyses. Cultures harvested at each of the five time points (0, 8, 12, 16 and 24 h) were gently mixed and the entire contents of each Serum bottle (approximately 80 mL) were poured into a thermos containing liquid N_2 while stirring to avoid clumping of the cultures. The snap-frozen samples were transferred into ziplock plastic bags and kept on dry ice before being stored at -85 ºC.

2.2.23 *In silico* analyses

2.2.23.1 *Butyrivibrio hungatei* MB2003 genome project

B. hungatei MB2003 was isolated from the rumen contents of fistulated Friesian dairy cows as described by Noel (2013) and sequenced as described by Palevich (2011).

Illumina sequencing

B. hungatei MB2003 cells were cultured anaerobically at 39 ºC in pre-warmed RM02 medium supplemented with 0.5% (w/v) cellobiose. Genomic DNA was extracted and DNA quality was determined as described previously. High-quality genomic DNA was sequenced on the Illumina MiSeq platform by Macrogen (Korea). Macrogen used the Illumina MiSeq Reagent Kit v2 (Illumina, USA) to construct shotgun libraries with 2 Kb paired-end inserts. Paired-end sequencing with 300 bp reads was performed, such that approximately 9.0 GB of clean sequence data was obtained, resulting in approximately

800-fold sequencing coverage. Two FASTQ files were provided by Macrogen, representing the forward and reverse reads, which were assembled and analysed using the Galaxy web-based comparative genomics and bioinformatics platform (Goecks *et al.*, 2010). FastQC (Andrews, 2010), a quality control tool for high throughput sequence data, was used to check the quality of the sequence data. DynamicTrim (Cox *et al.*, 2010), a read trimmer designed specifically for filtering of Illumina high throughput sequence data, was used on the FASTQ files to remove reads that did not pass the quality threshold ($P = 0.05$ or Phred score of $Q \leq 20$) containing unreadable or low quality bases, adapter sequences or duplicate read-pairs.

De novo assembly and genome closure

The MB2003 genome was sequenced, but proved difficult to close. To support genome closure and improve genome quality, the *B. hungatei* MB2003 genome sequence was assembled *de novo*. *De novo* genome assembly was performed using the Velvet (Zerbino and Birney, 2008) assembler programme that is specially designed for short read sequencing technologies, such as Illumina MiSeq and 454 Titanium. VelvetOptimiser version 2.5.5 (Table 2.11), a multi-threaded wrapper Perl script that automates and optimises the Velvet assembly process was used to convert the reads into *k*-mers (substrings of length *k* in DNA sequence data) using *velveth* which were then assembled into high quality unique contigs using *de Bruijn* graphs *via velvetg*. A *k*-mer length of 63 was used for the *de novo* assembly of the *B. hungatei* MB2003 genome sequence. The assembled genome sequence was aligned to the reference 454 Titanium *B. hungatei* MB2003 genome sequence using Geneious version 8.1 (Kearse *et al.*, 2012), which resulted in complete genome closure and verification of genome assembly. Assembly validation was confirmed by PFGE as described above.

Genome analysis and curation

Genome atlas visualisations were constructed using GENEWIZ (Jensen *et al.*, 1999) and genome content was visualised using Artemis (Rutherford *et al.*, 2000). GC skew analysis was performed by DNAPlotter (Carver *et al.*, 2009) and synteny plots were generated using MUMmer3.0 (Delcher *et al.*, 1999) (Table 2.11). The resulting images were saved as PostScript files and visualised using the Gsview program. Horizontal gene transfer studies were performed using Darkhorse (Podell and Gaasterland, 2007), GC% content (Rice *et al.*, 2000) and the Codon Adaptation Index (Sharp and Li, 1987). ORFs

displaying no significant sequence similarity to any previously described gene were annotated as "hypothetical" proteins, whereas those displaying significant sequence similarity to a gene of unknown function were annotated as "conserved hypothetical" proteins. Metabolic pathway reconstructions were performed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto, 2000) combined with an extensive review of the literature. The genome sequence was prepared for National Center for Biotechnology Information (NCBI) submission using Sequin (Benson *et al.*, 2015). The adenine residue of the start codon of the chromosomal replication initiator protein *DnaA1* (bhn_I0001) gene was chosen as the first base for the *Butyrivibrio hungatei* MB2003 genome. The nucleotide sequences of the *B. hungatei* MB2003 chromosome, chromid (BhuII), megaplasmid (pNP144) and plasmid (pNP6) have been deposited in GOLD under accession number Gp0115331.

Differential BLAST analysis (DBA)

To determine the target genes for the real time qPCR assays, a comparison of the complete *B. hungatei* MB2003 and *B. proteoclasticus* B316 genomes was conducted using DBA (Altermann, 2012) to identify unique and conserved genome regions. The reference genome of *B. hungatei* MB2003 was subjected to analysis against the *B. proteoclasticus* B316 BLASTP database using GAMOLA (Altermann and Klaenhammer, 2003). The B316 amino acid database employed the complete set of open reading frames (ORFs) or ORFeomes used in the FGD analysis (dbB316), while the second database was comprised of the MB2003 database (dbMB2003) also generated *via* FGD and excluding hits to dbB316. *E*-values of best BLASTP hits for both database sets were consolidated into an empirically determined *E*-value trust level range ([$T_{E\text{-value}}$]) and their respective differential calculated as follows: $\Delta=(T_{dbMB2003} - T_{dbB316})$. A conservation cut-off of 0 was employed and the results were visualised using Microsoft Excel software. The gene selected as a universal primer/probe assay for the *Butyrivibrio* strains was butyrate kinase, *buk* (bpr_I2323) gene. For the strain-specific assays, the cell wall binding domain-containing protein, *CBD* (bpr_I0264) gene was selected for B316, and the α-L-arabinofuranosidase, *arf51C* (bhn_I1532) gene for MB2003.

2.2.23.2 Comparative analysis of the pan-genome datasets

ANI and AF computation

ANI and the fraction of orthologous genes, AF, were used as complementary measures of genetic relatedness based on gene content between the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes. The algorithm implemented to determine ANI and AF between two genomes is a modification of the method proposed by Konstantinidis and Tiedje (Konstantinidis and Tiedje, 2005), in combination with the method described by Goris *et al*. using a custom Perl script (Goris *et al.*, 2007). ANI is a measure of nucleotide-level genomic similarity between the coding regions of two genomes (G1 and G2) using the high performance similarity search tool NSimScan and PSimScan (Kaznadzey *et al.*, 2013). First, the best bidirectional hits (BBHs) between a genome pair (G1 and G2) were computed as pairwise BBHs of genes having 70% or more nucleotide identity and at least 70% coverage of the shorter gene. Each genome sequence served as a reference genome, and the resulting ANI values were averaged.

The gANI of genome A (G1) to genome B (G2) is defined as the sum of the percent identity multiplied by the alignment length for all BBHs, divided by the sum of the lengths of the BBH genes (Figure 2.3A). The AF is computed by dividing the sum of the lengths of all BBH genes by the sum of the length of all the genes in G1 (Figure 2.3B). This computation is performed separately in both directions (i.e. from G1 to G2 and from G2 to G1). The code to perform gANI and AF computation is available at https://ani.jgi-psf.org/html/download.php. The AF and ANI were calculated for the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes to determine species cut-offs. All genomes were made publicly available from the JGI's Integrated Microbial Genomes (IMG) and the coding density filter was applied in order to minimise the impact of having draft genomes and errors in annotation on predicted coding sequences. In order to identify species ANI and AF that determine whether genomes in a pair belong to the same species, only the subset of high-quality genome pairs were utilised. The ANI cut-off of ≥ 96.5 and AF cut-off of ≥ 0.6 were used to define species.

A The Average Nucleotide Identity (ANI) is computed as:

$$gANI_{G1 \mapsto G2} = \frac{\sum_{bbh}(\text{Percent Identity} * \text{Alignment length})}{\text{lengths of BBH genes}}$$

B The Alignment Fraction (AF) is computed as:

$$AF_{G1 \mapsto G2} = \frac{\text{lengths of BBH genes}}{\sum \text{length of genes in genome 1}}$$

Figure 2.3. Equations for calculation of ANI and AF. A, average nucleotide identity (ANI) computation. B, alignment fraction (AF) computation.

Functional genome distribution (FGD)

FGD is a tool for comparative microbial genomics analysis and the interpretation of genetic diversity of bacteria (Altermann, 2012). FGD investigates the overall similarity levels between microbial genomes based on amino acid sequences of their predicted ORFeomes, which correspond to the coding sequences (CDSs) of genes, in a genome), which ultimately defines the degree of similarity of the genomes. Publicly available draft phase genomes were downloaded in FASTA format from the IMG genome database (Markowitz *et al.*, 2013), concatenated using a universal spacer-stop-spacer sequence and automatically annotated using GAMOLA (Altermann and Klaenhammer, 2003). The in-house closed genomes of *B. proteoclasticus* B316, *B. hungatei* MB2003 and draft genome of *P. xylanivorans* MA3014, have been manually annotated using GAMOLA (Altermann and Klaenhammer, 2003). Predicted ORFeomes of all genomes were subjected to an FGD analysis and the resulting distance matrix was imported into MEGA6 (Tamura *et al.*, 2013). The functional genome distribution was visualised using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method (Sneath and Sokal, 1962; Jones *et al.*, 1992).

Genome alignment

MUMmer 3 is a whole genome alignment tool that compares the alignment of contigs across a reference genome in the form of a dot plot diagram (Delcher *et al.*, 1999). Synteny plots were generated using the *mummerplot* utility. The plots reveal regions of exact matches between the pair of genomes compared and thus is an indicator of conservation between the two genomes. In a MUMmer plot, the reference genome is plotted along the x axis (*B. proteoclasticus* B316) and the query genome is on the y axis (draft *Butyrivibrio* genome). Wherever the two sequences match, a coloured dot is plotted on the chart, the forward matches are displayed in blue and the reverse matches are displayed in red. A straight line of dots starting from bottom left to the top right means that the two regions are exactly identical. A genome plot showing the presence of dots (blue and red combined) on a diagonal pattern indicates that the two genomes contain many similar sequences and thus their genomes share many similar genes. The Gsview (Womble, 1999) program was used to visualise the generated MUMmer plot.

Amino acid and codon usage analyses

Amino acid and codon usage analyses were performed using the CMG-biotools (Vesth *et al.*, 2013) package, under default parameters. The input amino acid and corresponding CDS files were downloaded from the IMG website to maintain consistent genome annotation results. The amino acid and codon usage were calculated using BioPerl modules that calculate the fraction of each amino acid or codon count of the total count of amino acids or codons. The bias in third position is found by counting the number of each base on each position in each codon, divided by the total number of codons. The bias in the third position between G/C and A/T is calculated as sum (GC) subtracted from sum (AT), so that 100% GC in third codon position is +1 and -1 for 100% AT. A comparison of the amino acid and codon usage between the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes involved clustering genomes with similar usage based on the genes identified using the Prodigal gene prediction program (Hyatt *et al.*, 2010). The percentage codon and amino acid usage was plotted in 2D heatmaps using *gplots* in R (Ihaka and Gentleman, 1996), reordering the organisms and the amino acids/codon to show the shortest distance between them. Dendograms were used to visualise the difference in usage between different strains.

Determination of the core- and pan-genomes

The genes representative of the *Butyrivibrio* and *Pseudobutyrivibrio* core- and pan-genome were determined by performing a BLAST cut-off of 50% identity and 50% coverage of the longest gene, using the CMG-biotools (Vesth *et al.*, 2013) package with default parameters. If two proteins within a genome meet the 50%/50% cut-off, they were clustered into one protein family. Protein families were extended *via* single linkage clustering. If a protein family included proteins from all genomes in the comparison, the family was designated as a core protein family. Subset genes such as species-group shared and unique subsets of genes within individual genomes, were identified by clustering results from the pan- and core-genome calculations. The procedure is based on mathematical set theory and works with intersections, unions and complementary data-sets as described by Vesth *et al*., (2013).

The IMG system

All JGI microbial genomes are sequenced and assembled at the JGI Production Genomics Facility (USA), with subsequent finishing at various JGI partner or collaborator institutions. Automated annotation for archaeal and bacterial genomes sequenced at JGI is provided by the Genome Analysis Pipeline (Lawrence Berkeley National Laboratory, USA). The DOE-JGI Microbial Genome Annotation Pipeline (MGAP) (Huntemann *et al.*, 2015), performs structural and functional annotation of bacterial and archaeal genomes included into the Integrated Microbial Genome (IMG) system (Markowitz *et al.*, 2013). The Expert Review (ER) version of the IMG system (Markowitz *et al.*, 2009), was used for comparative analysis and revision of the *Butyrivibrio* and *Pseudobutyrivibrio* genome annotations. Genomes are identified in IMG using an internally generated unique object identifier. In addition, individual genomes are associated with the NCBI Genomes Project Identifier and taxonomic lineage *via* NCBI's Taxonomy (domain, phylum, class, order, family, genus, species, and strain), and each genome sequence dataset submitted for annotation was required to be associated with an analysis project specified in the Genomes OnLine Database (Reddy *et al.*, 2014).

For every genome, IMG incorporates its primary genome sequence information recorded in RefSeq (Tatusova *et al.*, 2014), including its organization into chromosomal replicons (finished genomes) and scaffolds or contigs (draft genomes), cross-referenced with their RefSeq Accession Identifiers. These RefSeq Gene Identifiers are linked to other NCBI resources, such as Entrez Gene (Maglott *et al.*, 2011). Genes in IMG model predicted coding sequences (CDSs) and functional RNAs, and are recorded with start/end coordinates. Genes have attributes containing data on their protein product name, functional role and participation in pathways, and relationship (sequence similarity) to other genes.

IMG automated annotation and analysis pipeline

Sequences are first compared with a database containing all the non-coding RNA genes in the Rfam database using BLAST, then sequences that have hits to genes belonging to an Rfam model are searched using the program INFERNAL (Nawrocki *et al.*, 2009). Protein-coding genes are predicted using the Prodigal v2.6.2 *ab initio* gene prediction program (Hyatt *et al.*, 2010), and models overlapping with CRISPRs and certain types of RNAs (e.g. rRNAs) are removed. Functional roles are assigned to genes based on Enzyme Nomenclature (EC numbers) (Webb, 1992), that are computed using

RPS-BLAST against the PRIAM database (Altschul *et al.*, 1997), as a complement to the native EC numbers collected *via* RefSeq. UniProt (Mulder *et al.*, 2008) is used to associate genes with additional annotations, such as COGs (Tatusov *et al.*, 2000), Pfams (Punta *et al.*, 2011), TIGRfams (Selengut *et al.*, 2007), InterPro (Apweiler *et al.*, 2001), while KEGG Orthology (KO) terms and Gene Ontology (GO) terms (Zhou and Zhu, 2011), were used to establish KO term associations. Identification of tRNAs was performed using tRNAScan-SE-1.23 (Lowe and Eddy, 1997). Ribosomal RNA genes (5*S*, 16*S* and 23*S*) are predicted using *hmmsearch* from the package HMMER 3.1b2 (Eddy, 2011). With the exception of tRNA and rRNA, all RNA gene models are synchronised with Rfam (Griffiths-Jones *et al.*, 2005). In addition, CRISPR elements are detected using the CRISPR recognition tool (CRT) (Bland *et al.*, 2007) and PILER-CR (Edgar, 2007), signal peptides are computed using SignalP (Petersen *et al.*, 2011), whereas transmembrane helices are computed using TMHMM (Möller *et al.*, 2001). All metabolic pathways were compiled using information from the MetaCyc (Caspi *et al.*, 2012) and KEGG (Kanehisa and Goto, 2000) databases. As part of IMG's online documentation, the IMG terms and pathways are detailed at http://img.jgi.doe.gov/pub/doc/img-terms.html.

CAZyme annotation

The CAZy database does not offer automated annotation of the CAZyme members for a given genome. Thus for a comprehensive and accurate annotation of the CAZyme families for the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes, the resource dbCAN was used (Yin *et al.*, 2012). The basis for using dbCAN is that it implements an automated and comprehensive annotation with clearly defined domain model signatures for all 345 CAZyme classes. In addition to the five main CAZyme classes (GHs, PLs, GTs, CEs and CBMs), additional domain modules are included in dbCAN (Bayer *et al.*, 2008): auxiliary activities (AA), dockerin, cohesion and S-layer homology (SLH) domains.

The putative proteomes of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* datasets *via* the IMG/ER pipeline, were downloaded using the JGI portal (Huntemann *et al.*, 2015). Predicted proteomes of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* datasets were subjected to automated annotation and assignment to CAZymes using the dbCAN database CAZy family-specific hidden Markov models (HMMs) (release 4.0, based on the CAZy database as of March 2015) (Yin *et al.*, 2012). The *hmmscan* functionality of the HMMER 3.0 software package (Finn *et al.*, 2011) was used to systematically screen

the dbCAN database (total of 345 CAZyme HMMs) for homologous CAZyme domains (modules) in the *Butyrivibrio* and *Pseudobutyrivibrio* datasets using a Linux-based server. The output was parsed using the following cut-offs: an *E*-value < $1e^{-3}$ of CAZymes based on family-specific HMMs was used as the cut-off for alignments shorter than 80 amino acids, while *E*-value < $1e^{-5}$ for alignments longer than 80 amino acids. These cut-off settings enabled short but significant CBM matches to be maintained. The HMM-based CAZyme annotation was chosen over the sequence similarity-based approaches to minimise dependence of the identification of putative CAZymes on the sequence similarity shared with known CAZymes. All dbCAN hits were clustered at a 100% sequence identity threshold using the CD-HIT Illumina algorithm (Huang *et al.*, 2010) to remove duplicates.

The results from the dbCAN analysis were tabulated in a matrix format and a frequency table was generated based on counts of individual CAZyme families for each of the 46 genomes. The Bray-Curtis distances (Bray and Curtis, 1957) of compositional dissimilarity and hierarchical cluster analysis using Ward's method (Ward, 1963), were calculated using the *vegdist* and *hclust* libraries in R (Ihaka and Gentleman, 1996), respectively. The resulting distances were compared between all genomes and across all 345 CAZyme families to generate heatmaps for the specific CAZyme classes as well as a holistic comparison across all classes. All statistical analyses were performed using the R Studio software (Racine, 2012). The modular architecture of CAZymes and predicted proteins with multi-modular CAZyme organisation in the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* datasets were determined by searching each query protein against the Pfam database (Finn *et al.*, 2013) using the Pfam domain search function in the IMG/ER portal. The domain structure of each putative CAZyme protein was manually verified.

2.2.23.3 RNA sequencing analysis

An overview of the RNA-seq *in silico* analysis, from analysis of the raw sequence data through to determination of differentially expressed genes, is presented in Figure 2.4.

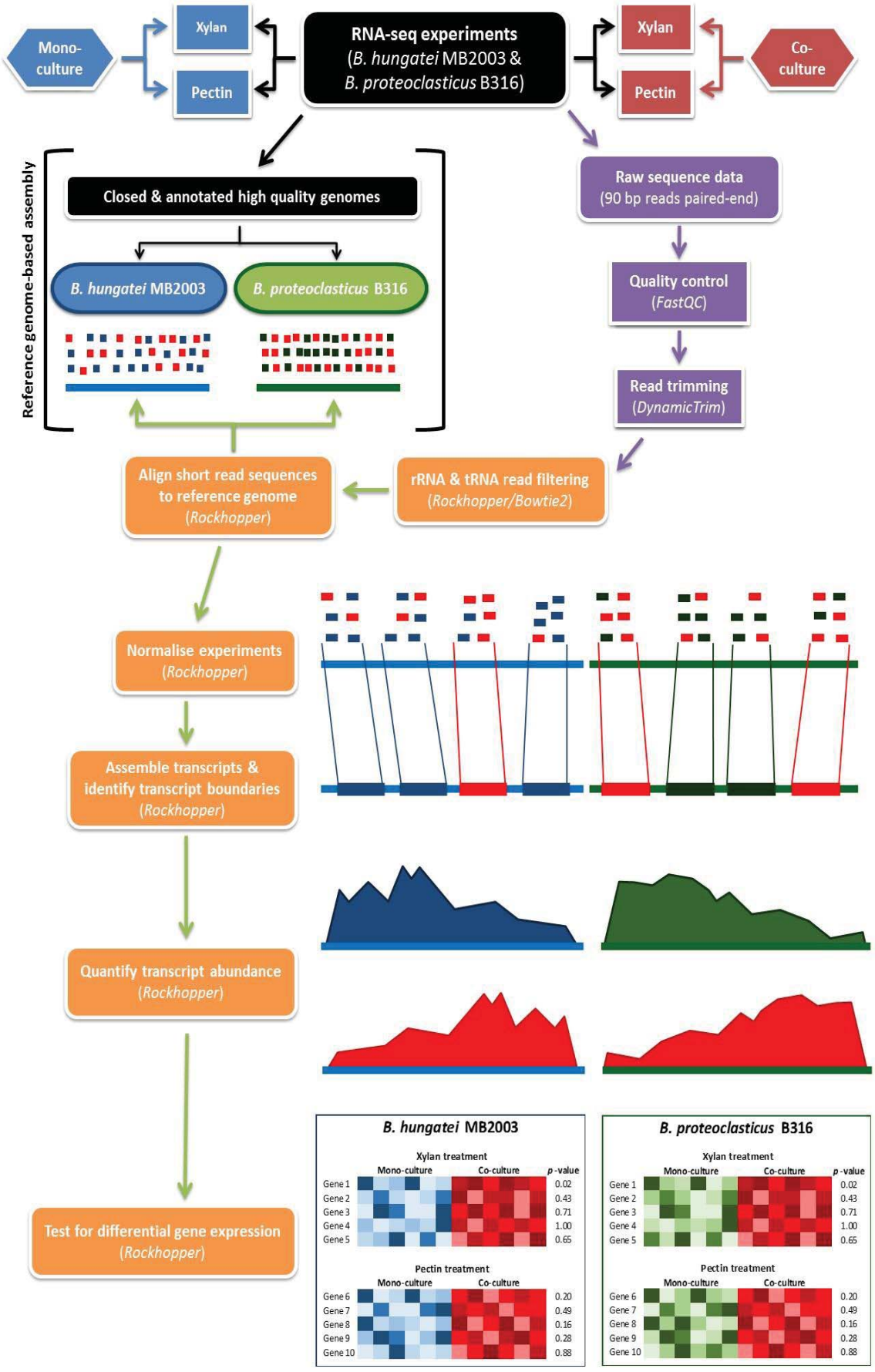


Figure 2.4. Overview of the workflow for RNA-seq *in silico* analysis. Software and programmes used at each stage are *italicised* and shown in brackets.

Illumina sequencing

High-quality total RNA was sent for sequencing on the Illumina HiSeq 2000 platform by BGI. Messenger RNA (mRNA) was enriched from total RNA using the Ribo-Zero™ rRNA Removal Kits for bacteria (Epicentre, USA) and prokaryotic strand-specific library transcriptome preparation was described above. Libraries with 200 bp inserts were constructed, paired-end sequencing was performed on each library and 90 bp reads were generated. BGI provided data that had been filtered to remove reads containing ≥ 10% unreadable bases, ≥ 20% low quality (≤ *Q*20) bases or adaptor contamination. For each sample, two FASTQ files were provided that represent the forward and reverse reads. To assess the quality of the FASTQ files, FastQC (Andrews, 2010) was used. DynamicTrim is a wrapper script and was used for quality checks, adapter trimming and trimming of paired-end and single-end data (Cox *et al.*, 2010). By default, the first 10 bp of all reads and any reads with a quality score ≤ *Q*28 were trimmed using DynamicTrim. For this analysis an error probability *p* or a quality value *Q* default setting of $p = 0.05$ (equivalent to quality score $Q \approx 13$) was applied.

Sequence read alignment and transcript assembly

Rockhopper is a system specifically designed for computational analysis of bacterial RNA-seq data (McClure *et al.*, 2013; Tjaden, 2015), and supports reference-based assembly of bacterial transcriptomes (http://cs.wellesley.edu/~btjaden/Rockhopper). The reference-based assembly from short read sequences involves aligning sequencing reads to a sequenced reference genome (Flicek and Birney, 2009; Martin and Wang, 2011). Reference-based assembly was preferred because it is a fast and relatively precise approach, but mainly because high-quality genome sequences were available for both *B. hungatei* MB2003 and *B. proteoclasticus* B316. Rockhopper was run on each set of FASTQ reads representing mono- and co-culture growth of *B. hungatei* MB2003 and *B. proteoclasticus* B316 on xylan and pectin conditions separately (Figure 2.4). The system utilises Bowtie 2 (Langmead and Salzberg, 2012), with default parameters, for the alignment of sequence reads in order to make an estimate of gene expression directly from alignment results. Illumina sequence datasets were used, and tables containing the coordinates of the tRNA and rRNA genes as well as the protein coding genes within both *B. hungatei* MB2003 and *B. proteoclasticus* B316 genomes were created. The tRNA genes and non-coding RNAs (ncRNA) were manually removed from the datasets and reads that did not align to rRNA, tRNA and ncRNA sequences were saved to a FASTQ

file for further analysis. For the reference-genome based assembly of the transcriptomes, Rockhopper was run on each set of FASTQ reads, using default parameters with the allowed mismatches and minimum seed length parameters for the transcripts were set to 0.02 and 0.33, respectively. In order to make the read counts comparable among the different samples analysed (or replicates of samples), normalization was performed for each individual gene by upper quartile normalization.

Differential gene expression

The Rockhopper output containing the total number of aligned reads, also calculates the reads per kilobase of gene per million reads mapped (RPKM), a common measure for quantifying gene expression. RPKM is a measure that sums the number of reads for a gene and divides by the gene's length and the total number of reads. Rockhopper reports the expression level of each transcript using RPKM, except that instead of dividing by the total number of reads, Rockhopper divides by the upper quartile of gene expression. Rockhopper then tests for differential expression by first obtaining a smoothed estimate of the variance for each gene that was expressed *via* local regression. A null hypothesis test was then performed based on a negative binomial distribution model (Robinson and Smyth, 2008; Anders and Huber, 2010). This is a statistical test that utilises a *p*-value to model the probability of observing a transcript's expression levels in different conditions (in this case mono- and co-culture) by chance. Because multiple tests are used to determine differential gene expression, *Q*-values (or *p*-adjusted values) were also reported that control for the false discovery rate (FDR) using the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995). Because Rockhopper was run for each set of FASTQ separately, the output data was grouped and compared based on the treatments (xylan or pectin) relative to the growth conditions (mono- versus co-culture gene expression) for *B. hungatei* MB2003 and *B. proteoclasticus* B316 separately. The cut-off used for a gene to be considered statistically differentially expressed (DEG) in each bacterium between the mono- and co-culture environments, was a *Q*-value ≤ 0.05 and a $\geq$ 2-fold $\log_2$ change in expression.

Statistical analysis of RNA-seq data

Two forms of multidimensional scaling (MDS) (Clarke, 1993) were used to analyse the RNA-seq data; a non-metric MDS (NMDS) (Faith *et al.*, 1987; Minchin, 1987), and a metric scaling in the form of Principal Coordinate Analysis (PCoA) (Gower, 1966). Stress (standardised residual sum of squares) was measured for the relationship between the original distance matrix associated with the sample replicates and the distances in dimensional space when carrying out NMDS analysis (Gower and Legendre, 1986). Bray-Curtis distances (Anderson, 2001) of the normalised counts were used to measure pairwise dissimilarity using *vegdist* function of the package *vegan* of R software (R Core Team, 2015). The relative power of NMDS depends on the resemblance method (dissimilarity measure) used as the basis of the analysis. Because the dissimilarity matrix was constructed using the Bray-Curtis distances, the analysis is more focused on compositional changes in sample identities, as opposed to the differences in the relative abundance of transcripts per sample (Petersen *et al.*, 2011). The *pcoa* function in the *ape* R package and the *metaMDS* function of the *vegan* package were used to create the configuration of all points in the two dimensional space.

Permutation multivariate analysis of variance uses distance matrices to define sources of variation among samples (Zapala and Schork, 2006). This resembles the 'homogeneity of variances' assumption required when doing univariate (one gene at a time) or nonparametric multivariate ANOVA (or "one-way MANOVA"). So if the *p*-value is large, that means that the 'homogeneity of variances' assumption is satisfied (Warton *et al.*, 2012). Correspondence analysis (CA) (Greenacre, 1993) was also utilised to explore and visualise the association between groupings of samples based on the culture condition (mono- (C1) and co-culture (C2)), treatment (xylan (X) and pectin (P)) and associated genes for B316 and MB2003. The R package *FactoMineR* was used to explore CA, while clustering, data manipulations and additional analysis of high-dimensional data sets were considered using packages such as *doBy* and *mixOmics*. Genes with zero or near-zero variances were first removed from the analyses using the *nearZeroVar* function of the *mixOmics* R package.

A Chi-square test of independence was carried out on the raw transcript data to assess whether paired observations on two variables, were independent of each other. In other words, test whether genes deemed differentially expressed were independent of culture condition (mono- and co-culture) and treatment (xylan and pectin). The goal was to obtain a global view of the gene expression data that is useful for interpretation. The

CA procedure utilised here can be simplified as follows. The frequencies for any row to column combination of categories are related to all other combinations based on the marginal frequencies, intimately connected with the Chi-square (χ^2) test (statistic) of independence. This yields a conditional expectation, very similar to an expected Chi-square value. Once obtained, these values are normalised, and a process much like PCA (Principal component analysis) defines the lower-dimensional solutions. The maximum number of new dimensions is equal to the minimum (number of rows, number of columns) - 1. Similar to the 'total variation' among the original variables in a PCA, in CA, the "total inertia" is decomposed. The "total inertia" is a measure of the total association between the rows and columns of the given contingency table, and equals to χ^2-statistic divided by the grand total. The low dimensions then simultaneously relate the rows and columns in a single graph, and the axes are called "principal axes". It should be noted here that the graph should be thought of as two different overlaid plots, one for rows and the other for columns.

To interpret the CA plot, each point should be observed relative to the origin or centroid, where each gene is plotted respective to its association within each culture condition/treatment category (PC1, PC2, XC1 and XC2). Therefore points that are in similar directions are positively associated and genes on opposite sides of the origin are negatively associated, and genes that are furthest from the origin exhibit the strongest associations. Meaning that for a positive association like up-regulation of a gene in a particular condition, the two points will lie on the same side of the centroid, and the larger the distance from the centroid, the stronger the association. A negative association like down-regulation will cause the column-point and the row-point to lie on opposite sides of the centroid. The CA analysis also identified the relative associations between genes in addition to which genes were most or least up- or down-regulated overall.

When interest is focussed on relationships or associations between a row point/category (PC1, PC2, XC1 and XC2) and all column categories (genes), a rather more complicated method that first draws a line ('direction vector') on the plot (of all row and column points in 2 or higher dimensions) through the origin and the point corresponding to the row point in question may be considered. This is followed by perpendiculars to this direction vector being dropped from each column position/point on the plot. The relative relationships associated with the column categories towards the row category of interest can be read off by traversing the direction vector through the row point of interest, looking at the positions at which the perpendiculars from the 'column'

points intersect it. A column point with an intersection on the direction vector on the same side (say, +ve side) of the origin as the row point of interest occurs more often in the sample than the average in the row categories overall, whereas one on the other side (say, -ve side) of the origin occurs less frequently than the average. In addition, the further from the origin on the +ve side such an intersection occurs, the greater the frequency of the column point in the sample. Conversely, the further out on the -ve side an intersection falls, the less frequent the column point in the sample. A matrix of similarity between each of the row categories and column categories can be constructed as the distances (+ve or -ve) of positions of column points on the above direction vectors in the multidimensional CA configuration. These similarities or measures of association can then be used to create 'Clustered Image Maps' (CIM) and 'Network graphs' (González *et al.*, 2012) to highlight the relationship between the rows and columns.

The Kruskal-Wallis (KW) analysis of variance by ranks of the *p*-values (Hollander *et al.*, 2013), was then performed on the normalised counts independently using the *kruskal.test* function in the R package. This analysis incorporated the ANOVA *t*-test and the Kruskal-Wallis (KW) pair-wise analysis of variance by ranks, which is a non-parametric test investigating whether samples originated from the same distribution (Conover, 1980). Assumptions were made when using this statistical test, however the KW test was selected because it applies to situations where two or more independent samples are compared with varying sample sizes.

Chapter 3

Selection and characterisation of candidate *Butyrivibrio* strains

3.1 Introduction

Because of their phenotypic and metabolic similarities, all *Butyrivibrio* and *Pseudobutyrivibrio* strains were once assigned to *Butyrivibrio fibrisolvens*. However, 16*S* rRNA sequencing of *Butyrivibrio fibrisolvens* strains derived from the rumen showed that they are genetically diverse and hence they were reclassified into two genera (*Butyrivibrio* and *Pseudobutyrivibrio)* containing multiple species. This chapter describes the phenotypic and genotypic characterisation of a number of *Butyrivibrio* and *Pseudobutyrivibrio* strains. In addition to two *Butyrivibrio* strains whose genome sequences have been completed (*B. hungatei* MB2003 (Palevich, 2011) and *B. proteoclasticus* B316 (Kelly *et al.*, 2010)), a set of strains was chosen from a large experiment that cultivated plant-adherent bacteria from rumen contents obtained from fistulated Friesian dairy cows (Noel, 2013). In this work, dilutions of rumen contents were incubated in a variety of media, resulting in 828 strains identified by partial 16*S* rRNA sequences as pure cultures (Noel, 2013). Of those, 219 were assigned to the genus *Butyrivibrio*.

Taxonomic analysis of the 219 *Butyrivibrio* strains based on the V1-V3 region of the 16*S* rRNA via Basic Local Alignment Sequence Tool (BLAST) and the RDP database (Cole *et al.*, 2009) grouped the *Butyrivibrio* strains into 19 separate clusters (Noel, 2013). Based on this analysis, representative strains were chosen for revival and confirmatory 16*S* rRNA gene sequence analysis based on the V1-V3 regions and alignment with the SILVA database via ARB (Pruesse *et al.*, 2007), followed by MEGA6 (Tamura *et al.*, 2013) and Geneious (Kearse *et al.*, 2012) software packages. The preliminary taxonomic analysis of the *Butyrivibrio* strains proved that approximately a quarter of the *Butyrivibrio* strains can be placed in the *B. fibrisolvens* group while the remainder formed a large cluster in which species boundaries were difficult to determine. However, a wider interpretation of this analysis suggests that there are three distinct *Butyrivibrio* species clusters, *B. fibrisolvens*, *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3.

As a prelude to genome sequencing of the *Butyrivibrio* strains, phenotypic and genotypic characterisations were carried out to ensure the strains selected were representative of the variability of this genus in the rumen environment. The phenotyping methods used were cell morphology, motility, carbon source utilisation and fermentation end products, while 16*S* ribosomal RNA gene sequencing and Pulsed Field Gel Electrophoresis (PFGE) were used for genotyping.

3.2 Selection and taxonomic assignment of *Butyrivibrio* strains

A selection of 5-10 strains from each of *B. fibrisolvens* and *Butyrivibrio* sp. groups 2 and 3 resulted in the characterisation of 30 strains in total. The 30 cultures selected were revived from frozen stocks using RM02 medium, genomic DNAs were extracted from each of the *Butyrivibrio* cultures, and almost full-length 16*S* rRNA gene sequences (>1400 bp) were obtained from PCR amplification products. The taxonomic assignment of the strains into species groups was done initially using BLASTN searches of the "16*S* ribosomal RNA sequences (Bacteria and Archaea)" non-redundant Database using the Megablast function, followed by comparing their 16*S* rRNA gene sequences with sequences from characterised strains and type strains (Table 3.1).

A phylogenetic tree was constructed using the almost full-length 16*S* rRNA gene sequences of all 30 *Butyrivibrio* strains, and 20 additional strains from previous work, including *Butyrivibrio hungatei* MB2003 and *Pseudobutyrivibrio xylanivorans* MA3014 (Palevich, 2011), *B. hungatei* (Kopecny *et al.*, 2003), *B. fibrisolvens* (Forster *et al.*, 1997), *B. proteoclasticus* B316 (Attwood *et al.*, 1996; Moon *et al.*, 2008), and *Pseudobutyrivibrio* species (Van Gylswyk *et al.*, 1996; Kopecny *et al.*, 2003) (Figure 3.1). The 16*S* rRNA sequences were aligned using ClustalW (Thompson *et al.*, 2003) and their phylogenetic relationships determined using MEGA6 software (Tamura *et al.*, 2013) using three methods; Neighbor-Joining (Saitou and Nei, 1987), Minimum Evolution (Rzhetsky and Nei, 1992) and Maximum Likelihood (Akaike, 1998), and corrected manually for alignment ambiguities. All three methods gave similar results and the Maximum Likelihood tree is shown (Figure 3.1). The *Butyrivibrio* sequences grouped into three main species clusters: Cluster 1 contained the sequences of type strains of *B. proteoclasticus* (B316) and *B. hungatei* (JK615) and *Butyrivibrio* species 3 group strains, Cluster 2 contained the *Butyrivibrio* species 2 strains with no representative type strains, while Cluster 3 contained the *B. fibrisolvens* type strain (D1) and closely related strains.

These relationships are well supported by Bootstrap analysis, Cluster 1 showing 99% bootstrap support, and Cluster 3 with 97% bootstrap support. This suggests that the *Butyrivibrio* 16*S* rRNA gene sequences can be divided broadly into the *B. fibrisolvens*, *Butyrivibrio* sp. 2 and the larger *Butyrivibrio* sp. 3 groups. Despite their similar phenotypic characteristics, *Butyrivibrio* and *Pseudobutyrivibrio* clearly differ from each other and from other rumen bacteria at the 16*S* rRNA gene level.

Table 3.1. Taxonomic assignments and 16*S* rRNA gene sequence analyses of *Butyrivibrio* strains.

Species	Strain	Taxonomic assignment (species (cluster) groups) *	Top BLASTN match to an type strain		
			Genus species	Strain	Identity match (%)
B. fibrisolvens	MD2001	*B. fibrisolvens* (B18)	*Butyrivibrio fibrisolvens*	D1	99
	AB2020	*B. fibrisolvens* (B18)	*Butyrivibrio fibrisolvens*	D1	99
	FE2007	*B. fibrisolvens* (B18)	*Butyrivibrio fibrisolvens*	D1	98
	ND3005	*B. fibrisolvens* (B18)	*Butyrivibrio fibrisolvens*	D1	98
	YRB2005	*B. fibrisolvens* (B18)	*Butyrivibrio fibrisolvens*	D1	98
	WTE3004	*B. fibrisolvens* (B18)	*Butyrivibrio fibrisolvens*	D1	98
	MC2013	*B. fibrisolvens* (B16)	*Butyrivibrio fibrisolvens*	D1	96
	NC3005	*B. fibrisolvens* (B17)	*Butyrivibrio fibrisolvens*	D1	93
Butyrivibrio sp. 2	XPD2002	*Butyrivibrio* sp. 2 (B13)	*Butyrivibrio hungatei*	JK615	96
	AE3006	*Butyrivibrio* sp. 2 (B11)	*Butyrivibrio proteoclasticus*	B316	96
	MB2005	*Butyrivibrio* sp. 2 (B11)	*Butyrivibrio proteoclasticus*	B316	96
	AC2005	*Butyrivibrio* sp. 2 (B13)	*Butyrivibrio proteoclasticus*	B316	96
	LC3010	*Butyrivibrio* sp. 2 (B12)	*Butyrivibrio hungatei*	JK615	95
	FB3011	*Butyrivibrio* sp. 1 (B9)	*Butyrivibrio proteoclasticus*	B316	96
	FC2001	*Butyrivibrio* sp. 2 (B13)	*Butyrivibrio hungatei*	JK615	97
	WCD3002	*B. proteoclasticus* (B11)	*Butyrivibrio proteoclasticus*	B316	96
	VCD2006	*B. proteoclasticus* (B11)	*Butyrivibrio hungatei*	JK615	97
	AD3002	*B. proteoclasticus* (B11)	*Butyrivibrio hungatei*	JK615	96
	WCD2001	*B. proteoclasticus* (B11)	*Butyrivibrio hungatei*	JK615	96
	AE3004	*B. proteoclasticus* (B11)	*Butyrivibrio proteoclasticus*	B316	95
Butyrivibrio sp. 3	XPD2002	*B. hungatei* (B3)	*Butyrivibrio hungatei*	JK615	97
	AE2015	*B. hungatei* (B2)	*Butyrivibrio proteoclasticus*	B316	96
	XBB1001	*B. proteoclasticus* (B6)	*Butyrivibrio proteoclasticus*	B316	98
	XPD2006	*Butyrivibrio* sp. 1 (B10)	*Butyrivibrio hungatei*	JK615	97
	VCB2001	*Butyrivibrio* sp. 1 (B9)	*Butyrivibrio proteoclasticus*	B316	96
	FCS006	*Butyrivibrio* sp. 1 (s1)	*Butyrivibrio hungatei*	JK615	96
	NC2007	*Butyrivibrio* sp. 1 (B8)	*Butyrivibrio proteoclasticus*	B316	96
	AE3009	*Butyrivibrio* sp. 1 (B10)	*Butyrivibrio hungatei*	JK615	96
	FCS014	*Butyrivibrio* sp. 1 (B10)	*Butyrivibrio proteoclasticus*	B316	97
	AE2032	*Butyrivibrio* sp. 1 (B10)	*Butyrivibrio proteoclasticus*	B316	95

* Taxonomic assignments represented first by the species group, followed by the designated cluster group based on Noel (2013), in brackets.

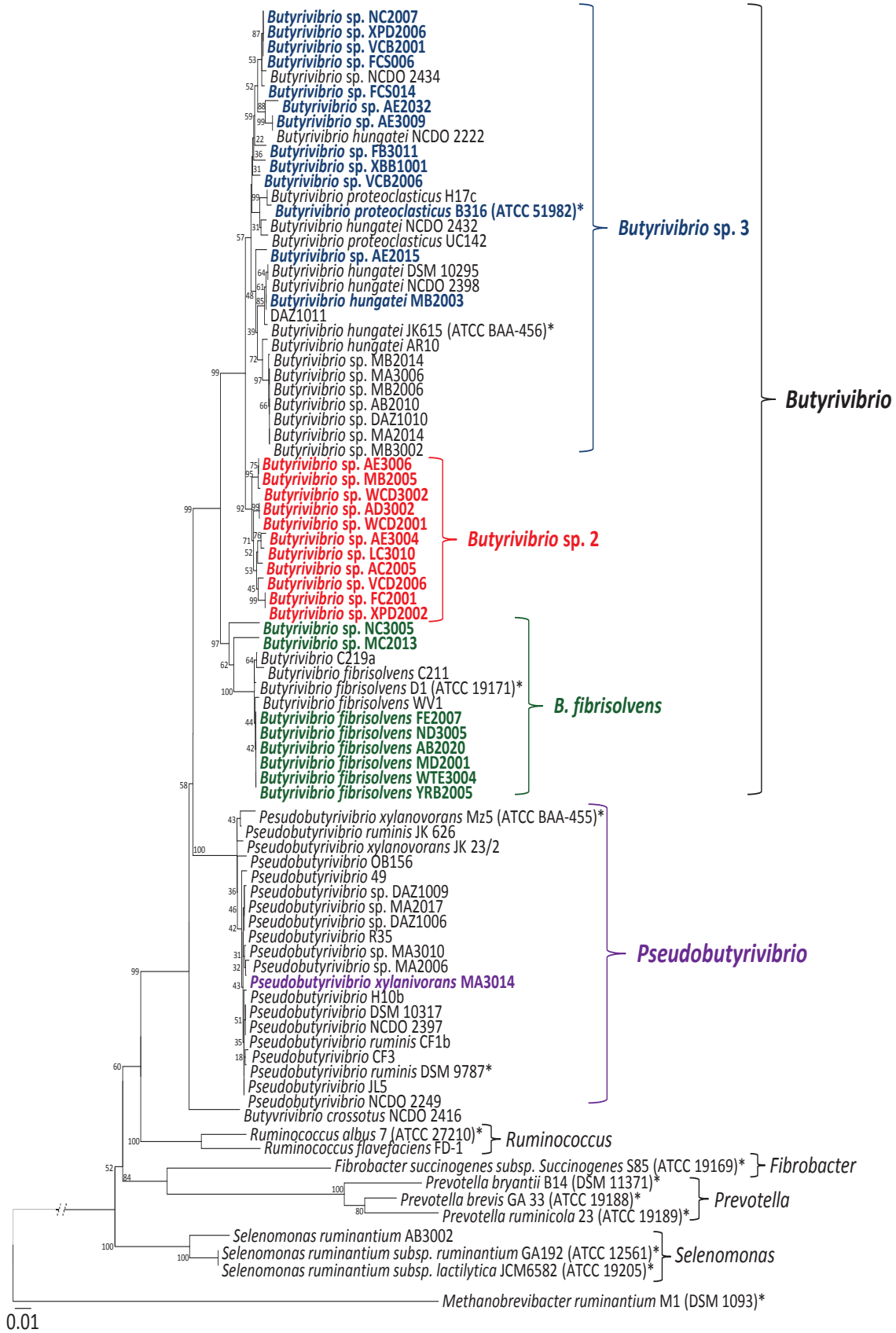


Figure 3.1. Phylogenetic tree of *Butyrivibrio* strains based on 16*S* rRNA full-length gene sequence data. Bar, indicates 0.01 nucleotide substitutions per site or 1% difference in nucleotide sequence. * indicates type strains. Strains shown in bold and coloured were selected for further characterisation.

3.3 Morphological characterisation

Thirty *Butyrivibrio* strains were revived in BY medium containing soluble oligosaccharides as the major carbon source. Microscopic evaluation of the cultures confirmed that each of the 30 strains were of a single cell type (Table 3.2). Colony morphologies were recorded after growth on plates (Table 3.3). Out of 30 strains, three (VCB2001, FCS014, AE2032) were slow-growing in the liquid and on solid media. All of the 30 strains displayed cell morphologies consistent with that of other *Butyrivibrio* strains and similar to *B. proteoclasticus* B316, which has slightly curved rods that form short chains or clumps, pairs of cells or long chains at a low frequency (Attwood *et al.*, 1996).

Despite being described originally as being Gram-negative based on the standard staining procedure (Bartholomew and Mittwer, 1952), *Butyrivibrio* cell walls have a Gram-positive cell wall structure (Cheng and Costerton, 1977; Beveridge, 2001). It is thought that *Butyrivibrio* appear Gram-negative when stained (Figures 3.2-3.4) because their cell walls become thinner (i.e. they have a thinner peptidoglycan layer) as they reach stationary phase (Beveridge, 2001).

Table 3.2. Growth and morphological characterisation of the *Butyrivibrio* strains.

Species	Strain	Culture revival observations		
		Morphological description	Gram Stain	Overnight growth
B. fibrisolvens	MD2001	short rods in chains & clumps	- ve	yes
	AB2020	clumping straight rods in chains	- ve	yes
	FE2007	small rods in chains	- ve	yes
	ND3005	small rods in chains	- ve	yes
	YRB2005	small rods in chains	- ve	yes
	WTE3004	straight rods in chains	- ve	yes
	MC2013	short rods in chains	- ve	yes
	NC3005	very long large rods in chains	- ve	yes
Butyrivibrio sp. 2	XPD2002	short rods	- ve	yes
	AE3006	thick short rods in chains	- ve	yes
	MB2005	short rods	- ve	yes
	AC2005	short round rods	- ve	yes
	LC3010	short rods in clumps	- ve	yes
	FB3011	short rods in clumps	- ve	yes
	FC2001	short rods in clumps	- ve	yes
	WCD3002	curved rods	- ve	yes
	VCD2006	short rods	- ve	yes
	AD3002	thick short rods	- ve	yes
	WCD2001	thick short rods	- ve	yes
	AE3004	short rods in small chains	- ve	yes
Butyrivibrio sp. 3	VCB2006	short rods in chains	- ve	yes
	AE2015	long rods in chains	- ve	yes
	XBB1001	short rods in chains & clumps	- ve	yes
	XPD2006	rods in chains	- ve	yes
	VCB2001 *	short rods in chains	- ve	**no ***
	FCS006	small rods in pairs	- ve	yes
	NC2007	clumps of rods	- ve	yes
	AE3009	short rods in clusters	- ve	yes
	FCS014 *	small rods in clumps	- ve	**no ***
	AE2032 *	chains of short rods	- ve	**no ***
Controls	B316	slightly curved rods in chains	- ve	yes
	23	short rods in chains	- ve	yes
	2B	cocci	+ ve	yes

* indicates slow growing cultures. *Butyrivibrio proteoclasticus* B316, *Prevotella ruminicola* 23 (Avguštin *et al.*, 1994) and *Streptococcus bovis* 2B (Iverson and Millis, 1976) cultures were used as controls.

Table 3.3. Colony morphology of the *Butyrivibrio* strains.

Species	Strain	Form	Elevation	Margin	Size (mm)	Colour	Surface	Opacity	Texture
B. fibrisolvens	MD2001	filamentous	convex	filamentous	10	white	smooth	opaque	butyrous
	AB2020	circular	convex	entire	3	white	smooth	opaque	butyrous
	FE2007	circular	convex	entire	2	white	smooth	opaque	butyrous
	ND3005	circular	convex	entire	1	white	smooth	opaque	butyrous
	YRB2005	circular	convex	entire	3	white	smooth	opaque	butyrous
	WTE3004	circular	convex	entire	3	white	smooth	opaque	butyrous
	MC2013	irregular	umbonate	undulate	2	white	rough	opaque	butyrous
	NC3005	irregular	umbonate	undulate	1	white	rough	opaque	butyrous
Butyrivibrio sp. 2	XPD2002	circular	convex	entire	2	white	smooth	opaque	butyrous
	AE3006	circular	convex	entire	1	white	smooth	opaque	butyrous
	MB2005	circular	convex	entire	1	white	smooth	opaque	butyrous
	AC2005	circular	convex	entire	1	white	smooth	opaque	viscous
	LC3010	circular	convex	entire	3	white	smooth	opaque	butyrous
	FB3011	circular	convex	entire	2	white	smooth	opaque	butyrous
	FC2001	circular	convex	entire	6	white	smooth	opaque	butyrous
	WCD3002	circular	convex	entire	1	white	smooth	opaque	butyrous
	VCD2006	circular	convex	entire	2	white	smooth	opaque	butyrous
	AD3002	circular	convex	entire	1	white	smooth	opaque	butyrous
	WCD2001	circular	convex	entire	10	white	smooth	opaque	butyrous
	AE3004	circular	convex	entire	2	white	smooth	opaque	butyrous
Butyrivibrio sp. 3	VCB2006	circular	convex	entire	1	white	smooth	opaque	butyrous
	AE2015	circular	convex	entire	3	white	smooth	opaque	butyrous
	XBB1001	circular	convex	entire	3	white	smooth	opaque	butyrous
	XPD2006	circular	convex	entire	3	white	smooth	opaque	butyrous
	VCB2001 *	circular	convex	entire	3	white	smooth	opaque	butyrous
	FCS006	circular	convex	entire	1	white	smooth	opaque	butyrous
	NC2007	circular	convex	entire	2	white	smooth	opaque	butyrous
	AE3009	irregular	umbonate	undulate	1	white	rough	opaque	butyrous
	FCS014 *	circular	convex	entire	1	white	smooth	opaque	butyrous
	AE2032 *	circular	convex	entire	0.5	white	smooth	opaque	butyrous
Controls	B316	circular	convex	entire	3	white	smooth	opaque	butyrous
	23	circular	convex	entire	3	white	smooth	opaque	butyrous
	2B	circular	convex	entire	3	white	smooth	opaque	butyrous

Controls were *Butyrivibrio proteoclasticus* B316, *Prevotella ruminicola* 23 and *Streptococcus bovis* 2B. * indicates slow growing cultures.

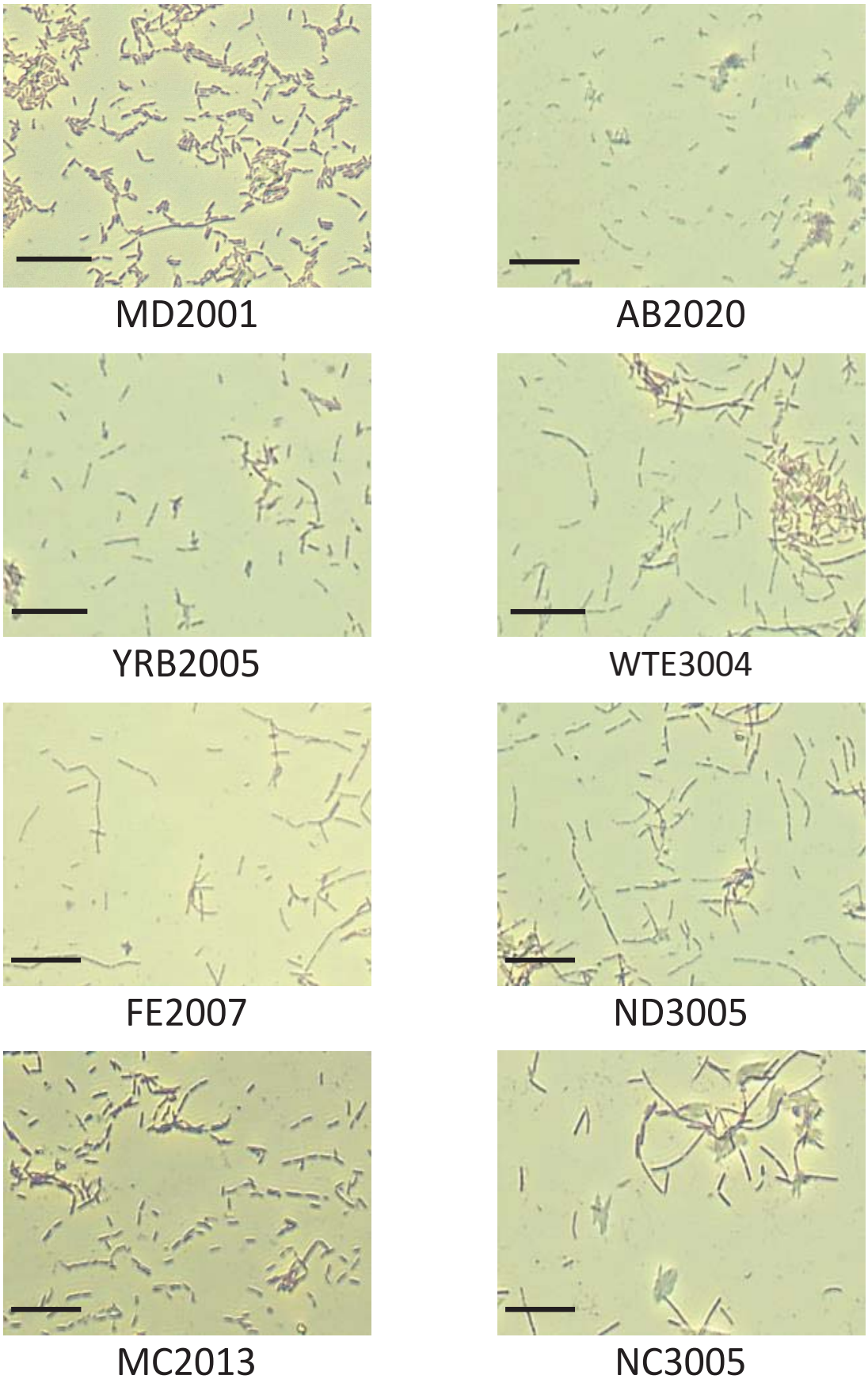


Figure 3.2. Light micrographs of Gram-stained *B. fibrisolvens* strains grown on RM02 media containing cellobiose. Micrographs taken at 100× magnification and bar represents 10 µm.

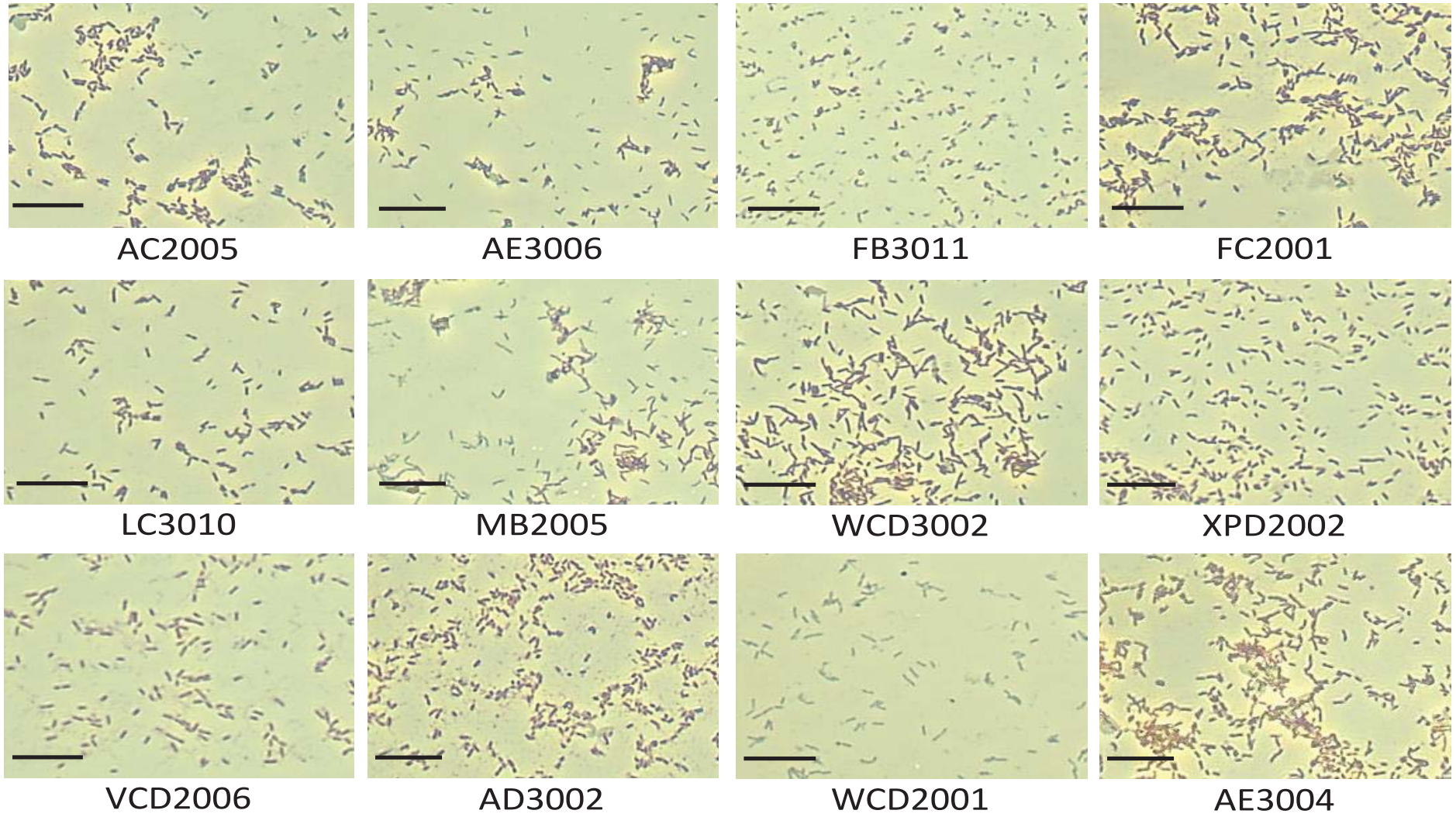


Figure 3.3. Light micrographs of Gram-stained *Butyrivibrio* sp. 2 strains grown on RM02 media containing cellobiose. Micrographs taken at 100× magnification and bar represents 10 µm.

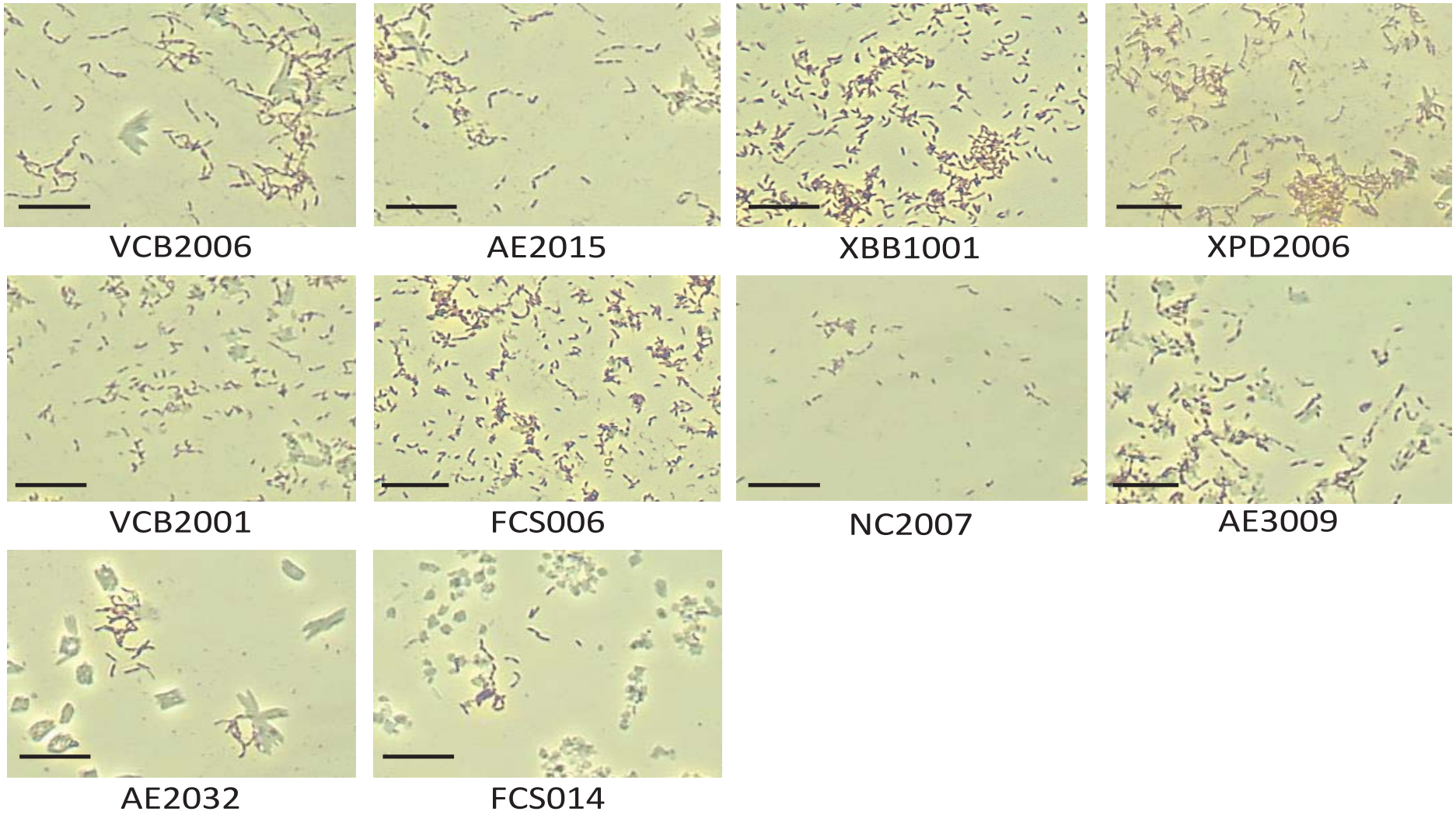


Figure 3.4. Light micrographs of Gram-stained *Butyrivibrio* sp. 3 strains grown on RM02 media containing cellobiose. Micrographs taken at 100× magnification and bar represents 10 µm.

3.4 Genome size estimation and *ApaI* restriction fragment length polymorphism (RFLP) assessed by PFGE

The intact genomic DNA of each *Butyrivibrio* strain was first analysed by PFGE to investigate the number of replicons (Figures 3.5-3.7), given that *Butyrivibrio* genomes have been previously shown to contain megaplasmids and secondary chromosomes (Palevich, 2011; Yeoman *et al.*, 2011). Furthermore, the RFLP pattern analysis using the rare-cutting restriction enzymes *Apa*I (recognition site 5'GGGCC|C3') or I-*Ceu*I (recognition site 5'TAGGGATAA|-CAGGGTAAT3'; within the 23*S* rRNA gene) (McClelland, 1981; Belfort and Roberts, 1997); followed by PFGE was undertaken (Appendix I Figures A1.1-A1.3 and A1.4-A1.6, respectively) to estimate the genome sizes of each strain (Appendix I Table A1.1) and to investigate RFLPs, including the number of rRNA operons.

Successful *Apa*I restriction and PFGE analyses were achieved for 6 of 8 *B. fibrisolvens*, 8 of 12 *Butyrivibrio* sp. 2 and 8 of 10 *Butyrivibrio* sp. 3 strains, and showed that all of the strains analysed have distinct restriction fragment patterns (Appendix I Figures A1.1-A1.3). In some *Butyrivibrio* strains endogenous DNAse activity caused general degradation of genomic DNA, resulting in smeared restriction patterns. The occurrence of the recognition site of the I-*Ceu*I restriction enzyme only within the 23*S* rRNA gene, provides additional information on the copy number of rRNA operons present in a genome. Furthermore, because these operons are most commonly the cross-over points in recombinations that lead to chromosomal rearrangements, it may be possible to derive these from the I-*Ceu*I restriction pattern (Marshall and Lemieux, 1992; Liu *et al.*, 1993). The PFGE analysis of I-*Ceu*I digests were successful with 7 of 8 *B. fibrisolvens*, 10 of 12 *Butyrivibrio* sp. 2 and 7 of 10 *Butyrivibrio* sp. 3 strains (Appendix I Figures A1.4-A1.6). Multiple bands were observed within the successful I-*Ceu*I digests, confirming the presence of multiple rRNA operons within each *Butyrivibrio* genome analysed. Also, common banding patterns among *Butyrivibrio* strains within the same taxonomic group were observed. For example, *B. fibrisolvens* strains YRB2005, WTE3004, AB2020 and MB2005, AC2005 as well as LC3010 and AE3004 from *Butyrivibrio* sp. 2 groups shared common band patterns within their respective species groups.

Extrachromosomal bands were identified in the PFGE analysis of uncut DNA from *B. fibrisolvens* strains FE2007, ND3005, YRB2005, WTE3004; *Butyrivibrio* sp. 2 strains XPD2002, AC2005, LC3010, WCD2001, AE3004; *Butyrivibrio.* sp. 3 strains VCB2006, AE2015, XBB1001, XPD2006, VCB2001, FCS006, NC2007, AE3009. The patterns produced from the PFGE analysis of uncut genomic DNAs indicate that extrachromosomal elements are a common characteristic of *Butyrivibrio*. The estimated sizes of the extrachromosomal elements (megaplasmids and/or plasmids) are summarised in Appendix I Table A1.1. The strains with the most extrachromosomal elements are *B. fibrisolvens*: MD2001 (3) and ND3005 (2), *Butyrivibrio* sp. 2: AC2005 (4), LC3010 (3) and AE3004 (2), and *Butyrivibrio* sp. 3: VCB2006 (2), XBB1001 (2) and NC2007 (2). The largest extrachromosomal element found was estimated at 869.2 Kb from XPD2002, the smallest was 99.1 Kb in size from AE3006, both strains are from *Butyrivibrio* sp. 2, while the average size range was 300-500 Kb. For the majority of *Butyrivibrio* strains, the draft genomes are in too many contigs to make meaningful comparisons from the genomes alone. The exception is *Butyrivibrio* sp. 2 strain AE3004 whose draft genome consists of six contigs, two of which match the 433.1 and 350.1 Kb bands from the PFGE analysis.

The uncut chromosomal bands were above the resolution range of the PFGE applied for analysis of the extrachromosomal elements, so their size could not be deduced (Figures 3.5-3.7). However, genome size estimations were made from by tallying all the fragment sizes from the *Apa*I and I-*Ceu*I digests (Appendix I Table A1.1). The estimated genome sizes for each strain from both *Apa*I and I-*Ceu*I digests were within +/- 500 Kb. The genome size estimates for *Butyrivibrio* range from approximately 3.5 Mb for *B. fibrisolvens* NC3005 and *Butyrivibrio* sp. 3 VCB2001, to 5.6 Mb for *Butyrivibrio* sp. 2 AC2005. This information contributes to the subsequent genome sequence analyses and will be discussed in more detail in Chapter 4.

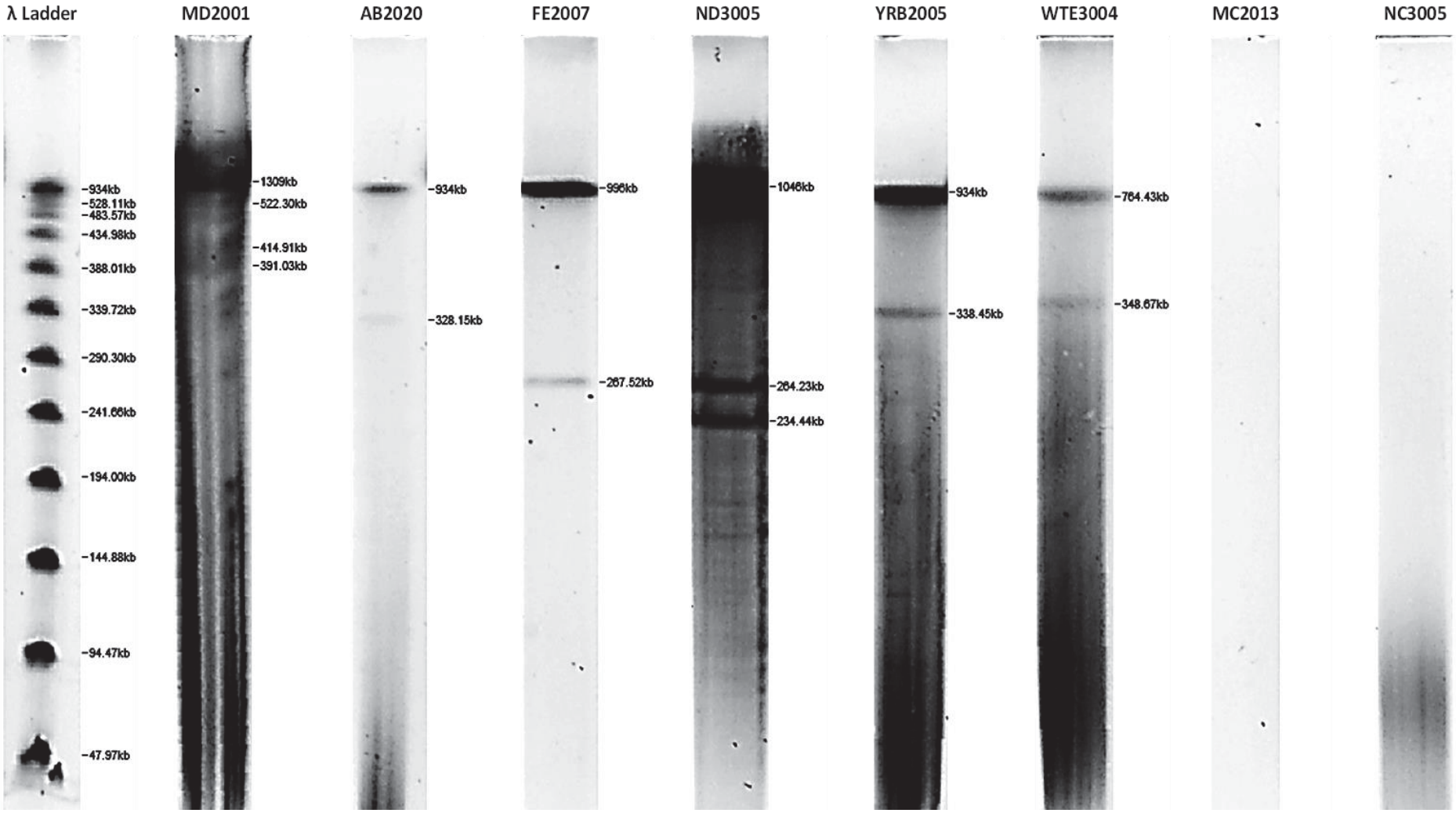


Figure 3.5. PFGE profiles of uncut *B. fibrisolvens* genomic DNAs. Genomic DNAs were separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

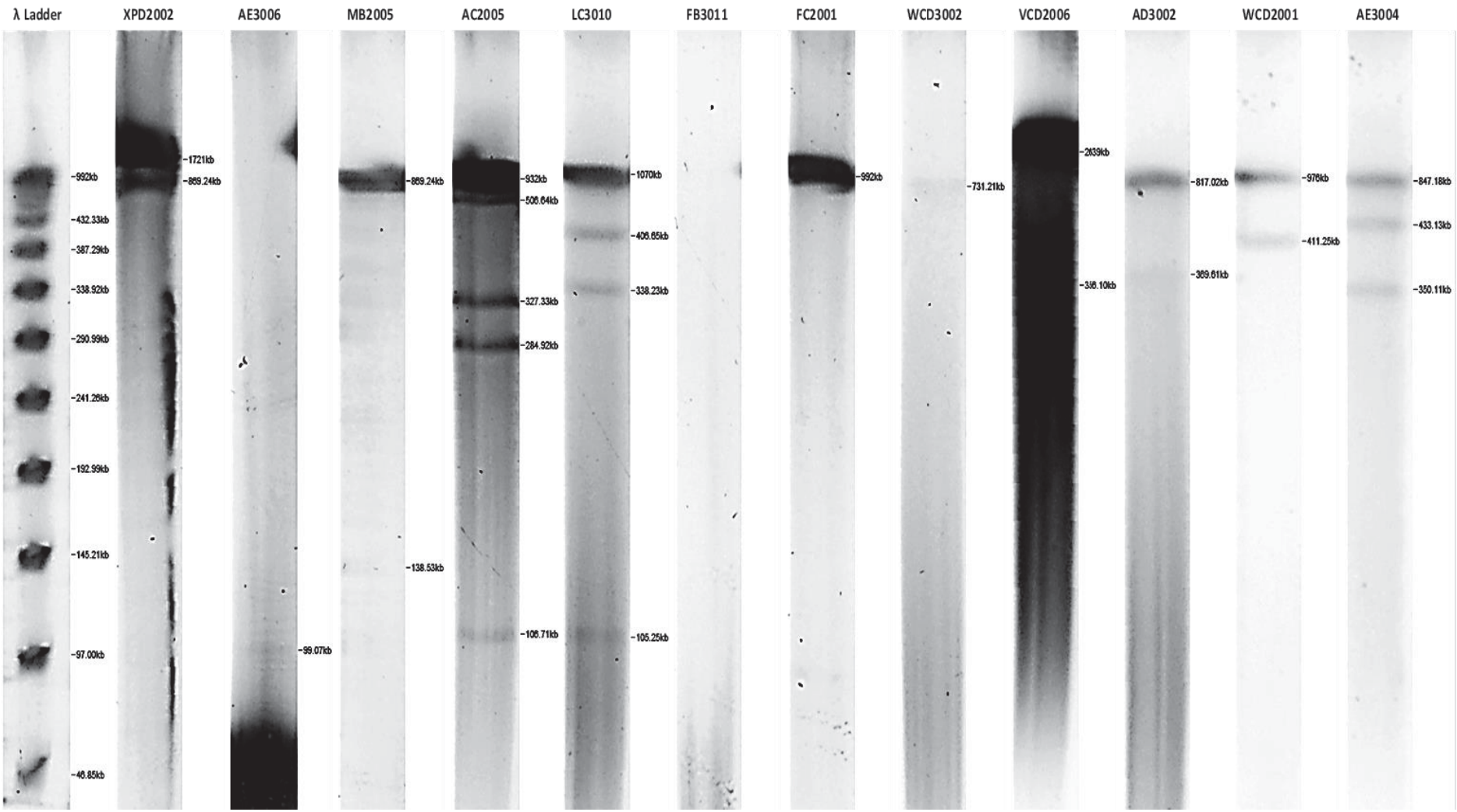


Figure 3.6. PFGE profiles of uncut *Butyrivibrio* sp. 2 genomic DNAs. Genomic DNAs were separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

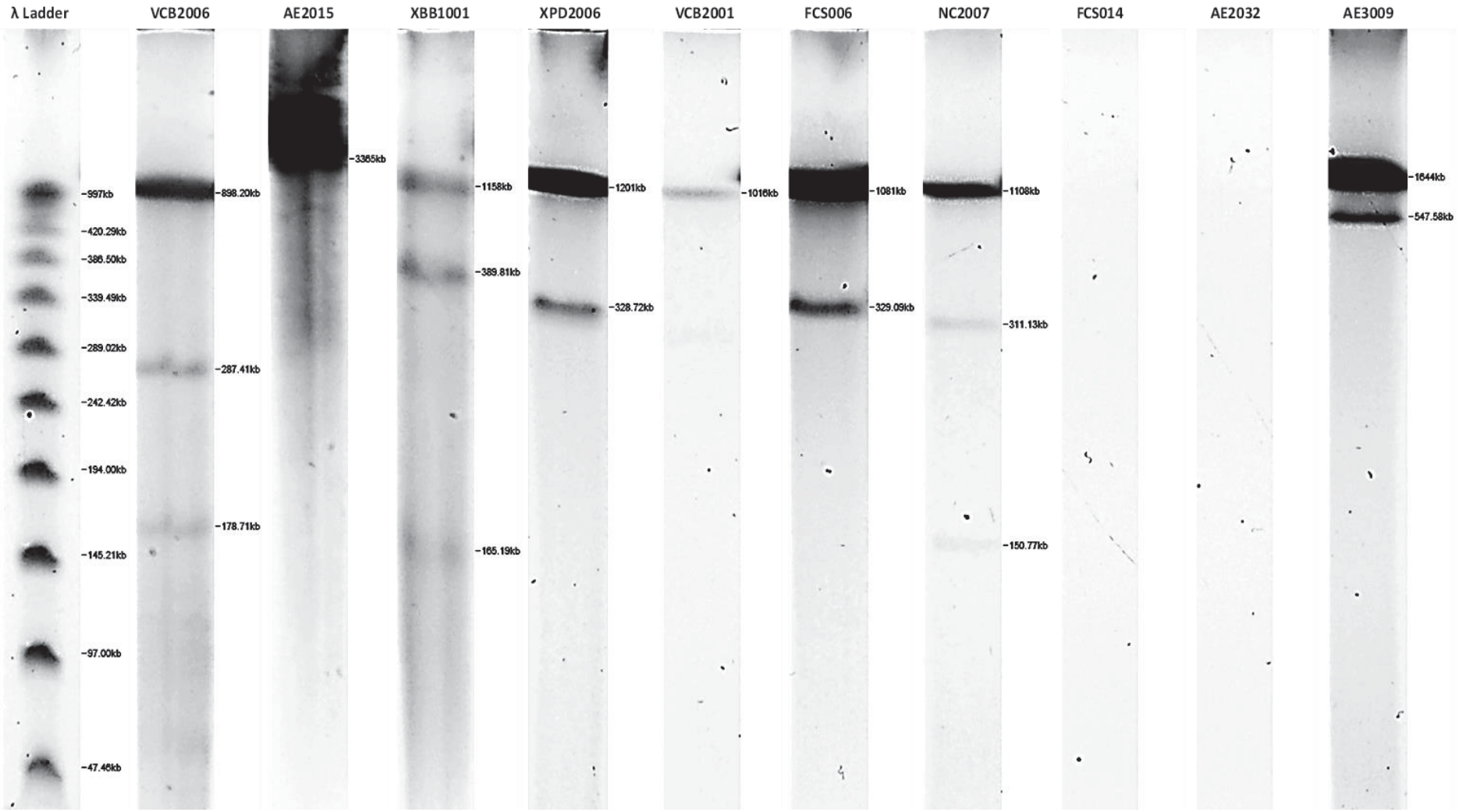


Figure 3.7. PFGE profiles of uncut *Butyrivibrio* sp. 3 genomic DNAs. Genomic DNAs were separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

3.5 Cell motility

To investigate the occurrence and function of flagella in the *Butyrivibrio* strains, a simple stab motility assay in semi-solid agar was carried out. If the cells are motile, their migration within the agar can be seen as zones of bacterial growth away from the inoculation stab. The non-motile controls showed clear, sharp lines of bacterial growth (Figure 3.8; the first three panels from the left), demonstrating the lack of movement for the known non-motile controls, *B. proteoclasticus* B316 (Attwood *et al.*, 1996), *B. hungatei* MB2003, and *Streptococcus bovis* 2B. In contrast, the fuzzy, indistinct line of bacterial growth shown by the motile control, *P. xylanivorans* MA3014 (Figure 3.8), indicates a motile phenotype, as the bacterial growth has moved from its initial single inoculation stab line outward in all directions. The motility assay found that only two *Butyrivibrio* strains, *B. fibrisolvens* MC2013 and *Butyrivibrio* sp. 2 strain FC2001, displayed a motile phenotype, whereas the remaining 28 strains were non-motile (Figures 3.9-3.11; Table 3.4).

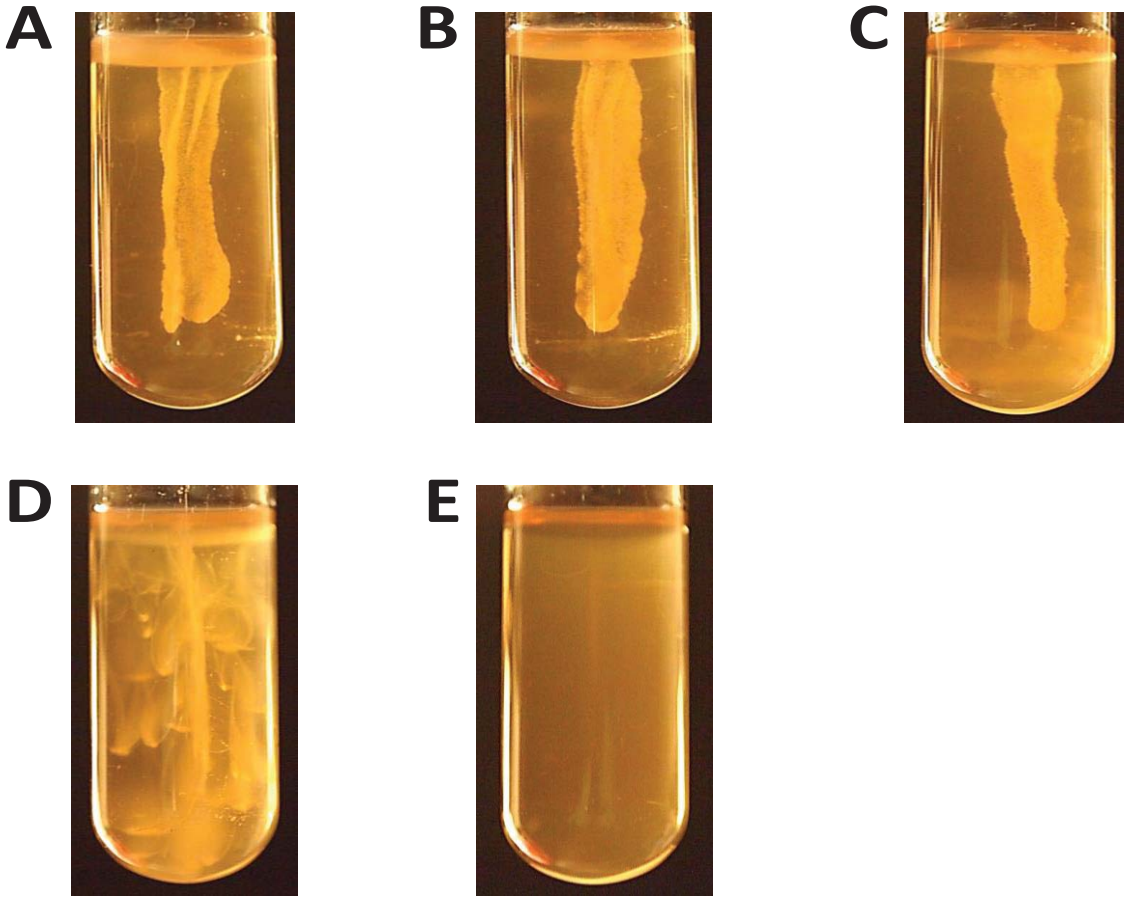


Figure 3.8. Motility assays of control strains. Motility assays of the non-motile A, *B. proteoclasticus* B316; B, *B. hungatei* MB2003; C, *Streptococcus bovis* 2B (left); the motile D, *P. xylanivorans* MA3014 (red); E, uninoculated control (far right).

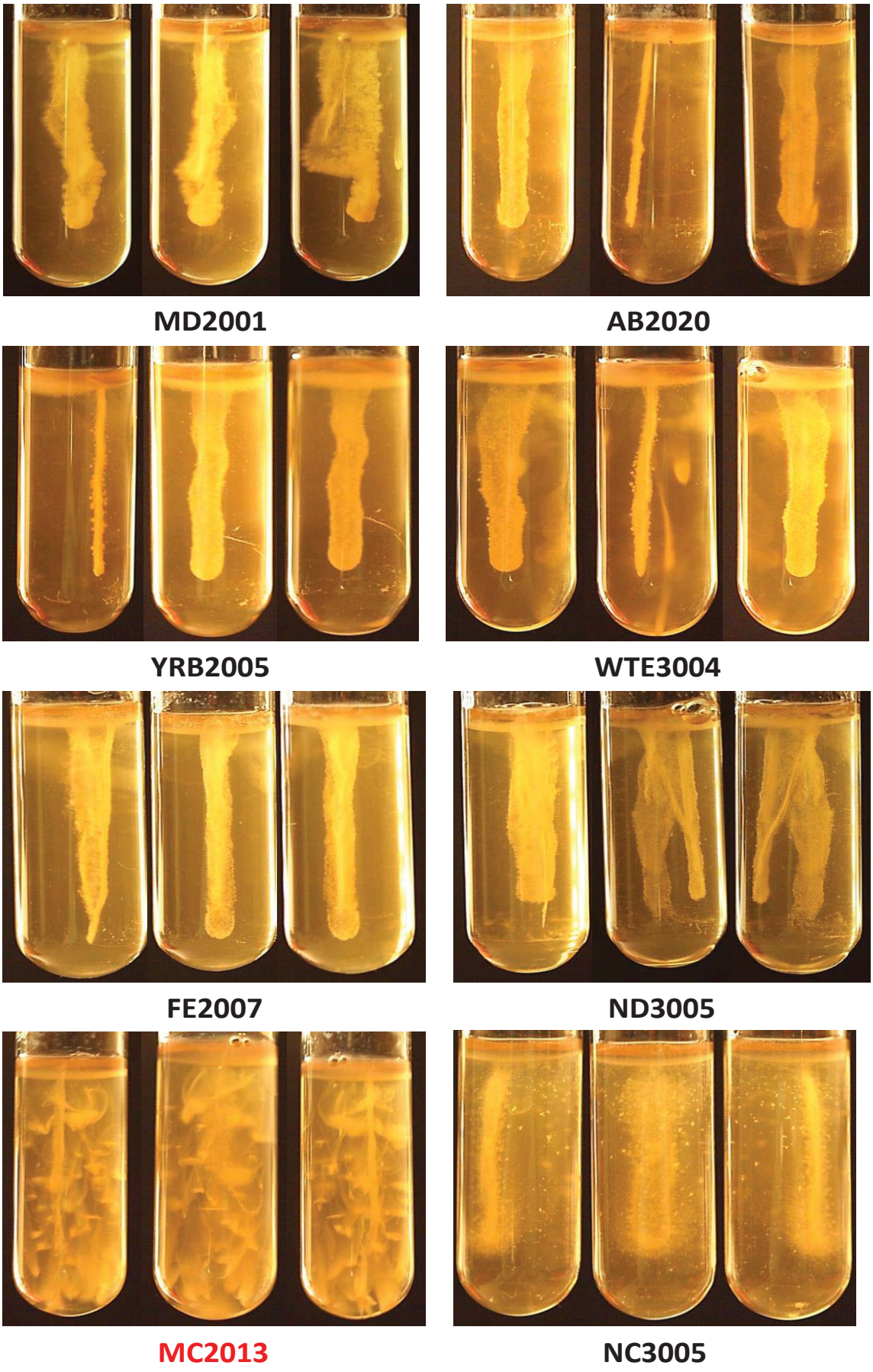


Figure 3.9. Motility assays of *B. fibrisolvens* strains. Motile strains coloured in red.

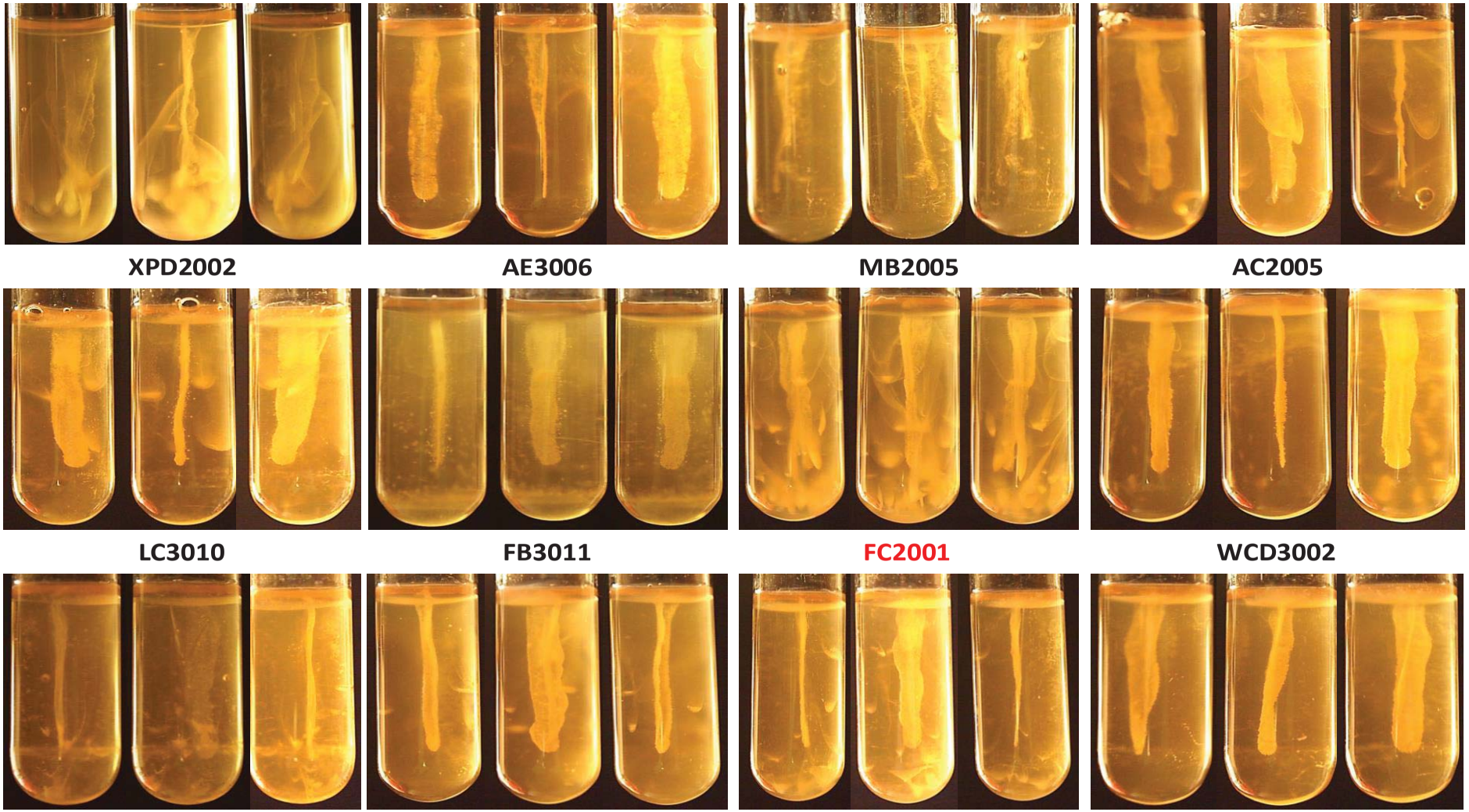


Figure 3.10. Motility assays of *Butyrivibrio* sp. 2 strains. Motile strains coloured in red.

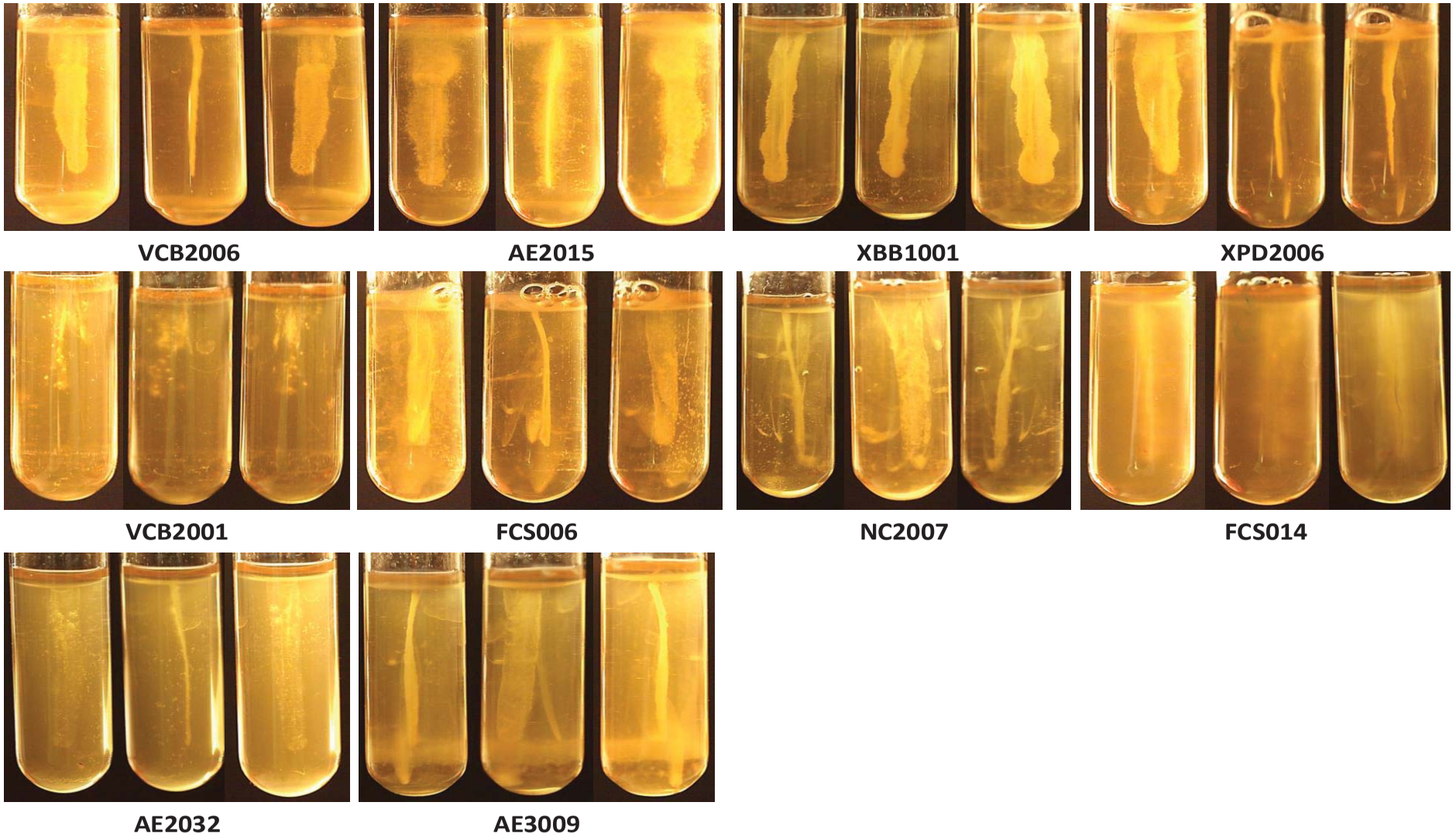


Figure 3.11. Motility assays of *Butyrivibrio* sp. 3 strains. Motile strains coloured in red.

Table 3.4. Motility assay of *Butyrivibrio* strains.

Species		Strain	Motility *
B. fibrisolvens		MD2001	-
		AB2020	-
		FE2007	-
		ND3005	-
		YRB2005	-
		WTE3004	-
		MC2013	+
		NC3005	-
Butyrivibrio sp. 2		XPD2002	-
		AE3006	-
		MB2005	-
		AC2005	-
		LC3010	-
		FB3011	-
		FC2001	+
		WCD3002	-
		VCD2006	-
		AD3002	-
		WCD2001	-
		AE3004	-
Butyrivibrio sp. 3		VCB2006	-
		AE2015	-
		XBB1001	-
		XPD2006	-
		VCB2001	-
		FCS006	-
		NC2007	-
		AE3009	-
		FCS014	-
		AE2032	-
Controls	*B. hungatei*	MB2003	-
	B. proteoclasticus	B316	-
	P. xylanivorans	**MA3014**	+
	S. bovis	2B	-

*, motile phenotype was indicated with a +; -, signified a non-motile phenotype.

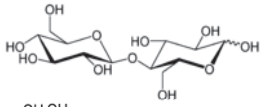
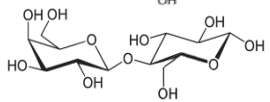
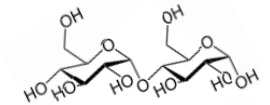
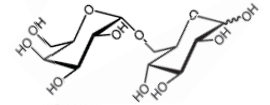
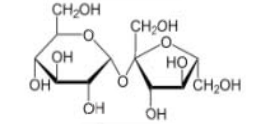
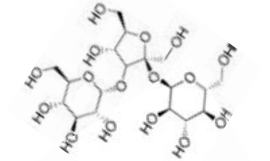
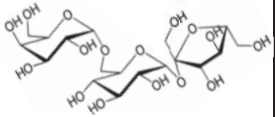
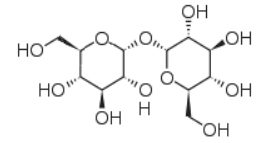
3.6 Carbon source utilisation by *Butyrivibrio* strains

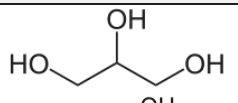
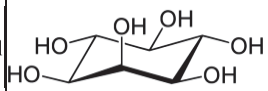
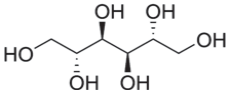
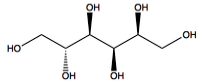
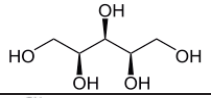
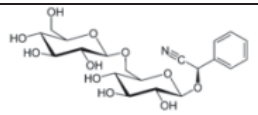
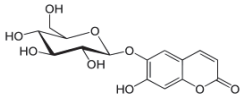
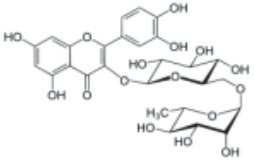
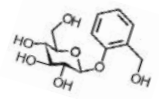
Growth assays were performed on 32 commonly used soluble and insoluble carbohydrate substrates for phenotyping, to determine the overall carbon-source utilisation patterns for each of the 30 *Butyrivibrio* strains (Table 3.5). Growth on soluble substrates was monitored by measuring OD_{600nm} of the cultures. Among the soluble substrates, the majority of the strains were able to utilise monosaccharides (galactose and glucose), trisaccharides (melezitose and trehalose), and the sugar alcohols (glycerol, myo-inositol, mannitol and sorbitol). There were distinct differences between the three *Butyrivibrio* species groups. *Butyrivibrio* sp. 2 displayed preferential utilisation of fructose, ribose and xylose among the monosaccharides and the trisaccharide trehalose. *B. fibrisolvens* and *Butyrivibrio* sp. 2 also preferred cellobiose and lactose among the disaccharides. *Butyrivibrio* sp. 2 and 3 were utilisers of the disaccharide sucrose. The soluble substrates that were not utilised by any of the tested strains were mannose, rhamnose, maltose, amygdalin, rutin and salicin. All of the substrates tested (including amygdalin, rutin and salicin), were selected based on their use as carbon sources in early studies that characterised the phenotype of different *Butyrivibrio* strains (Bryant, 1986; Hespell, 1992b).

To verify the utilisation of insoluble carbon sources, the growth and metabolic activity of cultures were assessed by measuring the concentration of VFAs produced, using gas chromatography. The gas chromatography method used did not measure formate and lactate production. An increase in concentration of acetate and butyrate correlated with bacterial growth, whereas BCVFAs and propionate were present only in trace amounts and have not been included in this analysis (Figure 3.12). From the VFA analysis, six *B. fibrisolvens* strains (MD2001, AB2020, FE2007, ND3005, YRB2005 and WTE3004) displayed similar growth patterns on insoluble substrates. In contrast, the remaining two *B. fibrisolvens* strains, MC2013 and NC3005, showed different substrate utilisation patterns. Strains MB2005, AC2005, LC3010, FB3011 and FC2001 from *Butyrivibrio* sp. 2, and FCS006, FCS014 and AE2032 of *Butyrivibrio* sp. 3, were not able to grow on any of the insoluble substrates according to the VFA production analysis. Overall, *Butyrivibrio* sp. 3 strains were able to utilise a wider range of insoluble substrates compared to *Butyrivibrio* sp. 2.

Table 3.5. Carbon source utilisation of the *Butyrivibrio* strains.

Substrate	Chemical Structre	*B. fibrisolvens*								*Butyrivibrio* sp. 2												*Butyrivibrio* sp. 3									
		MD2001	AB2020	FE2007	ND3005	YRB2005	WTE3004	MC2013	NC3005	XPD2002	AE3006	MB2005	AC2005	LC3010	FB3011	FC2001	WCD3002	VCD2006	AD3002	WCD2001	AE3004	VCB2006	AE2015	XBB1001	XPD2006	VCB2001 *	FCS006	NC2007	AE3009	FCS014 *	AE2032 *
Controls	Positive Control	+	+	+	+	+	+	-	+	+	-	-	+	+	+	+	-	+	+	+	+	+	+	++	+	+	+	+	-	-	-
	Innoculation Control	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Uninnoculated Media	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Monosaccharides	**Arabinose**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
	Fructose	+	-	-	+	-	+	++	+	+	+	++	++	++	+	++	+	++	++	++	+	-	+	-	-	-	-	-	-	-	-
	Galactose	+	+	+	+	+	+	+	+	+	+	++	+	+	+	+	-	++	+	+	+	+	+	+	+	-	-	+	+	-	-
	Glucose	++	++	++	++	++	++	+	++	++	+	++	++	++	++	++	++	++	++	++	++	+	++	+	+	+	+	+	+	-	-
	Mannose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Rhamnose	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Ribose	+	++	-	-	++	++	++	++	++	-	-	++	++	++	++	-	++	+	++	++	+	+	-	-	-	+	-	+	-	-
	Xylose	+	+	-	-	+	+	-	+	+	+	++	+	+	-	+	+	+	+	+	++	-	+	-	-	-	-	-	-	-	-

	Substrate	Chemical Structre	*B. fibrisolvens*								*Butyrivibrio* sp. 2												*Butyrivibrio* sp. 3									
			MD2001	AB2020	FE2007	ND3005	YRB2005	WTE3004	MC2013	NC3005	XPD2002	AE3006	MB2005	AC2005	LC3010	FB3011	FC2001	WCD3002	VCD2006	AD3002	WCD2001	AE3004	VCB2006	AE2015	XBB1001	XPD2006	VCB2001 *	FCS006	NC2007	AE3009	FCS014 *	AE2032 *
Disaccharides	Cellobiose		++	++	++	++	+	++	-	-	++	+	-	+	++	++	++	-	++	++	++	++	-	++	-	-	+	-	-	-	-	-
	Lactose		+	++	++	++	++	++	++	-	++	++	-	++	++	++	++	-	++	++	++	++	-	-	-	-	-	-	+	-	-	-
	Maltose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Melibiose		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-
	Sucrose		-	+	-	-	-	+	+	-	+	+	+	+	+	+	+	-	++	+	+	-	+	+	+	+	+	+	+	+	-	-
Trisaccharides	Melezitose		++	++	++	+	++	++	+	++	++	+	++	++	++	+	++	++	++	++	++	++	-	++	++	+	+	+	+	+	+	-
	Raffinose		-	++	-	-	+	++	+	-	-	-	-	-	++	++	-	-	-	-	+	++	-	-	-	+	+	+	+	+	-	-
	Trehalose		+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	++	+	+	+	-	+	-	+	-	+	-	-	-	-

Substrate		Chemical Structre	*B. fibrisolvens*								*Butyrivibrio* sp. 2												*Butyrivibrio* sp. 3									
			MD2001	AB2020	FE2007	ND3005	YRB2005	WTE3004	MC2013	NC3005	XPD2002	AE3006	MB2005	AC2005	LC3010	FB3011	FC2001	WCD3002	VCD2006	AD3002	WCD2001	AE3004	VCB2006	AE2015	XBB1001	XPD2006	VCB2001 *	FCS006	NC2007	AE3009	FCS014 *	AE2032 *
Sugar Alcohols	Glycerol		++	++	+	+	++	+	+	+	-	+	++	+	++	+	+	+	++	+	+	-	+	+	+	+	-	+	+	-	-	-
	myo-Inositol		+	+	+	+	+	+	+	+	+	+	++	+	+	+	+	+	+	+	++	+	+	+	+	+	-	+	+	+	-	-
	Mannitol		+	+	+	+	+	++	+	+	+	+	+	+	++	+	+	+	+	-	+	++	+	+	+	+	+	+	+	+	-	-
	Sorbitol		+	++	++	++	++	++	+	++	++	+	++	++	++	+	++	++	++	++	++	++	+	++	+	+	+	+	+	+	-	-
	Xylitol		-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glycosides	Amygdalin		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Esculin		-	++	-	-	++	-	-	-	-	-	-	-	++	-	-	-	-	-	-	++	-	-	-	-	-	-	-	+	-	-
	Rutin		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Salicin		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

ΔOD_{600nm} readings of 0.5-1.0 were scored as ++, 0.2-0.5 scored as +, and 0-0.2 scored as -. * indicates slow growing cultures.

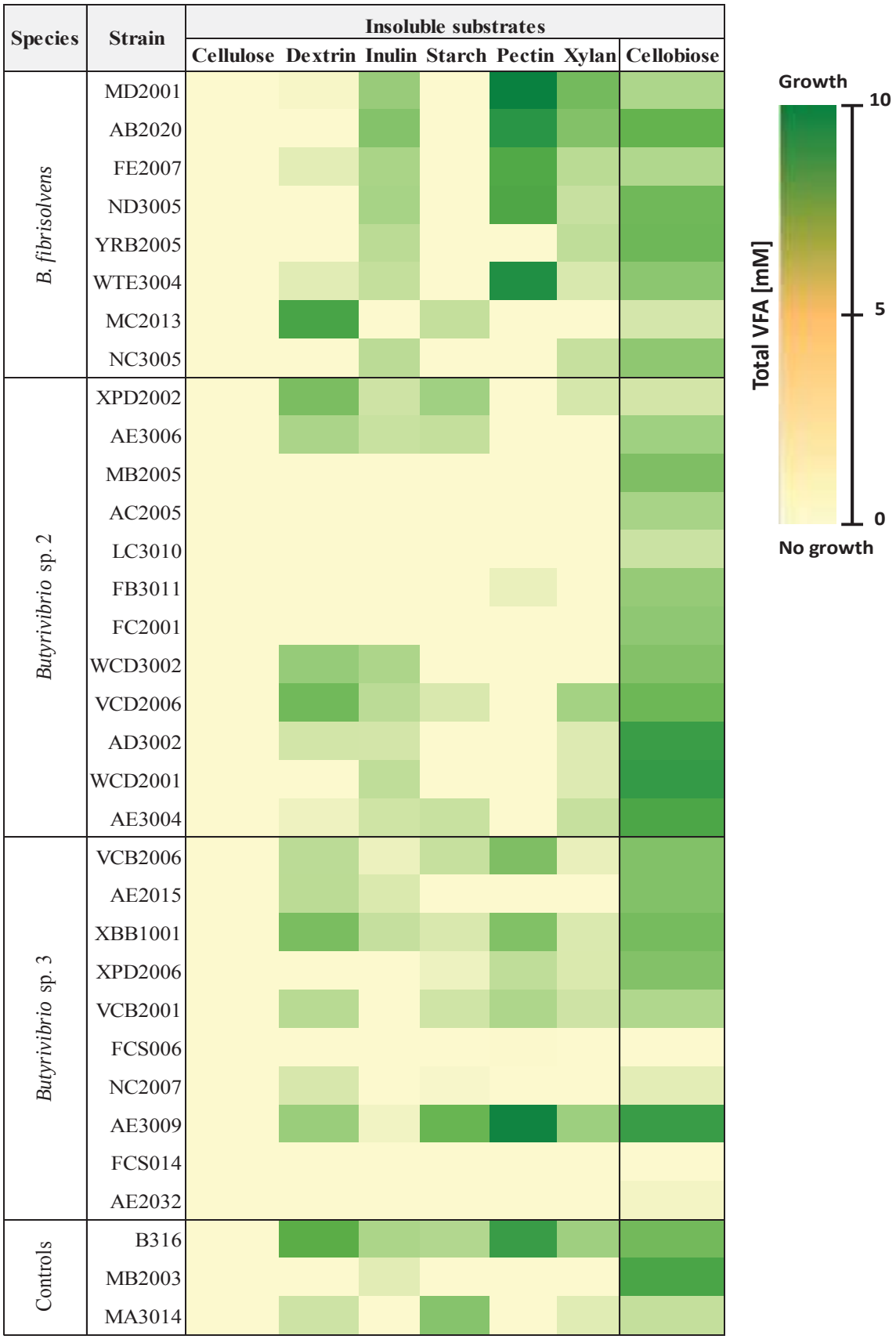


Figure 3.12. End product analysis of *Butyrivibrio* strains grown on insoluble carbon sources. End products are expressed as the average of the total VFA concentrations (mM) of two duplicate cultures with background inoculum and insoluble substrate VFAs in media controls subtracted.

The VFA data also showed that *Butyrivibrio* utilisation of polysaccharides and their ability to assimilate the degradation products of these polysaccharides was variable with clear differences identified between species groups based on the phylogenetic analysis. For example, the *B. fibrisolvens* MD2001, AB2020, FE2007, ND3005, YRB2005 and WTE3004 strains displayed similar growth patterns, in particular their ability to utilise pectin, xylan and xylose, whereas *B. fibrisolvens* sp. MC2013 and NC3005 expressed different growth patterns that coincide with their phylogenetic placements within this cluster (Figure 3.1). All *Butyrivibrio* sp. 2 strains could utilise xylose and strains AE3004, XPD2002 and VCD2006 displayed similar growth patterns on all insoluble substrates analysed that is also reflected by their phylogenetic similarities. The *Butyrivibrio* sp. 3 strains B316, VCB2006, XBB1001 and AE3009 that grouped closely at the genome and phylogenetic levels displayed similar growth patterns across all insoluble substrates analysed, and curiously only AE2015 could utilise xylose. Almost all *Butyrivibrio* strains analysed could utilise the glucose which is a predominant degradation products of starch, dextrin and inulin, and all *Butyrivibrio* sp. 2 strains could utilise fructose that is also abundant in inulin. The monosaccharides arabinose, rhamnose and galactose are among the main components of pectin and xylan, but none of the strains tested could utilise arabinose (except XPD2006) and rhamnose, while galactose utilisation was observed in almost all strains tested. Furthermore only the *Butyrivibrio* sp. 3 strain VCB2001 that could not utilise either galactose or xylose, was able to grow on both pectin and xylan. This suggests that the galactose and xylose are important components of the complex polysaccharides xylan and pectin, and that the composition and distribution of side chains are important for their degradation and utilisation.

Interestingly, different SCVFAs were predominant metabolic end-products of different insoluble growth substrates. Butyrate and acetate were the predominant VFA products and appeared to be characteristic for cultures grown on dextrin, inulin, starch, xylan and pectin. The characteristic VFA product for cultures grown on cellobiose was butyrate. Our analysis of the VFA data has identified distinct differences in growth patterns, not only between the different species groups, but also amongst individual *Butyrivibrio* strains.

3.7 Analysis of fermentation end products after growth on cellobiose

The fermentation end products released after growth on cellobiose for all 30 *Butyrivibrio* strains were analysed. For this analysis, an additional method was used to derivatise formate, lactate and succinate, which could not be detected using the VFA method from the previous section. Cellobiose was used as substrate for this experiment, as it is a disaccharide that all strains could utilise (Table 3.6). As expected, most of the *Butyrivibrio* strains produced large amounts of formate (average 11 mM), acetate (average 5 mM) and *n*-butyrate (average of 6 mM). Eleven strains (MD2001, AB2020, YRB2005, WTE3004, AE3009, AC2005, LC3010, FC2001 and AE2015) and the slow growers (AE2032 and FCS014) also produced lactate (average 3 mM). In addition, glucose was detected, especially in slow growing cultures, which indicates incomplete fermentation of the cellobiose substrate.

Table 3.6. End product analysis of *Butyrivibrio* strains.

Species	Strain	End Product (mM) [A]							
		Ethanol	Lactate	Formate	Acetate	Propionate	*n*-Butyrate	Succinate	*iso*-valeric
B. fibrisolvens	MD2001	1.2	5.7	6.5	4.4	1.0	3.6	0.1	0.0
	AB2020	1.8	5.5	15.3	6.2	1.1	8.3	0.0	0.0
	FE2007	1.4	1.1	8.6	4.5	1.0	4.1	0.0	0.0
	ND3005	1.4	1.6	14.3	5.9	1.1	7.5	0.0	0.0
	YRB2005	0.8	8.6	9.5	4.7	1.0	5.3	0.0	0.0
	WTE3004	0.8	11.0	7.8	4.5	1.0	4.7	0.0	0.0
	MC2013	0.0	0.2	1.2	3.0	1.1	1.9	0.0	0.0
	NC3005	1.0	0.3	6.0	4.5	1.1	2.4	0.1	0.0
Butyrivibrio sp. 2	XPD2002	1.5	2.7	14.3	5.3	1.1	7.2	0.0	0.0
	AE3006	1.5	0.2	10.6	3.5	1.1	6.3	0.0	0.0
	MB2005	1.4	0.3	14.2	4.6	1.2	7.7	0.0	0.0
	AC2005	1.2	8.9	9.2	4.1	1.1	5.9	0.1	0.0
	LC3010	1.0	4.8	7.7	3.0	1.1	4.1	0.0	0.0
	FB3011	1.3	0.9	5.3	5.8	3.4	2.8	0.4	0.0
	FC2001	0.9	3.2	15.0	5.9	1.4	9.4	0.1	0.0
	WCD3002	0.9	0.3	6.4	3.9	1.1	3.0	0.0	0.0
	VCD2006	0.4	0.2	1.0	2.2	0.7	1.0	0.1	0.0
	AD3002	0.9	2.2	14.1	4.4	1.2	10.4	0.0	0.0
	WCD2001	1.1	2.3	12.9	4.8	1.2	10.7	0.0	0.0
	AE3004	1.4	0.2	19.6	5.5	1.2	10.4	0.1	0.0
Butyrivibrio sp. 3	VCB2006	1.0	2.3	13.3	5.2	1.2	6.7	0.1	0.0
	AE2015	1.3	10.5	14.1	5.8	1.2	7.0	0.1	0.0
	XBB1001	0.8	0.3	21.0	7.0	1.2	10.5	0.1	0.0
	XPD2006	0.7	2.5	8.9	3.0	1.1	4.7	0.1	0.0
	VCB2001 *	0.9	0.8	5.3	3.2	1.1	3.3	0.1	0.0
	FCS006	1.1	0.2	6.7	3.6	1.1	4.2	0.0	0.0
	NC2007	1.4	0.3	5.4	3.6	1.2	3.7	0.1	0.0
	AE3009	1.4	3.5	19.2	6.8	1.2	9.1	0.1	0.0
	AE2032 *	0.0	5.5	0.0	0.0	0.0	0.0	0.1	0.0
	FCS014 *	0.0	4.0	0.0	0.0	0.0	0.0	0.1	0.0

[A], End products are expressed as the average VFA concentrations (mM) of two duplicate cultures with background VFAs in media controls (RM02 or BY) subtracted. All strains were grown in RM02, except the slow growing strains which were grown in BY, indicated with an *.

3.8 Discussion

Butyrivibrio is a major group of fibre-degrading rumen bacteria, whose activity in the rumen provides VFAs for nutrition of its animal host. This chapter describes *Butyrivibrio* strains encompassing the spectrum found in the rumen, based on 16*S* rRNA sequence analysis. The ultimate goal of this work is to improve current knowledge of rumen function by understanding how plant polysaccharides are broken-down within the rumen. In the rumen, bacteria that physically attach to the ingested plant material are recognised to have an important role in fibre degradation (McAllister *et al.*, 1994; Michalet-Doreau *et al.*, 2001). The initial selection of cultures for this project was focused on the large clusters of *Butyrivibrio* strains isolated from the study of the fibre-adherent fraction of rumen contents from dairy cows (Noel, 2013), from which 30 strains were selected (Figure 3.1).

A comprehensive analysis of the full-length *Butyrivibrio* 16*S* rRNA gene sequences produced a tree consisting of three distinct *Butyrivibrio* clusters. Previous 16*S* rRNA gene sequence analyses of 48 *Butyrivibrio* strains indicated that rumen-derived *B. fibrisolvens* strains clearly fell into two phylogenetic groups, *Butyrivibrio* and *Pseudobutyrivibrio* (Forster *et al.*, 1997), but the authors concluded that "phenotypic data that support division of the genus *Butyrivibrio* are not available" (Willems *et al.*, 1996). The *B. fibrisolvens* cluster for this study consists of 8 strains and the remaining 22 strains cluster into the *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3 groups. These final 30 strains were designated as *Butyrivibrio* strains, and form the basis of all the subsequent work reported in this thesis.

The PFGE analyses of genomic DNAs digested with *Apa*I and I-*Ceu*I produced unique banding patterns for 25 and 23 of the analysed *Butyrivibrio* strains, respectively (Appendix I Figures A1.1-A1.6), providing convincing evidence for differences at the genomic level between these organisms. PFGE of uncut genomic DNA (Figures 3.5-3.7), also demonstrated the presence of many extrachromosomal elements which is consistent with previous reports of megaplasmids in *Butyrivibrio* (Yeoman *et al*., 2011). These elements varied greatly in number and size among the 30 strains. Combined with previous findings that *B. hungatei* MB2003 (Palevich, 2011), and *B. proteoclasticus* B316 (Kelly *et al.*, 2010), have multi-replicon genomes, the findings of multiple strains containing more than one large replicon confirm that this type of genome arrangement is common among rumen *Butyrivibrio* species. The reason for the existence of large extrachromosomal replicons in organisms within the *Butyrivibrio* genus is still unclear.

Most of the genes encoding core metabolic processes in two sequenced strains, MB2003 and B316, are encoded on the chromosome or chromid (secondary chromosome), while the megaplasmids and small plasmid encode no core metabolic functions, and encode largely hypothetic genes with unassigned functions.

During the annotation and analysis of the MB2003 genome (Palevich, 2011), several genes were identified that were thought to encode flagellar function. The motility assay showed that only two out of 30 strains, *B. fibrisolvens* MC2013 and *Butyrivibrio* sp. 2 strain FC2001 displayed a motile phenotype (Table 3.4). The performance of a bacterium in a motility assay does not however imply the absence of a flagellum. It is known that *B. proteoclasticus* B316 contains a flagellum but is non-motile (Attwood *et al.*, 1996); this is putatively correlated with an insert in the cluster of genes encoding components of the motor ATPase. It is therefore possible to hypothesise that these insertions exist in most non-motile strains analysed in this chapter. It is possible that *Butyrivibrio* use their flagella for adherence rather than motility, as shown for some archaea (Shimoyama *et al.*, 2009). It is known that *B. hungatei* MB2003 and *P. xylanivorans* MA3014 adhere to plant substrates (Palevich, 2011) and the potential role of flagella in this process warrants further investigation, along with morphological evidence in order to visualise or deny the existence of flagella in different *Butyrivibrio* strains and species.

Butyrivibrio is one of a small number of rumen genera capable of efficiently utilising the insoluble, complex polysaccharide xylan, a component of hemicellulose which can make up to 20-35% of the plant biomass. However, some of the *Butyrivibrio* strains characterised in this thesis have demonstrated an inability to utilise both xylan and pectin. Therefore determining the carbon sources that *Butyrivibrio* are able to utilise is therefore a useful guide to understanding their metabolic capabilities. Previous work regarding substrate utilisation of different *Butyrivibrio* strains concluded that glucose, cellobiose, maltose and esculin are universally used substrates for *Butyrivibrio* (Kopecny *et al.*, 2003). In addition, comparison of *B. proteoclasticus* B316 to its closest relatives, *B. hungatei* and *Butyrivibrio* sp. NCDO 2222 (Moon *et al.*, 2008), showed its ability to utilise amygdalin, glycogen, mannose and melezitose. *B. proteoclasticus* B316, in addition, is able to utilise fructose, inulin, mannose, rhamnose and trehalose.

To fully analyse their carbon source utilisation, *Butyrivibrio* strains were screened for growth on soluble and insoluble substrates. Table 3.5 shows that the *Butyrivibrio* strains displayed best growth on some monosaccharides (galactose and glucose),

trisaccharides (melezitose and trehalose) and sugar alcohols (glycerol, myo-inositol, mannitol and sorbitol), while monosaccharides (mannose and rhamnose), the disaccharide maltose and most glycosides (amygdalin, rutin and salicin) did not support growth. Early characterisation of *Butyrivibrio* indicated that some strains had the ability to degrade cellulose and most strains were able to digest xylan and pectin substrates, which are major components of plant fibre (Bryant and Small, 1956a; Hungate, 1966). However, previous phenotypic characterisation of *B. proteoclasticus* B316 stated that it was unable to degrade crystalline cellulose (Attwood *et al.*, 1996; Moon *et al.*, 2008). Our results of the VFA analysis confirmed the growth characteristics of *B. proteoclasticus* B316 and its ability to utilise xylan, pectin, starch, inulin and dextrin substrates (Figure 3.12), but also its inability to utilise cellulose. There was absence of growth on cellulose by all 30 *Butyrivibrio* strains as well as the *P. xylanivorans* MA3014. Therefore, the VFA data supports previous findings that *Butyrivibrio* and *Pseudobutyrivibrio* are able to utilise a wide range of insoluble substrates but are not cellulose-degrading bacteria.

The VFA data also showed *Butyrivibrio* use of polysaccharides within a single species is variable. Our analysis of 30 *Butyrivibrio* strains, designated into three species groups, has revealed distinct differences in growth patterns, both between the different species groups, but also amongst individual strains (Figure 3.12). For example, the six *B. fibrisolvens* strains (MD2001, AB2020, FE2007, ND3005, YRB2005 and WTE3004) that are closely related based on the phylogenetic analysis, displayed similar growth patterns, in particular their ability to utilise inulin, pectin and xylan. Interestingly, none of these six *B. fibrisolvens* strains were able to utilise starch for growth. The MC2013 and NC3005 strains expressed different growth patterns and these differences coincide with their phylogenetic divergence from all other *B. fibrisolvens* strains.

Each of the *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3 strains that were able to utilise insoluble substrates, displayed a distinct growth pattern. Overall, *Butyrivibrio* sp. 2 group had the largest number of strains that were not able to utilise any of the insoluble substrates despite growth on cellobiose, namely MB2005, AC2005, LC3010, FB3011 and FC2001. Interestingly, the *Butyrivibrio* sp. 2 strains were not able to utilise pectin for growth at all (except for FB3011). Interestingly, rhamnose and arabinose which are breakdown products of pectin were not able to be utilised by *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3 strains (except for XPD2006), and all *Butyrivibrio* strains (except for WCD3002, VCB2001, FCS006, FCS014 and AE2032) were able to utilise galactose, and xylose utilisation varied greatly among all three *Butyrivibrio* species groups (Table 3.5).

The VFA results for FCS006, FCS014 and AE2032 of *Butyrivibrio* sp. 3 showed that these strains cannot utilise cellobiose for growth and did not display growth on any of the insoluble substrates. This suggests that FCS006, FCS014 and AE2032 have specific substrate and growth condition requirements that will require future investigation. Also, VFA results for the other slow growing strain VCB2001 corroborated the initial growth characteristics (Table 3.5), and its ability to utilise starch, pectin and xylan. Importantly, *Butyrivibrio* sp. 3 strain AE3009, displayed the most impressive growth pattern across all insoluble substrates analysed. Furthermore the polysaccharide degrading profiles of AE3009 and B316 were comparable, where AE3009 was able to degrade and utilise xylan, pectin, starch and cellobiose more efficiently. These results suggest that some *Butyrivibrio* strains are able to breakdown complex polysaccharides, while others survive in the rumen by assimilating their degradation products for growth.

The ability of *Butyrivibrio* as a genus to utilise a wide variety of substrates, especially xylan and pectin, suggests that these microorganisms play an important role in hemicellulose degradation in the rumen, and act as significant energy suppliers to ruminants. In particular, the large production of VFAs by *B. proteoclasticus* B316 on a range of insoluble substrates, indicates the ability of certain *Butyrivibrio* to switch substrate utilisation from the simple cellobiose substrate to insoluble polysaccharides in mono-culture. Therefore it is hypothesised that certain *Butyrivibrio* strains are unable to initiate significant degradation of the insoluble polysaccharides alone and rely on more specialised *Butyrivibrio* to initiate the degradation process, resulting in the release of soluble sugars, for which they compete. This hypothesis is investigated in Chapter 5, where the xylan and pectin utilisation abilities of *B. hungatei* MB2003 and *B. proteoclasticus* B316 in co-culture are explored.

Based on observations from this and previous studies it appears that *Butyrivibrio* can be grouped into both lactate (mainly *B. fibrisolvens*), and non-lactate producing strains. Also, the observed diversity in fermentation products within each *Butyrivibrio* cluster suggest differences in the metabolic pathways of each strain, and further highlights the importance and adaptive role of *Butyrivibrio* in the digestion of fibrous constituents of animal feed. To gain a more complete understanding of the metabolic capacity of *Butyrivibrio* species, a comparison of their growth on a range of soluble substrates, end-product and metabolite analyses must be compared in parallel with whole genome sequences, as well as gene and protein expression profiling. These comparisons are made in Chapter 4.

Chapter 4

Butyrivibrio and *Pseudobutyrivibrio* pan-genome analyses

4.1 Introduction

An enormous genetic diversity exists amongst the strains belonging to the *Butyrivibrio* and *Pseudobutyrivibrio* genera (Edwards *et al.*, 2004). Yet the important adaptations that serve to divide these genera into different niches within the rumen and how this diversity is expressed in terms of plant polysaccharide breakdown, are still not defined. This chapter focuses on detailed analysis of the *Butyrivibrio* and *Pseudobutyrivibrio* pan-genomes, encompassing 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains, most of which have been isolated from pasture-fed New Zealand dairy cows. The rapidly growing number of genome sequences creates a great opportunity to improve the overall understanding of a group of organisms through the analysis of genomic data using a 'pan-genome' approach (Tettelin *et al.*, 2005). Originally a 'pan-genome' described the entire genome space occupied by any number of a defined set of clonally distinct groups of organisms or strains classified within a single species or closely related group of species (Medini *et al.*, 2005). Here the pan-genome strategy was used to identify the genome space occupied by two related genera, *Butyrivibrio* and *Pseudobutyrivibrio*, that were all once grouped as a single species, *B. fibrisolvens*.

This work defined three parts of the pan-genome; the core, which is the genes present in all of the sequenced strains, the variable part, composed of the genes present in some but not all of the strains, and the unique part, which constitutes the genes that are specific to single strains. The determination of the core- or pan-gene set of these rumen microbes will provide insights into the genes involved in breakdown of complex polysaccharides. Along with the 30 *Butyrivibrio* strains that were phenotypically characterised (Chapter 3), genome sequences of an additional 10 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains were included in the comparative genomic analyses. These 46 genomes are representative of each of the three putative *Butyrivibrio* species groups defined in Chapter 3 (*B. fibrisolvens*, *Butyrivibrio* sp. 2, and *Butyrivibrio* sp. 3) and the *Pseudobutyrivibrio* genera. The additional *Butyrivibrio* and *Pseudobutyrivibrio* strains included in comparative genomic analyses are shown in Table 4.1.

The revival, growth and large-scale genomic DNA extractions for the *Butyrivibrio* and *Pseudobutyrivibrio* strains are detailed in Section 2.2. Sequencing was carried out at the Joint Genome Institute (JGI) CA, USA. Upon submission of the sequence data, each genome was first associated with a sequencing project in the GOLD database (Pagani *et al.*, 2012). The genomes were compared and analysed using the IMG data warehouse (Markowitz *et al.*, 2013), which provides the tools for functional annotations and

structural modelling. Conserved and essential genes were identified through functional annotation, CAZyme analysis, and *in silico* prediction of the *Butyrivibrio* and *Pseudobutyrivibrio* glycobiomes.

Table 4.1. Additional strains for pan-genome analyses.

Species	Strain Name	Origin
Butyrivibrio sp. 2	*Butyrivibrio* sp. WCE2006	Bovine rumen
	Butyrivibrio sp. NC2002	Bovine rumen
Butyrivibrio sp. 3	*B. hungatei* NK4A153 [B]	Ovine rumen
	Butyrivibrio sp. AE2005	Bovine rumen
	Butyrivibrio sp. AE3003	Bovine rumen
	Butyrivibrio sp. LB2008	Bovine rumen
	B. hungatei MB2003 [A]	Bovine rumen
	B. proteoclasticus P6B7	Bovine rumen
	B. proteoclasticus B316 [A]	Bovine rumen
	B. proteoclasticus FD2007	Bovine rumen
	Butyrivibrio sp. MC2021	Bovine rumen
Pseudobutyrivibrio	*P. xylanivorans* MA3014 [A]	Bovine rumen
	P. ruminis HUN009	Bovine rumen
	P. ruminis CF1b [B]	Ovine caecum
	P. ruminis AD2017	Bovine rumen
	Pseudobutyrivibrio sp. LB2011	Bovine rumen
	Pseudobutyrivibrio sp. MD2005	Bovine rumen

[A], *B. hungatei* MB2003 (Palevich, 2011), *B. proteoclasticus* B316 (Kelly *et al.*, 2010), and *P. xylanivorans* MA3014 (Palevich, 2011) are additional strains whose genome sequences have already been analysed. [B], *B. hungatei* NK4A153 (Kenters *et al.*, 2011) and *P. ruminis* CF1b (Moon *et al.*, 2008) are additional strains isolated from sheep. They were also included in the pan-genome comparisons and analyses.

4.2 Genomic features of *Butyrivibrio* and *Pseudobutyrivibrio*

The genome and annotation features of each of the 46 *Butyrivibrio* and *Pseudobutyrivibrio* genome sequences chosen for this study are summarised in Appendix II Tables 4.2 and A2.1. The genome sizes varied greatly amongst *Butyrivibrio* ranging from ~3.37 Mb (*B. hungatei* NK4A153) to ~5.13 Mb (*Butyrivibrio* sp. 2 AC2005) with an overall average among 46 *Butyrivibrio* and *Pseudobutyrivibrio* genomes of 4.14 Mb (Table 4.2). The average *Butyrivibrio* genome size was 4.14 Mb; the individual species genome size averages were 4.40 Mb for *B. fibrisolvens*, 4.38 Mb for *Butyrivibrio* sp. 2 and 3.9 Mb for *Butyrivibrio* sp. 3. These genome sizes are much higher compared to the average genome size of *Pseudobutyrivibrio*, which is 3.07 Mb. *Pseudobutyrivibrio* genome sizes ranged from 2.80 Mb (*P. ruminis* HUN009) to 3.56 Mb (*P. xylanivorans* MA3014). Overall, *Butyrivibrio* genomes are 27% larger and contain 24% more genes than *Pseudobutyrivibrio* genomes which is a statistically significant finding. The %G+C varies from a low of 36% GC for *B. fibrisolvens* NC3005 to a high of 46% GC for *Butyrivibrio* sp. 3 FCS014. The average %G+C of *B. fibrisolvens*, *Butyrivibrio* sp. 2, *Butyrivibrio* sp. 3 and *Pseudobutyrivibrio* groups were 40%, 41%, 42% and 39%, respectively.

Table 4.2. General features of *Butyrivibrio* and *Pseudobutyrivibrio* genomes.

Species		Strain	Genome size (bp)	Scaffold count	Gene count	CDS count	tRNA count	G+C (%)	Sequencing method	GOLD ID
B. fibrisolvens	1	MD2001	4,738,960	66	4,022	3,934	61	40	Illumina HiSeq 2000	Gp0024723
	2	AB2020	4,735,565	170	4,107	4,021	56	40	Illumina HiSeq 2000	Gp0025891
	3	FE2007	4,315,442	65	3,732	3,652	62	40	Illumina HiSeq 2000	Gp0024605
	4	ND3005	4,461,744	160	3,862	3,793	51	39	Illumina HiSeq 2000	Gp0025890
	5	YRB2005	4,651,802	81	3,997	3,906	56	40	Illumina HiSeq 2000	Gp0026002
	6	WTE3004	4,665,426	87	4,000	3,897	62	40	Illumina HiSeq 2000	Gp0024604
	7	MC2013	3,813,073	48	3,274	3,180	62	44	Illumina HiSeq 2000	Gp0024692
	8	NC3005	3,871,453	93	3,415	3,326	56	36	Illumina HiSeq 2000	Gp0024691
Butyrivibrio sp. 2	9	AE3006	4,147,770	41	3,657	3,582	45	41	Illumina HiSeq 2000	Gp0025921
	10	MB2005	4,178,539	42	3,694	3,630	41	41	Illumina HiSeq 2000	Gp0025673
	11	WCD3002	4,254,596	50	3,762	3,692	47	41	Illumina HiSeq 2000	Gp0025671
	12	AD3002	4,316,529	73	3,946	3,874	56	42	Illumina HiSeq 2000	Gp0025720
	13	WCD2001	4,276,742	59	3,900	3,833	47	42	Illumina HiSeq 2000	Gp0024573
	14	VCD2006	4,705,279	66	4,111	4,018	66	41	Illumina HiSeq 2000	Gp0025997
	15	AE3004	4,470,907	6	4,049	3,965	56	40	Illumina HiSeq 2000, PacBio RS	Gp0036580
	16	LC3010	4,588,273	66	4,095	4,004	65	40	Illumina HiSeq 2000	Gp0024603
	17	WCE2006	4,528,646	94	4,001	3,917	60	39	Illumina HiSeq 2000, 2500	Gp0047467
	18	AC2005	5,126,371	74	4,764	4,681	64	40	Illumina HiSeq 2000	Gp0025885
	19	FC2001	4,575,583	39	4,156	4,082	45	42	Illumina HiSeq 2000	Gp0025884
	20	XPD2002	4,442,963	47	4,047	3,957	61	42	Illumina HiSeq 2000	Gp0025998
	21	NC2002	3,407,272	8	3,031	2,971	45	40	Illumina HiSeq 2000, 2500	Gp0047466

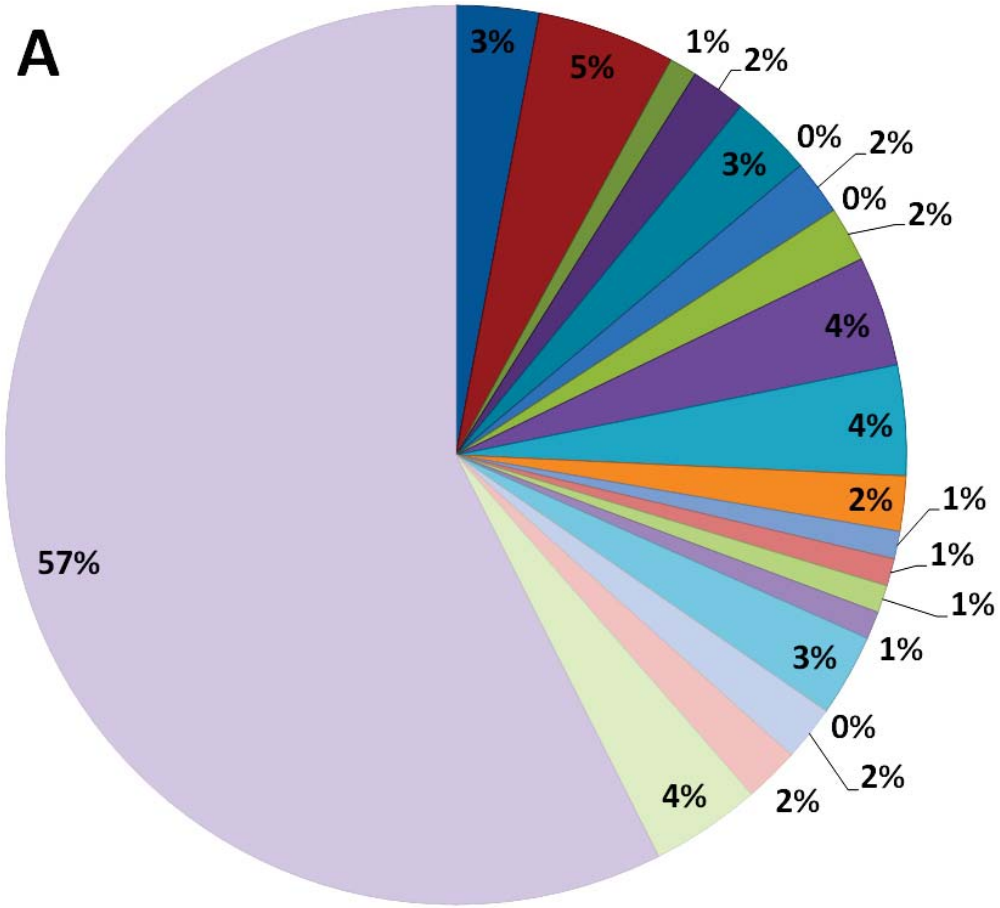
Species		Strain	Genome size (bp)	Scaffold count	Gene count	CDS count	tRNA count	G+C (%)	Sequencing method	GOLD ID
Butyrivibrio sp. 3	22	NK4A153	3,372,542	31	3,164	3,072	50	41	Illumina HiSeq 2000	Gp0024636
	23	AE2005	3,712,078	41	3,471	3,384	48	41	Illumina HiSeq 2000, 2500	Gp0036556
	24	AE3003	3,464,030	1	3,254	3,160	55	41	Illumina HiSeq 2000, 2500	Gp0036576
	25	LB2008	3,718,842	30	3,490	3,415	55	41	Illumina HiSeq 2000, 2500	Gp0036685
	26	MB2003 *	3,386,314	4	3,087	3,022	48	40	454 Pyrosequencing, Illumina MiSeq	Gp0115331
	27	AE2015	3,711,555	73	3,275	3,203	51	40	Illumina HiSeq 2000	Gp0025962
	28	P6B7	3,927,843	53	3,556	3,487	48	42	Illumina HiSeq 2000, 2500	Gp0047465
	29	B316 *	4,404,886	4	3,879	3,811	50	40	454 Pyrosequencing	Gp0003213
	30	FD2007	3,880,043	43	3,342	3,252	60	40	Illumina HiSeq 2000	Gp0036667
	31	VCB2006	3,947,728	40	3,488	3,402	54	41	Illumina HiSeq 2000	Gp0025889
	32	XBB1001	3,912,001	39	3,445	3,351	57	41	Illumina HiSeq 2000	Gp0024688
	33	AE3009	4,218,197	103	3,630	3,521	63	44	Illumina HiSeq 2000	Gp0025999
	34	MC2021	4,362,523	32	3,660	3,569	65	44	Illumina HiSeq 2000, 2500	Gp0036702
	35	XPD2006	3,888,106	31	3,523	3,456	48	43	Illumina HiSeq 2000	Gp0026001
	36	VCB2001	4,016,637	43	3,637	3,564	53	44	Illumina HiSeq 2000	Gp0026000
	37	FCS006	3,795,729	12	3,339	3,276	48	44	Illumina HiSeq 2000	Gp0024690
	38	NC2007	4,000,398	54	3,627	3,562	46	44	Illumina HiSeq 2000	Gp0024689
	39	FCS014	4,089,922	22	3,699	3,592	72	46	Illumina HiSeq 2000, PacBio RS	Gp0036578
	40	AE2032	3,686,941	6	3,357	3,253	52	43	Illumina HiSeq 2000, 2500	Gp0036579
Pseudobutyrivibrio	41	*P. xylanivorans* MA3014	3,553,207	5	3,216	3,132	57	40	454 Pyrosequencing, Illumina MiSeq	Gp0115764
	42	*P. ruminis* HUN009	2,804,489	2	2,593	2,522	52	39	Illumina HiSeq 2000, PacBio RS	Gp0036553
	43	*P. ruminis* CF1b	3,271,126	16	2,996	2,939	43	40	Illumina HiSeq 2000	Gp0026003
	44	*P. ruminis* AD2017	2,912,101	21	2,640	2,577	47	38	Illumina HiSeq 2000	Gp0025674
	45	*Pseudobutyrivibrio* sp. LB2011	2,807,604	9	2,624	2,561	51	39	Illumina HiSeq 2000, 2500	Gp0036682
	46	*Pseudobutyrivibrio* sp. MD2005	3,086,437	26	2,860	2,743	66	38	Illumina HiSeq 2000, 2500	Gp0047103

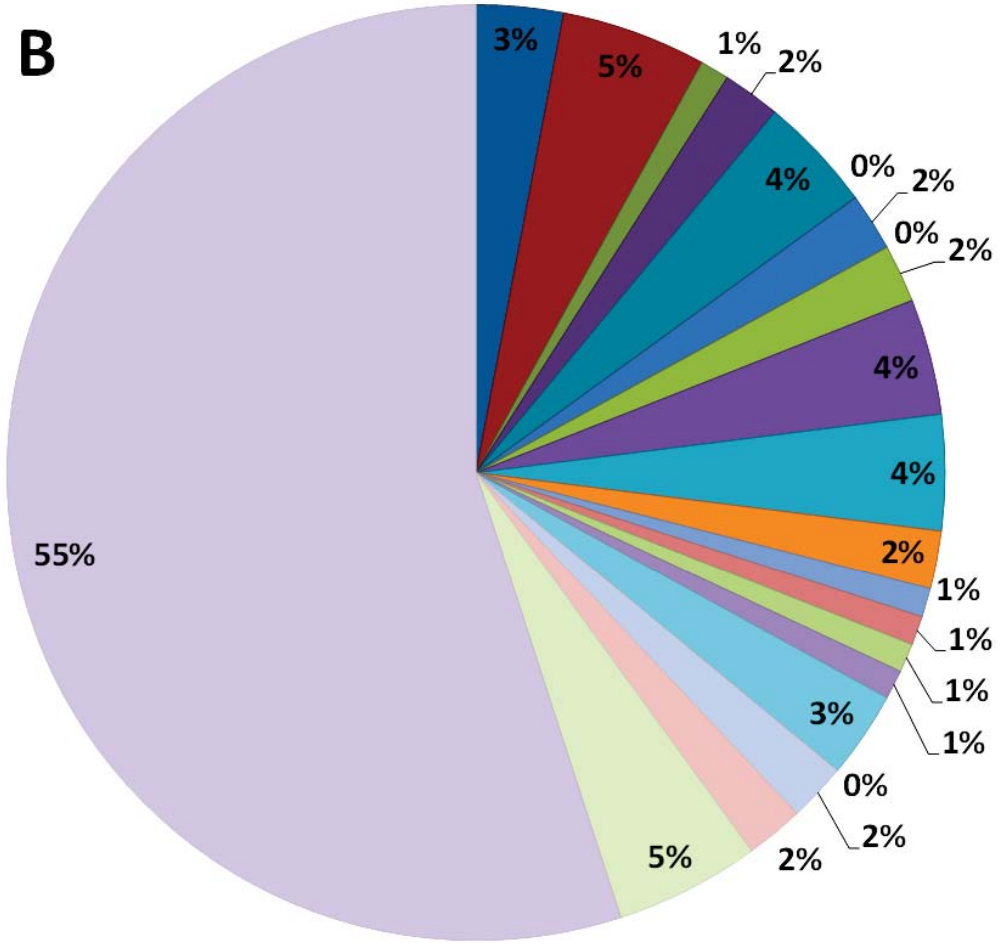
* indicates strains that have closed genomes. The genomes of all other strains are in draft status.

A comparison of gene functional annotations was conducted for the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes using the Pfam database, prokaryotic COGs, TIGRfams and KEGG pathways (Figures 4.1-4.2 and Appendix II Figures A2.1-A2.2, respectively). There are very few differences between *Butyrivibrio* and *Pseudobutyrivibrio* across the four analyses of gene function. The Pfam comparison showed ~55% of the protein coding genes have unclassified functions, with 4% function unknown and 4% with only general function predictions. All of the genomes are substantially enriched relative to the average bacterial genome, with proteins involved in carbohydrate transport and metabolism, cell wall biogenesis and translation (Figure 4.1).

The COG classification counts were highest for carbohydrate transport and metabolism at 12% and 11%, respectively (Figure 4.2). For TIGRfams, 21% and 19% of protein coding sequences were either unknown function or unclassified for the *Butyrivibrio* and *Pseudobutyrivibrio*, respectively and the largest single TIGRfam category was protein synthesis, with 13% of *Butyrivibrio* CDSs and 14% of *Pseudobutyrivibrio* CDSs encoding these functions (Figure A2.1). There were several functional groups with 4-8% classifications: amino acid biosynthesis, biosynthesis of cofactors, cell envelope, cellular processes, DNA metabolism, energy metabolism, regulatory functions, signal transduction, transport and binding proteins.

KEGG pathway analyses resulted in 29% and 27% unclassified for *Butyrivibrio* and *Pseudobutyrivibrio*, respectively (Figure A2.2). Both had classifications of 20% for global (metabolic pathways and biosynthesis of secondary metabolites) and overview (carbon and fatty acid metabolism, biosynthesis of amino acids) maps, 8% for carbohydrate metabolism, and 6% for amino acid metabolism and membrane transport functions. Overall, *Butyrivibrio* and *Pseudobutyrivibrio* appear to encode similar functional potential despite the differences in the genome sizes between these two genera.


A
3%
5%
1%
2%
0%
2%
3%
0%
2%
4%
4%
2%
1%
1%
1%
1%
3%
0%
2%
4%
2%
57%


B
3%
5%
1%
2%
4%
0%
2%
0%
2%
4%
4%
2%
1%
1%
1%
1%
3%
0%
2%
5%
2%
55%

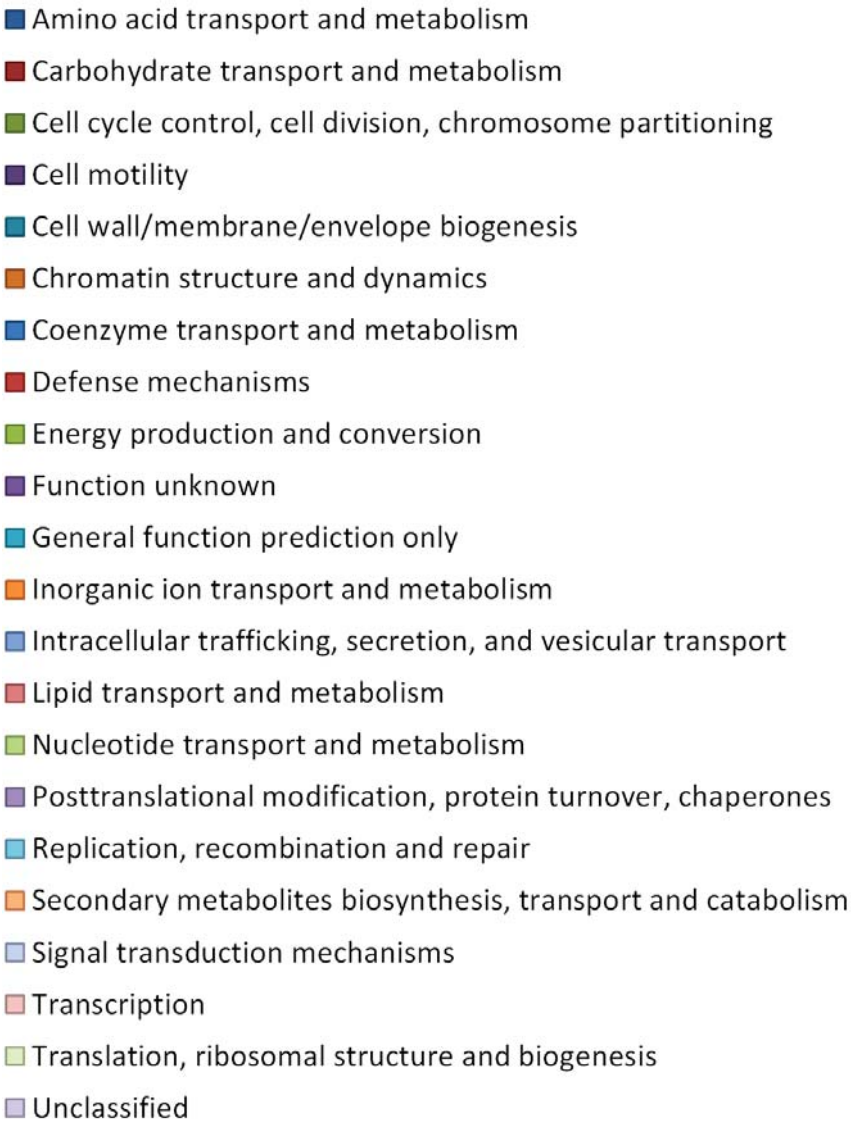
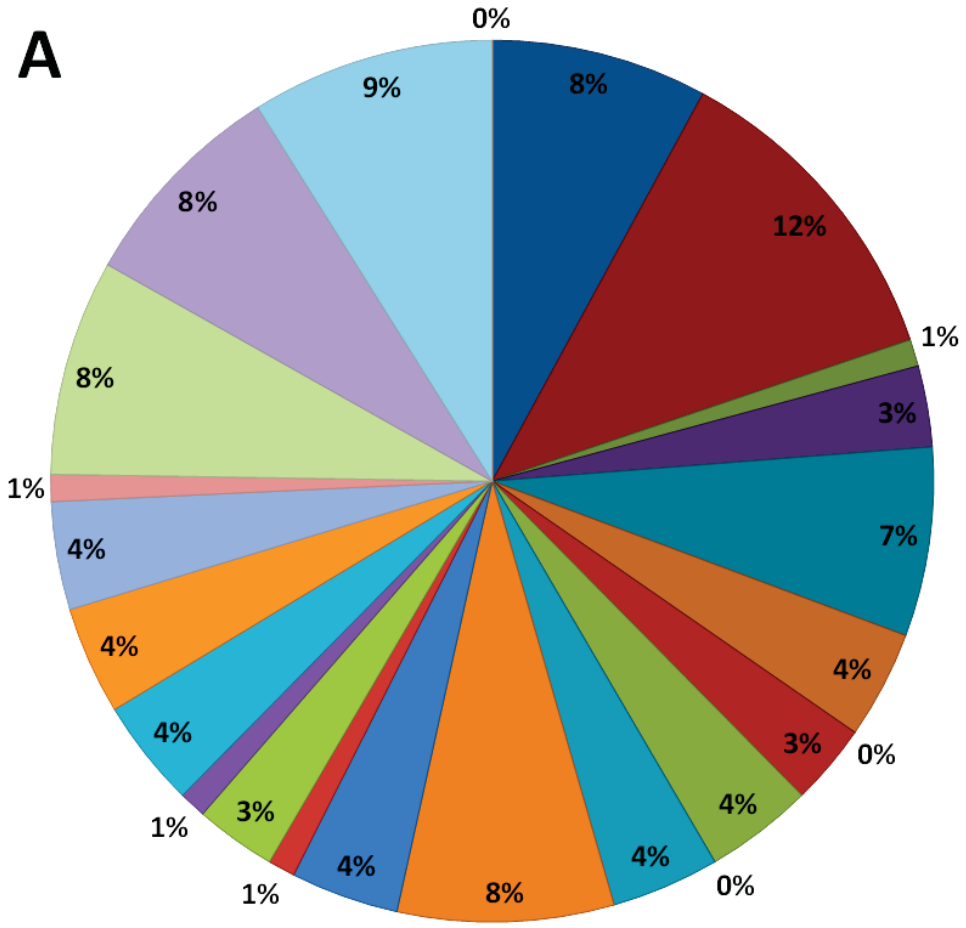
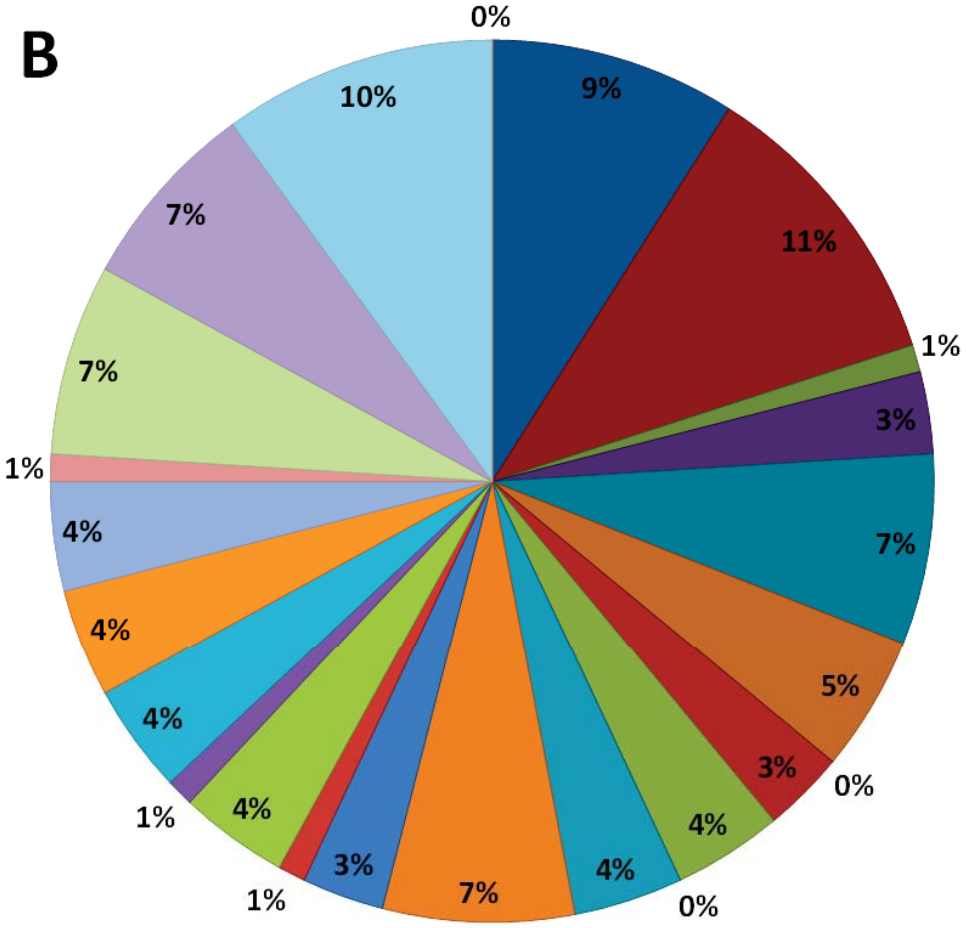


Figure 4.1. Pfam analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes. A, *Butyrivibrio* genomes; B, *Pseudobutyrivibrio* genomes.


A
0%
8%
12%
1%
3%
7%
4%
0%
3%
4%
4%
0%
4%
8%
4%
1%
3%
1%
4%
4%
4%
1%
8%
8%
9%


B
0%
9%
11%
1%
3%
7%
5%
0%
3%
4%
4%
0%
4%
7%
4%
1%
3%
1%
4%
4%
4%
1%
7%
7%
10%

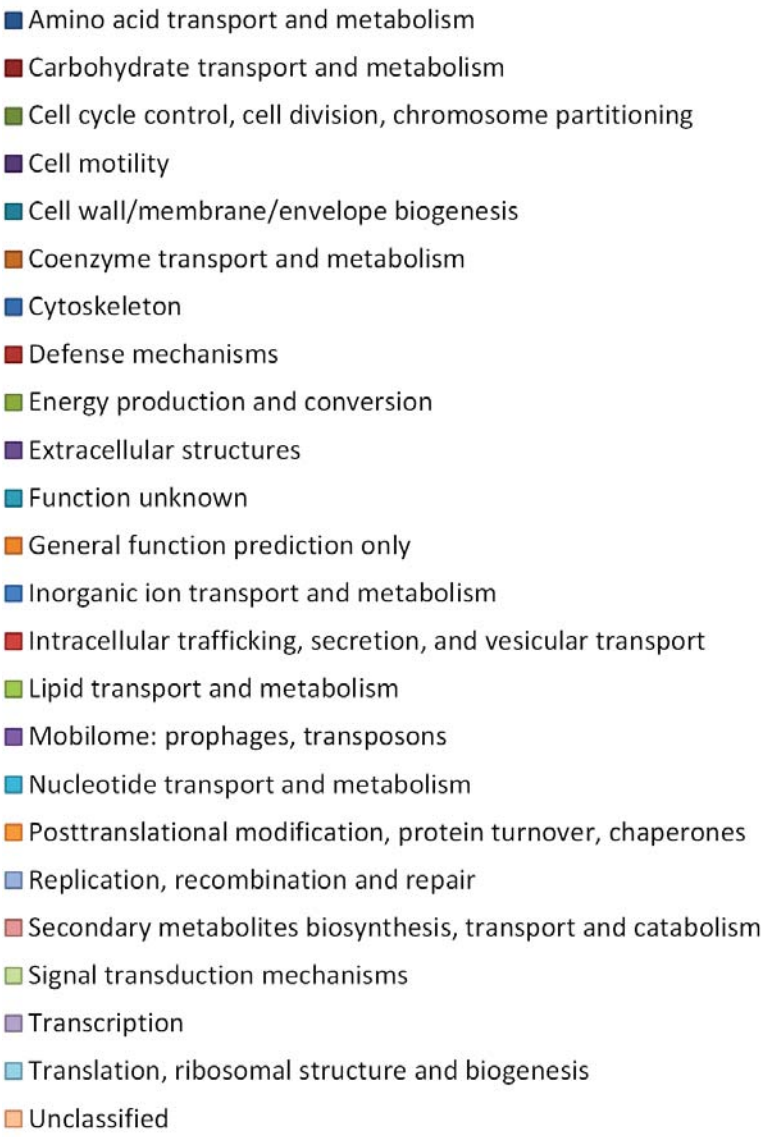


Figure 4.2. Prokaryotic COGs analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes. A, *Butyrivibrio* genomes; B, *Pseudobutyrivibrio* genomes.

4.3 Genome-based reconstruction of phylogeny

Three methods were used to investigate the similarity between the *Butyrivibrio* and *Pseudobutyrivibrio* genomes, including FGD and ANI. The FGD analysis created an initial representation of the functional placement of *Butyrivibrio* and *Pseudobutyrivibrio* genomes (Figure 4.3). The *Butyrivibrio* cluster groupings created by the FGD analysis are consistent with the phylogenetic tree of *Butyrivibrio* strains based on 16*S* rRNA full length gene sequence data, shown previously in Figure 3.1. The genomes were next compared using a correlation clustering or grouping approach. A pairwise comparison of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes across the four major protein/functional family types (Pfam, COG, TIGRfam and KO) was carried out and the results are displayed in a matrix format (Figures 4.4 and A2.3). Each cell of the matrix displays the correlation coefficient between the genomes on the corresponding row and column, expressed as a percentage. From this analysis it is apparent that the initial 16*S* rRNA gene-based species grouping and genome similarity comparisons are further supported by the correlation clustering of the genomes across the protein/functional family types. The *B. fibrisolvens*, *Butyrivibrio* sp. 2, *Butyrivibrio* sp. 3 and *Pseudobutyrivibrio* cluster groups also appear functionally defined and separate from one another. Within each of the three *Butyrivibrio* cluster groups, there are strains that display apparent genomic divergence and are clearly separate groupings. Examples of these are; *B. fibrisolvens* MC2013 and NC3005, and strains AE3003, AE2005, LB2008 and NK4A153 from *Butyrivibrio* sp. 3 (Figure 4.3).

The ANI and AF values between the genome pairs are shown in Figure 4.5. The thresholds of 95 or 96% ANI are recommended for recognising novel species and as a boundary for species demarcation (Richter and Rosselló-Móra, 2009; Kim *et al.*, 2014). The identity of whole-genome nucleotide sequences between the proposed cluster groups of *Butyrivibrio* and *Pseudobutyrivibrio* varied considerably. The mutual ANI values for the pairs of the six *B. fibrisolvens* strains MD2001, AB2020, FE2007, ND3005, YRB2005 and WTE3004 were between 96% and 98%. The AF values for these pairs were also very high ranging from 0.8 to 0.9. These results support the initial FGD analysis that groups these six strains separate from other 34 *Butyrivibrio* strains. Strains MC2013 and NC3005, that were loosely placed with the *B. fibrisolvens* group based on the FGD analysis, produced ANI values between 72% and 74% and AF values of 0.3 to 0.4 amongst this group.

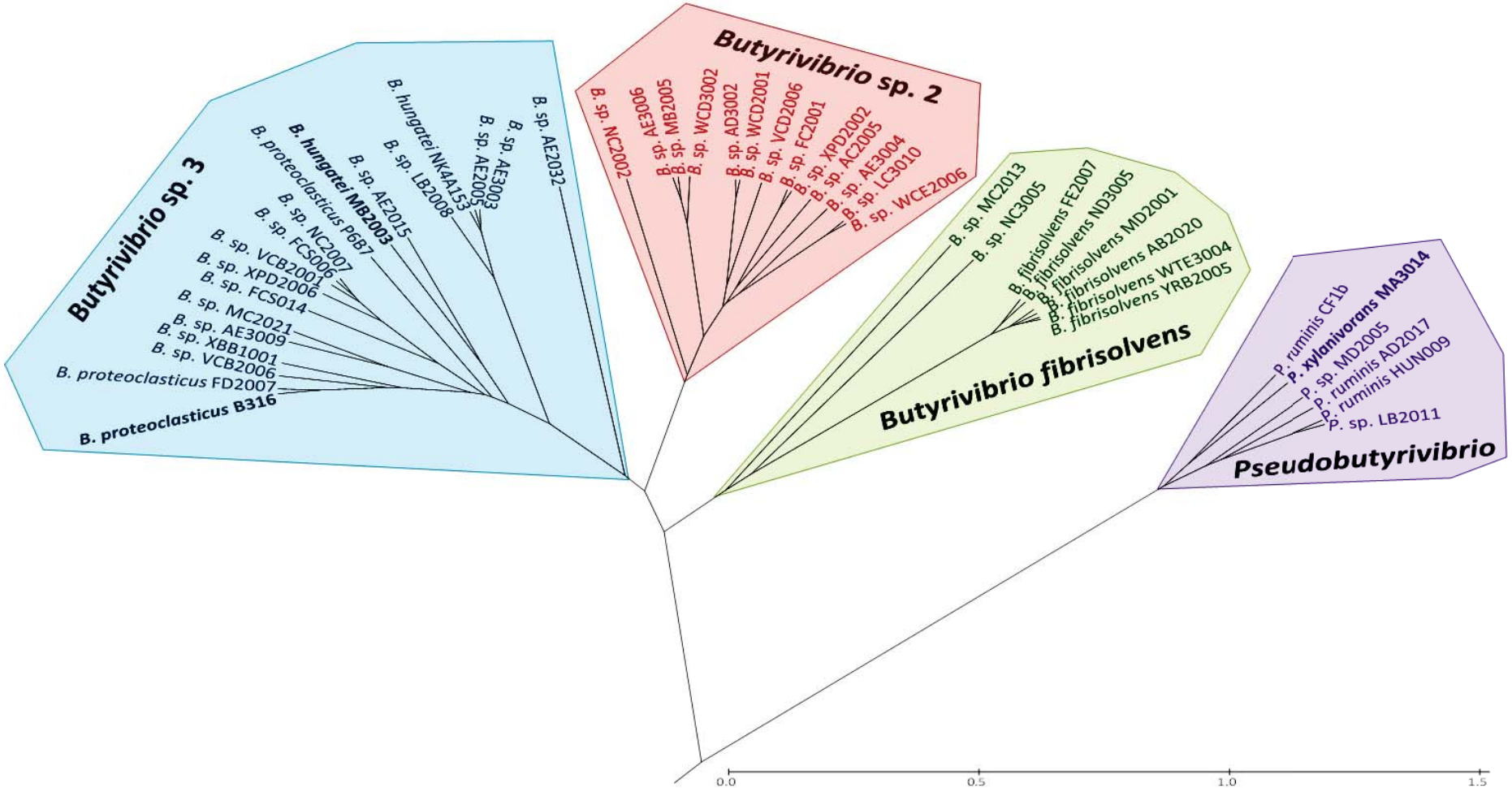


Figure 4.3. FGD of *Butyrivibrio* and *Pseudobutyrivibrio* genomes. Predicted ORFeomes of all 46 genomes were subjected to an FGD analysis and the resulting distance matrix was imported into MEGA6 (Tamura *et al.*, 2013). The functional distribution was visualised using the UPGMA method (Sneath and Sokal, 1962). The tree is drawn to scale, with branch lengths in the same units as those of the functional distances used to infer the distribution tree. Bar represents nucleotide substitutions per site.

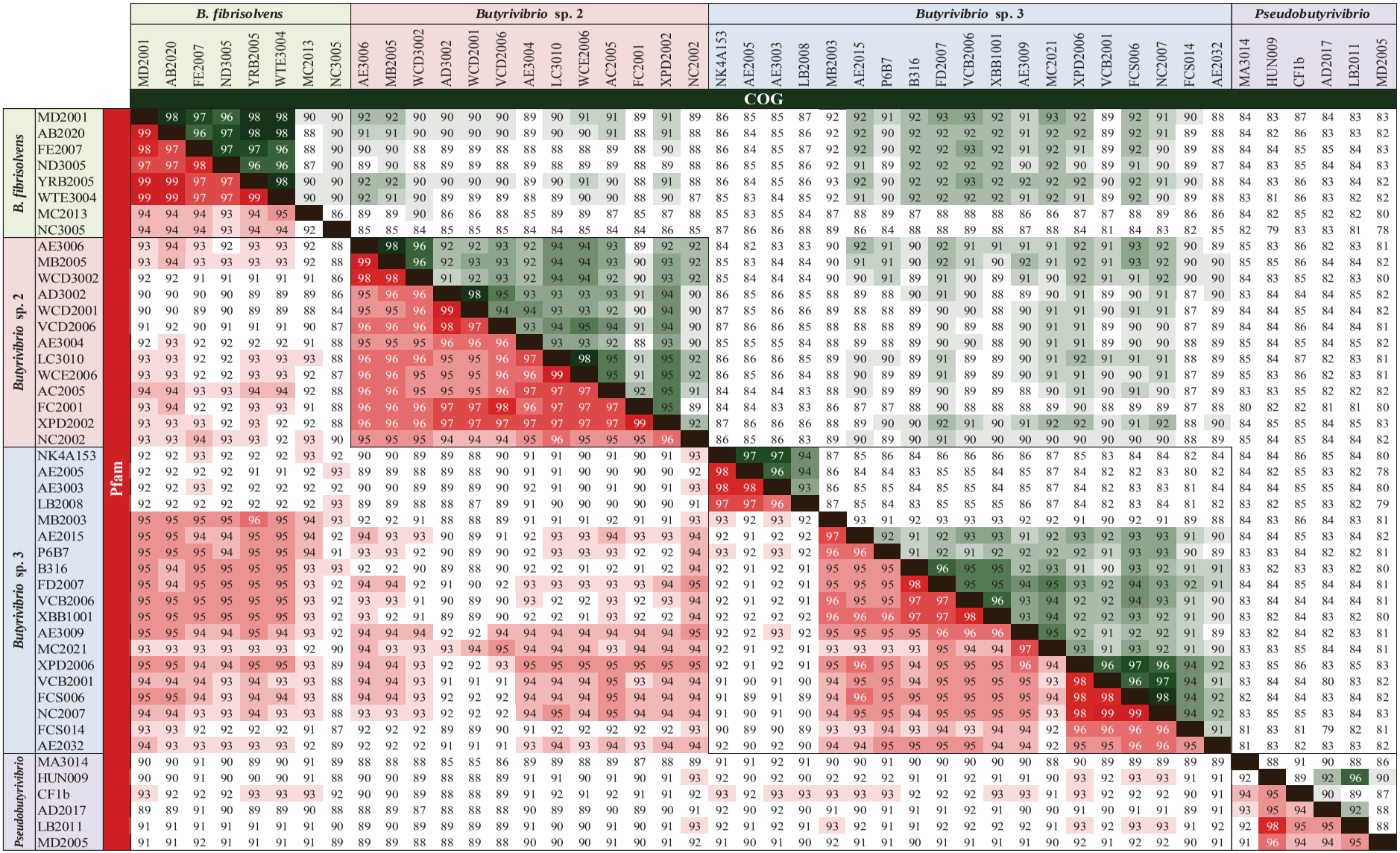

B. fibrisolvens
Butyrivibrio sp. 2
Butyrivibrio sp. 3
Pseudobutyrivibrio
COG
Pfam
MD2001 AB2020 FE2007 ND3005 YRB2005 WTE3004 MC2013 NC3005
AE3006 MB2005 WCD3002 AD3002 WCD2001 VCD2006 AE3004 LC3010 WCE2006 AC2005 FC2001 XPD2002 NC2002
NK4A153 AE2005 AE3003 LB2008 MB2003 AE2015 P6B7 B316 FD2007 VCB2006 XBB1001 AE3009 MC2021 XPD2006 VCB2001 FCS006 NC2007 FCS014 AE2032
MA3014 HUN009 CF1b AD2017 LB2011 MD2005

Figure 4.4. Genome clustering comparison of Pfam and COG domains from *Butyrivibrio* and *Pseudobutyrivibrio*. Each cell of the matrix displays the correlation coefficient between the genomes in the corresponding row and column. The correlation coefficient is computed based on the similarity of the functional characterization of the genomes. For this analysis the correlation coefficients have been converted to percentages. The diagonal correlations of the genomes with themselves are always 1.00 or 100%, and are thus left as black squares. The 46 genomes have been arranged in the order in which they are grouped in the FGD analysis to provide reference points for comparisons. The three *Butyrivibrio* and one *Pseudobutyrivibrio* designated cluster groups have been outlined.

The ANI and AF values for these two strains were of similar value amongst all other remaining strains, suggesting that MC2013 and NC3005 belong to distinct groups, different from all other analysed *Butyrivibrio* strains. *Butyrivibrio* sp. 2 and 3 groupings from the FGD analysis were supported by the ANI and AF analysis, their values between the groups were 73% to 74% and 0.2 to 0.4, respectively. Among the *Butyrivibrio* sp. 2 group, there were several strains with high ANI and AF values: 89% to 98% and 0.8 to 0.9, respectively, for strains AE3006, MB2005 and WCD3002. The strain pairs, AD3002 and WCD2001, LC3010 and WCE2006, FC2001 and XPD2002 had ANI identities of 98%, 99% and 99%, respectively, and AF values of 0.9 each. Similarly, small groups and pairs of highly similar strains were found in *Butyrivibrio* sp. 3 group (e.g. NK4A153, AE2005, AE3003 and LB2008 in one group, and XPD2006, VCB2001, FCS006 and NC2007 in another). Among *B. proteoclasticus* strains, B316 and FD2007 were highly related, according to the ANI identities of 85% to 98%, and AF of 0.6 to 0.9, respectively.

The *Pseudobutyrivibrio* cluster group, consisting of only 6 strains, had ANI identity pairs ranging from 78% to 93% and AF from 0.5 to 0.8. The only pair within this cluster group that had a high ANI and AF were *P. ruminis* HUN009 and *Pseudobutyrivibrio* sp. LB2011 with 93% and 0.8, respectively. The ANI and AF identities of the 6 *Pseudobutyrivibrio* strains compared with the 40 *Butyrivibrio* strains were between 70% to 71% and ~0.2, respectively.

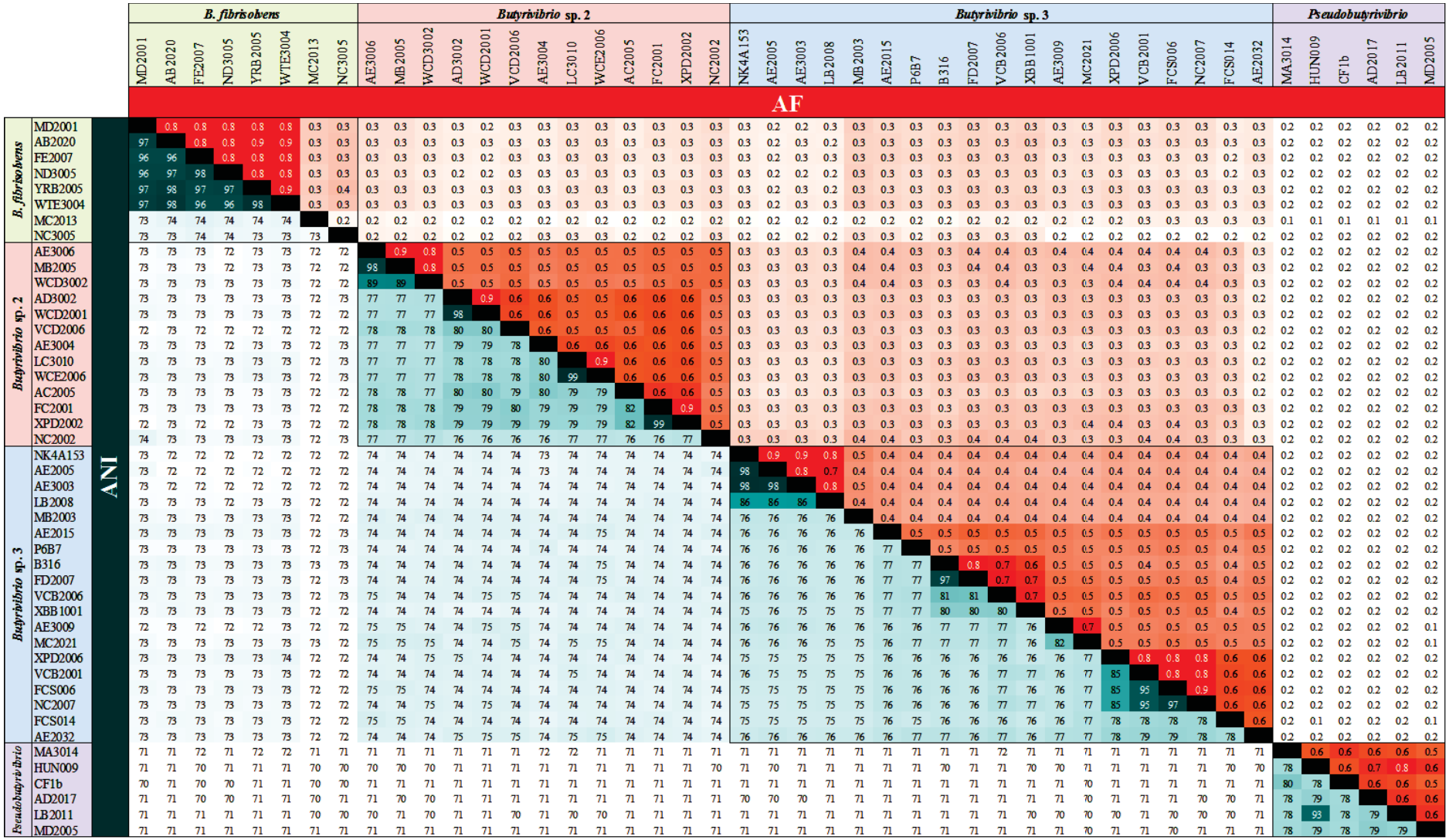


Figure 4.5. ANI and AF analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes.

4.4 Synteny, amino acid and codon usage comparison

A MUMmer plot was used to align the incomplete draft *Butyrivibrio* genomes with the closed *B. proteoclasticus* B316 genome (Appendix II Figures A2.4-A2.6). From this analysis, only strains ND3005 and AB2020 from *B. fibrisolvens* produced a strong diagonal pattern indicative of strong synteny (Figure A2.4). Strains XBB1001, VCB2006 and FD2007 from *Butyrivibrio* sp. 3 groups produced straight lines when compared with B316 (Figure A2.6). This suggests that a large proportion of the genomes of these five strains have very similar regions in B316. A closer examination of the *Butyrivibrio* sp. 3 genomes revealed that only the FD2007 genome has a chromid similar to B316.

Among the *Butyrivibrio* sp. 2 strains, approximately half produced a diagonal pattern (Figure A2.5), as did the majority of *B. fibrisolvens* (Figure A2.4) and *Butyrivibrio* sp. 3 strains (Figure A2.6). Some plots also showed an X-shaped pattern known as an X-alignment. This X-shape is symmetrical about the putative origin of replication of the two compared genomes and indicates that matching sequences tend to occur at the same distance from the origin/terminus of replication but not necessarily on the same side of the origin. Examples of these occurrences were observed in *Butyrivibrio* sp. 3 group strains such as AE2015, AE3003 and AE2032.

The amino acid and codon usage of *Butyrivibrio* and *Pseudobutyrivibrio* were compared to determine if the genomes cluster based on similar usage (Appendix II Figures A2.7-A2.8). The most frequently used amino acids were isoleucine and leucine, along with serine, aspartate, glutamate, alanine, lysine, valine and glycine. Amongst the cluster groups, serine was frequently used amongst *B. fibrisolvens*, glutamate and alanine among *Butyrivibrio* sp. 2, 3 and *Pseudobutyrivibrio*, with lysine usage frequent among *Butyrivibrio* sp. 3. The strains with distinct codon usage profiles are *B. fibrisolvens* strains NC3005 and MC2013, strain VCD2006 of the *Butyrivibrio* sp. 2 and strains AE2015 and MB2003 of *Butyrivibrio* sp. 3.

Among the 46 genomes, the most frequently used codons were GAG, GAT, GAA, CTT and ATC (Figure A2.8). Among the cluster groups, the codons ATT, AAT and GCA were frequent among *Pseudobutyrivibrio*, AAA and AAT among *Butyrivibrio* sp. 2 and 3 (especially abundant in strain NC3005), and TAT for *B. fibrisolvens*. The strains with differing codon usage profiles are *B. fibrisolvens* strains NC3005 and MC2013, and *Butyrivibrio* sp. 2 VCD2006.

4.5 Pan-genome analysis

The pan-genome of the *Butyrivibrio* and *Pseudobutyrivibrio* genera was determined using BLAST analysis and was used to define the core, variable and unique gene families. The pan- and core-genomes are represented by 29,105 and 602 orthologous gene families, respectively (Figure 4.6). The "core-genome" predominantly consists of housekeeping, carbohydrate metabolism, and transport genes. Figure 4.7 is a flowerplot diagram illustrating the number of unique, cluster group-specific and core orthologous gene families (also based on BLAST analysis) in each of the *Butyrivibrio* and *Pseudobutyrivibrio* strains. A total of 602 orthologous gene families represent the "core-genome" set (gene families are shared amongst all genomes) and is shown in the centre of the diagram. The sum of the four cluster group-specific orthologous gene families does not equal the 602 "core-genome" set, because BLAST analysis for each cluster group consisted of only the genomes within that cluster group.

The cluster group with the highest number of cluster group-specific genes is the *Pseudobutyrivibrio* with 471 genes. Examples of the predicted functions that these genes encode are; flagella biosynthesis, signal transduction, acetolactate synthase (ALS), accessory Sec system proteins, diguanylate cyclase (GGDEF), phosphoesterase, cell division protein (FtsA), helicase, β-glucosidase or GH3 protein, and many hypothetical proteins. The strains with the highest number of unique genes per genome were *B. fibrisolvens* strains MC2013 and NC3005, *Butyrivibrio* sp. 2 strains AC2005, AE3004 and VCD2006, *Butyrivibrio* sp. 3 strains B316, P67B, *B.* sp. AE2015, LB2008, FCS014 and AE2032, *P. xylanivorans* MA3014, *P. ruminis* CF1b and *Pseudobutyrivibrio* sp. MD2005.

The analysis identified considerable variation in the conservation of orthologous gene families, both between and within the *Butyrivibrio* and *Pseudobutyrivibrio* pan-genomes. This suggests a degree of specialisation within these bacteria, in addition to a set of genes required for polysaccharide degradation in the rumen. Thus, next an investigation and comparison into the polysaccharide-degrading capabilities of the *Butyrivibrio* and *Pseudobutyrivibrio* genomes follows.

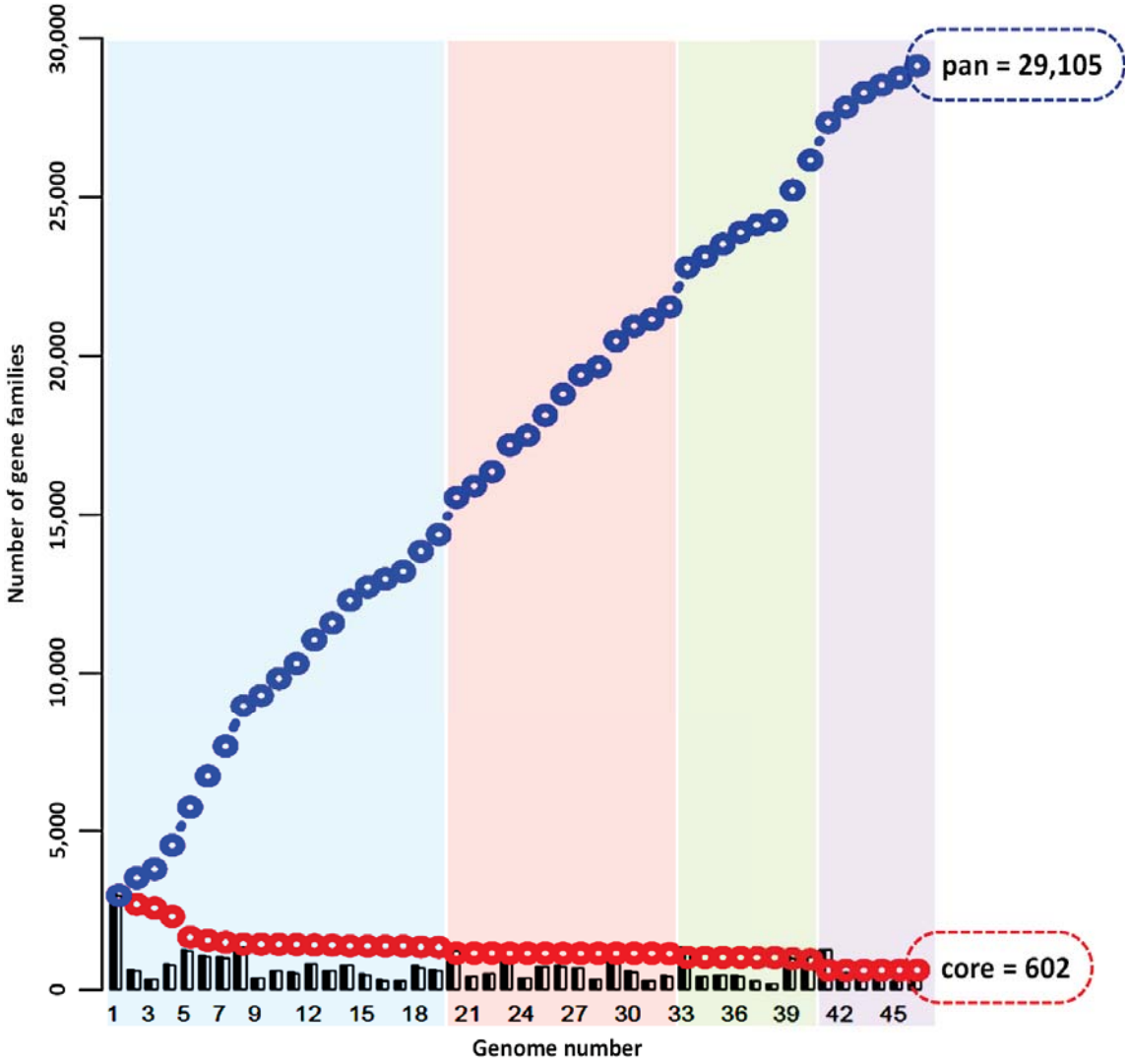


Figure 4.6. Core- and pan-genomes of *Butyrivibrio* and *Pseudobutyrivibrio* defined using BLAST analysis. The number of gene families belonging to the pan-genome shown as blue circles and the core-genome shown in red circles. Black bars, number of new genes in genome; white bars, number of new gene families in genome. Genome numbers correspond to the order of genomes in Table 4.2.

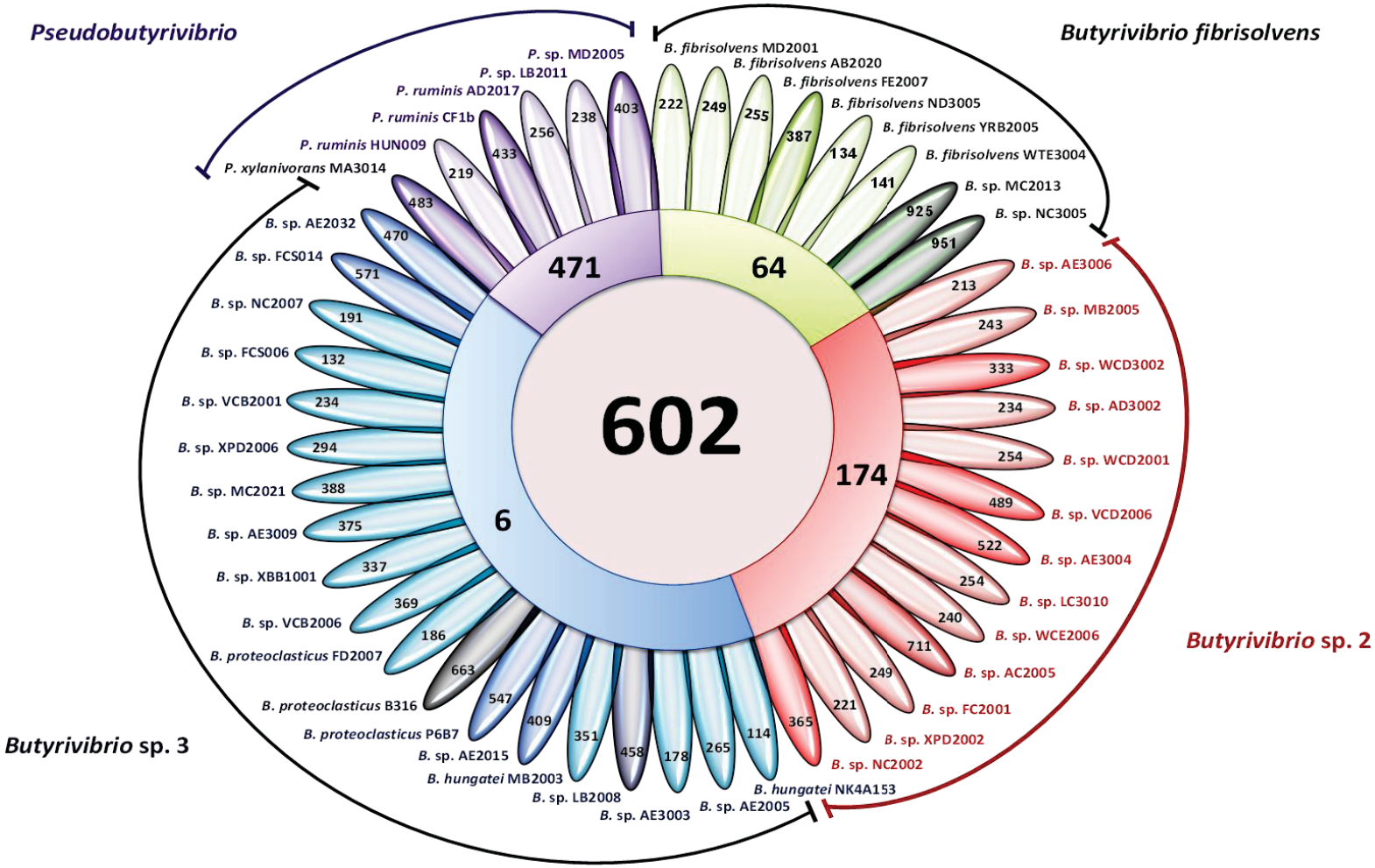


Figure 4.7. Flowerplot diagram of unique, group-specific and core gene families in the *Butyrivibrio* and *Pseudobutyrivibrio* genomes. The core genome is shown in the centre circle. Each coloured segment represents the number of gene families shared among the four species groups, and the outer petals represent unique gene families for individual genomes.

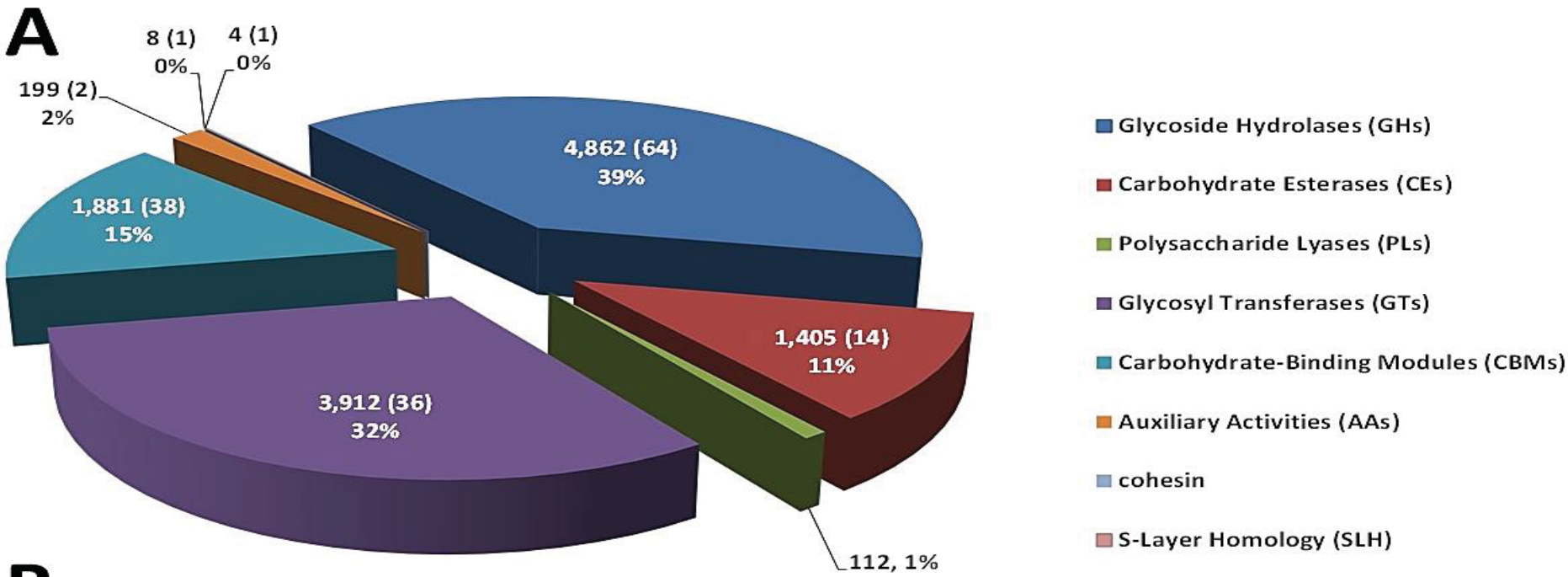
4.6 Glycobiomes of *Butyrivibrio* and *Pseudobutyrivibrio*

As described in Chapter 1, the functional domains of proteins involved in the breakdown or synthesis of complex carbohydrates are grouped into six functional classes: GHs, CEs, PLs, GTs, CBMs and SLH collectively known as CAZymes. The GTs control carbohydrate assembly whereas the GHs, PLs and CEs direct carbohydrate breakdown. The classification and assignment of CAZyme families and identification of the functional domains, make it possible to predict the range of carbohydrate substrates that can be utilised by these bacteria (Cantarel *et al.*, 2012). The polysaccharide-degrading capabilities of the *Butyrivibrio* and *Pseudobutyrivibrio* genomes were therefore determined based on CAZyme analysis using the dbCAN resource (Yin *et al.*, 2012).

In total, 163 CAZyme families were identified in the *Butyrivibrio* and *Pseudobutyrivibrio* pan-genome, consisting of 64 GHs, 14 CEs, 7 PLs, 38 CBMs, 36 GTs, two AAs, one SLH and cohesin (Figure 4.8). A comparison of the mean scores across each CAZyme family indicated the most abundant CAZyme gene for each cluster group. *B. fibrisolvens* had the highest number of GTs and CBMs, *Butyrivibrio* sp. 2 had the most GHs, CEs and AAs, and *Butyrivibrio* sp. 3 had the most PLs. *Pseudobutyrivibrio* had the fewest number of CAZymes across all categories per genome, but as indicated earlier, also had the smallest genomes. The CAZymes that are the most prominent in the analysed genomes are GHs, GTs, CBMs, CEs and PLs; they accounted for 98% of the combined *Butyrivibrio* and *Pseudobutyrivibrio* CAZymes. Most of the genomes analysed contained more than 200 CAZymes each, with 12 of the *Butyrivibrio* genomes having more than 300 CAZymes per genome (Figure 4.9).

The CAZyme families that were detected in all 46 genomes examined (Figure 4.10) were: GHs 2, 3, 5, 13, 18, 25, 32, 36, 43, 77 and 94; CEs 1 and 4; GTs 2, 4, 5, 28, 35 and 51; CBMs 34, 48 and CBM50. Some CAZymes, however, were less abundant, for example GHs 73, 37, 117, 15, 46, 12, 24, 64, 74, 81, 85, 87 and 128; CEs 5 and 15; PLs 6, 10 and 21; GTs 56, 9, 94, 73, 76, 92, 7, 87, 22, 30, 60 and 90; CBMs 42, 56, 12, 20, 47, 51, 21, 22, 28, 39, 46, 59 and 65; and SLH, occurred only in five or less genomes.

A



B

Species	Genomes	GH		CE		PL		GT		CBM		AA		cohesin		SLH		Total
		Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean	
Butyrivibrio fibrisolvens	8	938	117.3	262	32.8	27	3.4	851	106.4	365	45.6	34	4.3	0	0.0	1	0.1	2,478
***Butyrivibrio* species 2**	13	1,564	120.3	439	33.8	7	0.5	1,148	88.3	544	41.8	68	5.2	2	0.2	0	0.0	3,772
***Butyrivibrio* species 3**	19	1,919	101.0	582	30.6	76	4.0	1,606	84.5	797	41.9	77	4.1	4	0.2	3	0.2	5,064
Pseudobutyrivibrio	6	441	73.5	122	20.3	2	0.3	307	51.2	175	29.2	20	3.3	2	0.3	0	0.0	1,069
Total	46	4,862	105.7	1,405	30.5	112	2.4	3,912	85.0	1,881	40.9	199	4.3	8	0.2	4	0.1	12,383

Figure 4.8. CAZyme composition of *Butyrivibrio* and *Pseudobutyrivibrio* genomes and cluster groups. A, CAZyme composition of *Butyrivibrio* and *Pseudobutyrivibrio* genomes. The first number corresponds to the total number of genes that contain the specific CAZyme family, second number in brackets represents total number of CAZyme families. B, Summary of the individual CAZy classes across the four cluster groups with mean calculated relative to the number of genomes in each cluster group.

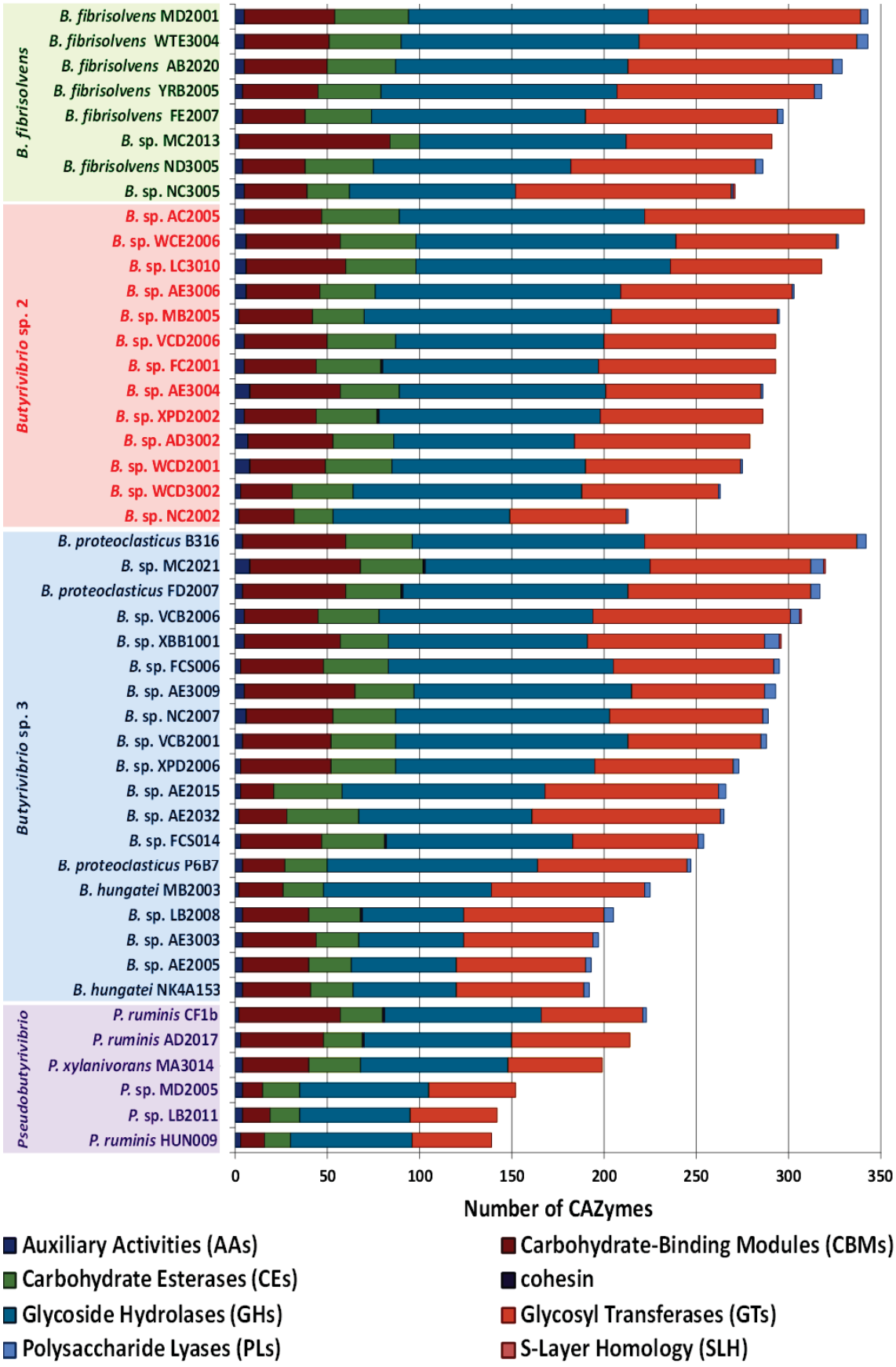


Figure 4.9. Comparative analysis of annotated *Butyrivibrio* and *Pseudobutyrivibrio* CAZymes. The numbers and type of CAZyme modules or domains were represented as coloured horizontal bars.

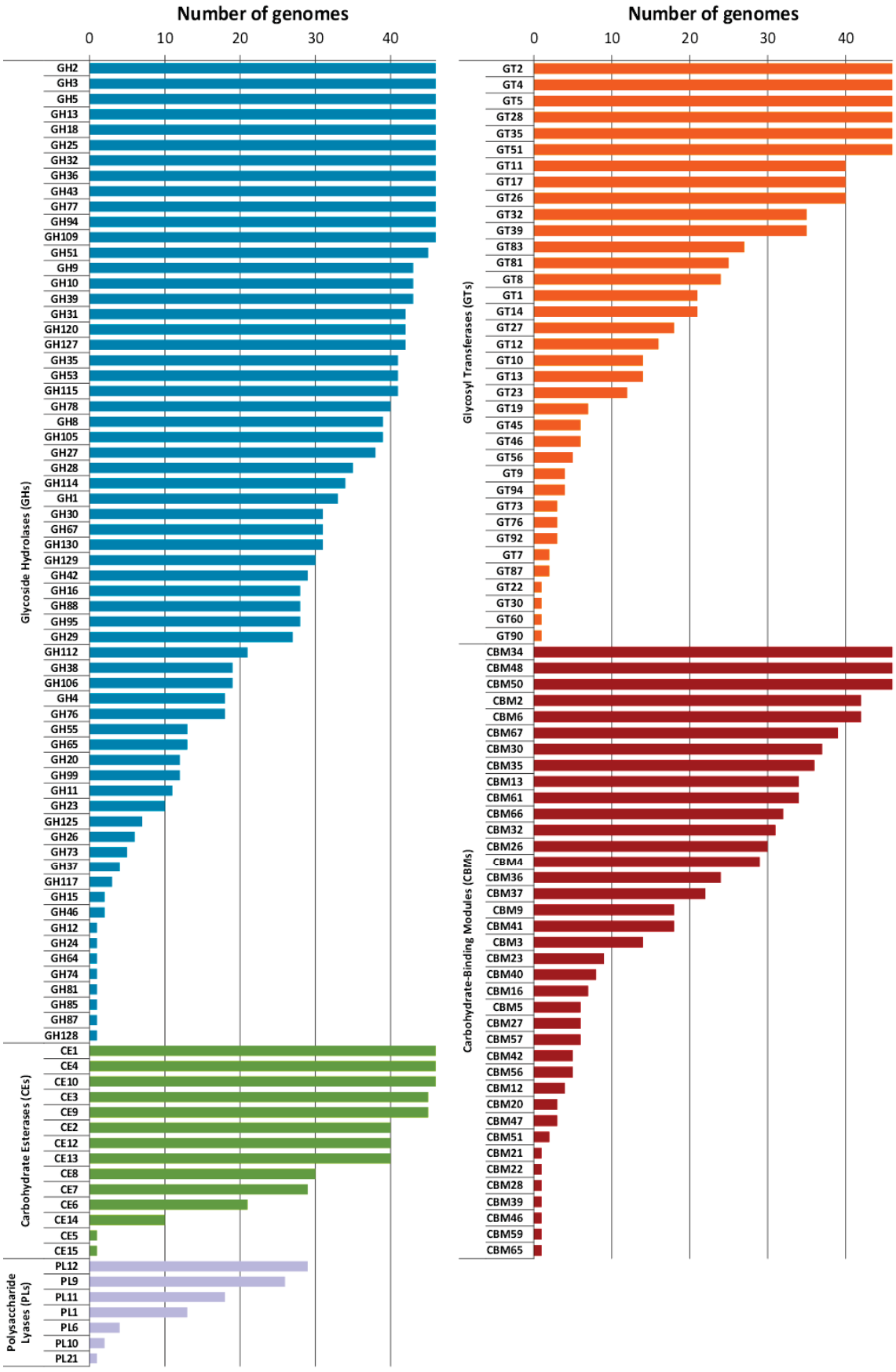


Figure 4.10. The distribution of each CAZyme class and family in *Butyrivibrio* and *Pseudobutyrivibrio* genomes. Coloured bars represent the total numbers of genomes that contain members of specific CAZyme family present in their genomes.

The distribution of some CAZyme families appeared to be group-specific (Table 4.3). For example 11 families (GHs 12, 64, 81, 87, 128, CE15, GTs 7, 90, CBMs 21, 39, 59), were only found in the *B. fibrisolvens*; families GH85, CE5, GT22, CBMs 47 and 65 were only found in *Butyrivibrio* sp. 2; families GH74, PLs 6, 10, 21, GT60, CBMs 20, 46 were only found in *Butyrivibrio* sp. 3, and families GHs 15, 24, GT30, CBMs 22, 28, 42 were only found in *Pseudobutyrivibrio*. Also, 25 families (GHs 26, 28, 30, 38, 65, 67, 88, 105, 112, 129, CE13, PLs 11, 12, GTs 1, 10, 11, 13, 17, 26, 45, 56, 83, CBMs 35, 41 and SLH), appeared to be *Butyrivibrio*-specific. A detailed list of all CAZy families and their known functions, produced from this analysis can be found in Appendix II Table A2.2.

The normalised relative abundances for all 163 CAZyme families were determined for the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes (Figure 4.11). The CAZyme profiles across all 163 CAZyme families varied greatly between the *Butyrivibrio* and *Pseudobutyrivibrio* genomes. *B. fibrisolvens* strains MC2013 and NC3005 and *Butyrivibrio* sp. 2 NC2002 appear to have unique CAZyme profiles and cluster separately from related strains that were members of the same group defined by ANI and AF analyses. The CAZyme-encoding gene complement of all other *Butyrivibrio* and *Pseudobutyrivibrio* strains grouped in the same manner as they did in the whole-genome-based (FGD, ANI) cluster grouping, including the *Butyrivibrio* sp. 3 that displayed two distinct groups of strains. Therefore, the CAZyme gene profiles of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* genomes overall corresponded reasonably well to the phylogenetic trees of *Butyrivibrio* strains based on 16*S* rRNA full length gene sequence data (Figure 3.1), and whole-genome sequence analyses based on FGD and ANI (Figures 4.3 and 4.5).

4.6.1 Carbohydrate-Binding Modules (CBMs)

CBMs are autonomously folding and functioning protein domains, and while they have no enzymatic activity *per se* (Cantarel *et al.*, 2009), they potentiate the carbohydrate-binding activity of the various proteins in which they occur. The most abundant CBMs were CBMs 6, 13, 2, 4, 48 and 34 (Figure 4.12). In contrast, CBMs 51, 65, 59, 46, 39, 28, 22 and 21 were found only in strains AE2032, AC2005, AE3006, NC3005, NC2007, MC2013, AD2017, CF1b and AB2020, respectively.

Table 4.3. Group specific CAZyme families.

CAZy Family	B. fibrisolvens	Butyrivibrio sp. 2	Butyrivibrio sp. 3	Pseudobutyrivibrio
GH11	0	3	7	1
GH12	1	0	0	0
GH15	0	0	0	2
GH20	1	11	0	0
GH23	0	3	4	4
GH24	0	0	0	1
GH26	1	1	4	0
GH28	14	25	38	0
GH30	11	12	20	0
GH37	3	1	0	0
GH38	5	11	8	0
GH46	1	0	1	0
GH55	7	0	8	0
GH64	1	0	0	0
GH65	4	2	10	0
GH67	1	14	17	0
GH73	0	5	1	0
GH74	0	0	1	0
GH81	1	0	0	0
GH85	0	1	0	0
GH87	1	0	0	0
GH88	8	5	15	0
GH105	14	23	60	0
GH112	6	1	14	0
GH117	0	1	1	1
GH125	1	6	0	0
GH128	1	0	0	0
GH129	9	14	9	0
CE5	0	1	0	0
CE13	11	15	19	0
CE14	0	3	4	4
CE15	1	0	0	0
PL6	0	0	4	0
PL10	0	0	2	0
PL11	10	1	15	0
PL12	11	7	16	0
PL21	0	0	1	0

CAZy Family	B. fibrisolvens	Butyrivibrio sp. 2	Butyrivibrio sp. 3	Pseudobutyrivibrio
GT1	6	7	12	0
GT7	4	0	0	0
GT10	7	1	9	0
GT11	34	46	71	0
GT12	0	4	14	0
GT13	4	2	12	0
GT17	8	14	19	0
GT22	0	1	0	0
GT26	10	15	23	0
GT30	0	0	0	1
GT45	2	1	3	0
GT46	5	0	3	0
GT56	1	2	2	0
GT60	0	0	1	0
GT76	3	1	0	0
GT83	3	41	26	0
GT87	0	0	1	2
GT90	1	0	0	0
GT92	1	0	2	0
GT94	1	3	0	0
CBM5	3	0	1	4
CBM16	9	0	1	0
CBM20	0	0	3	0
CBM21	1	0	0	0
CBM22	0	0	0	1
CBM27	0	9	1	0
CBM28	0	0	0	1
CBM35	34	34	42	0
CBM39	1	0	0	0
CBM41	1	8	10	0
CBM42	0	0	0	5
CBM46	0	0	1	0
CBM47	0	3	0	0
CBM51	0	1	2	0
CBM57	0	2	4	0
CBM59	1	0	0	0
CBM65	0	1	0	0
SLH	1	1	3	0

Numbers refer to the sum of genes that contain specific CAZy families for each cluster group.

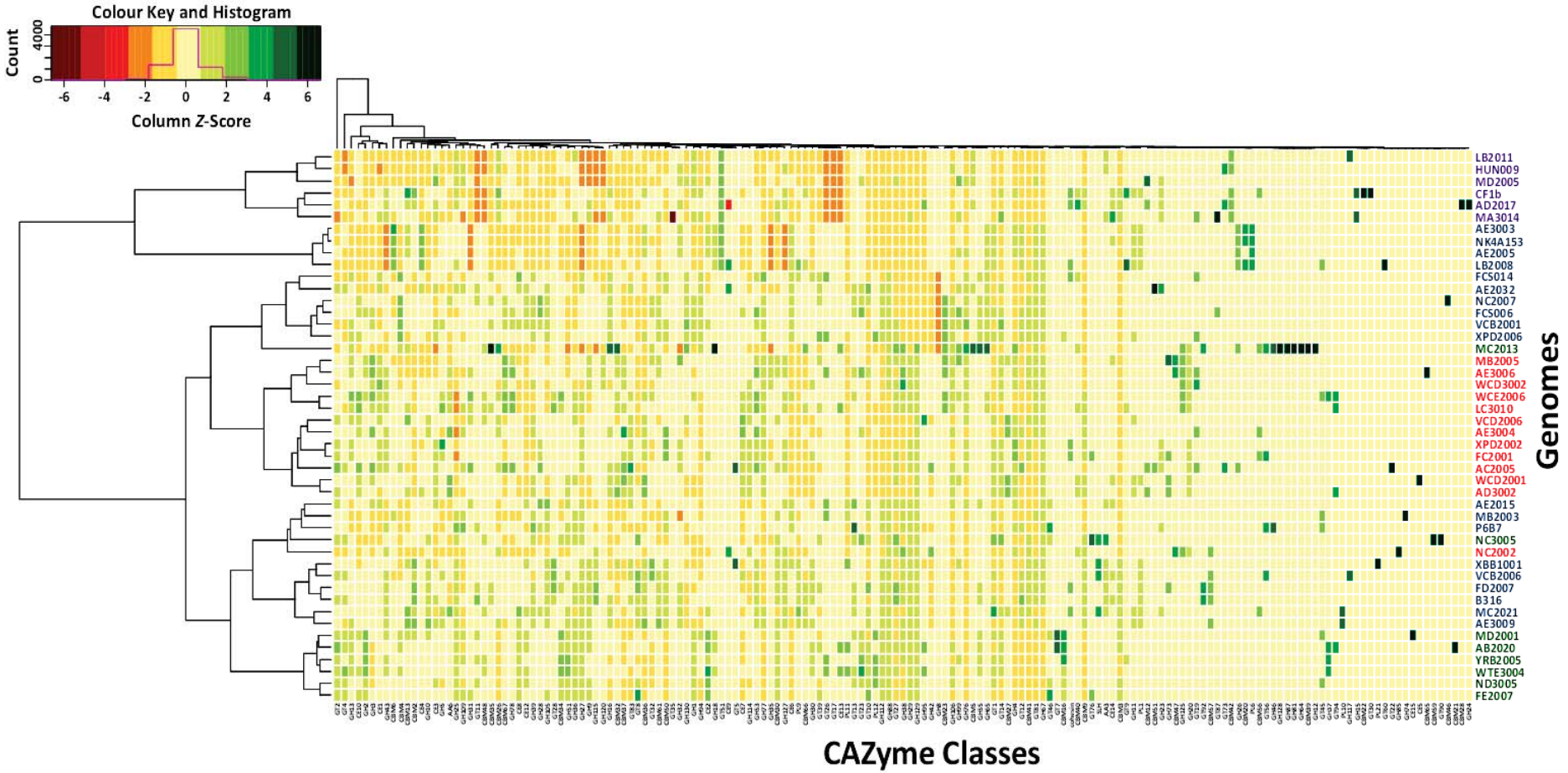


Figure 4.11. Heatmap of normalised relative abundances for CAZyme families determined for *Butyrivibrio* and *Pseudobutyrivibrio* genomes. The relative normalised inferred CAZy gene family abundances per genome (*Z*-score) using a heat colour scheme (red to green), indicating low to high relative abundance. The Bray-Curtis distances (Bray and Curtis, 1957) of compositional dissimilarity and hierarchical cluster analysis using Ward's method (Ward, 1963), were used to calculate normalised abundance. Genome names are coloured to represent *B. fibrisolvens* in green, *Butyrivibrio* sp. 2 in red, *Butyrivibrio* sp. 3 in blue and *Pseudobutyrivibrio* in purple.

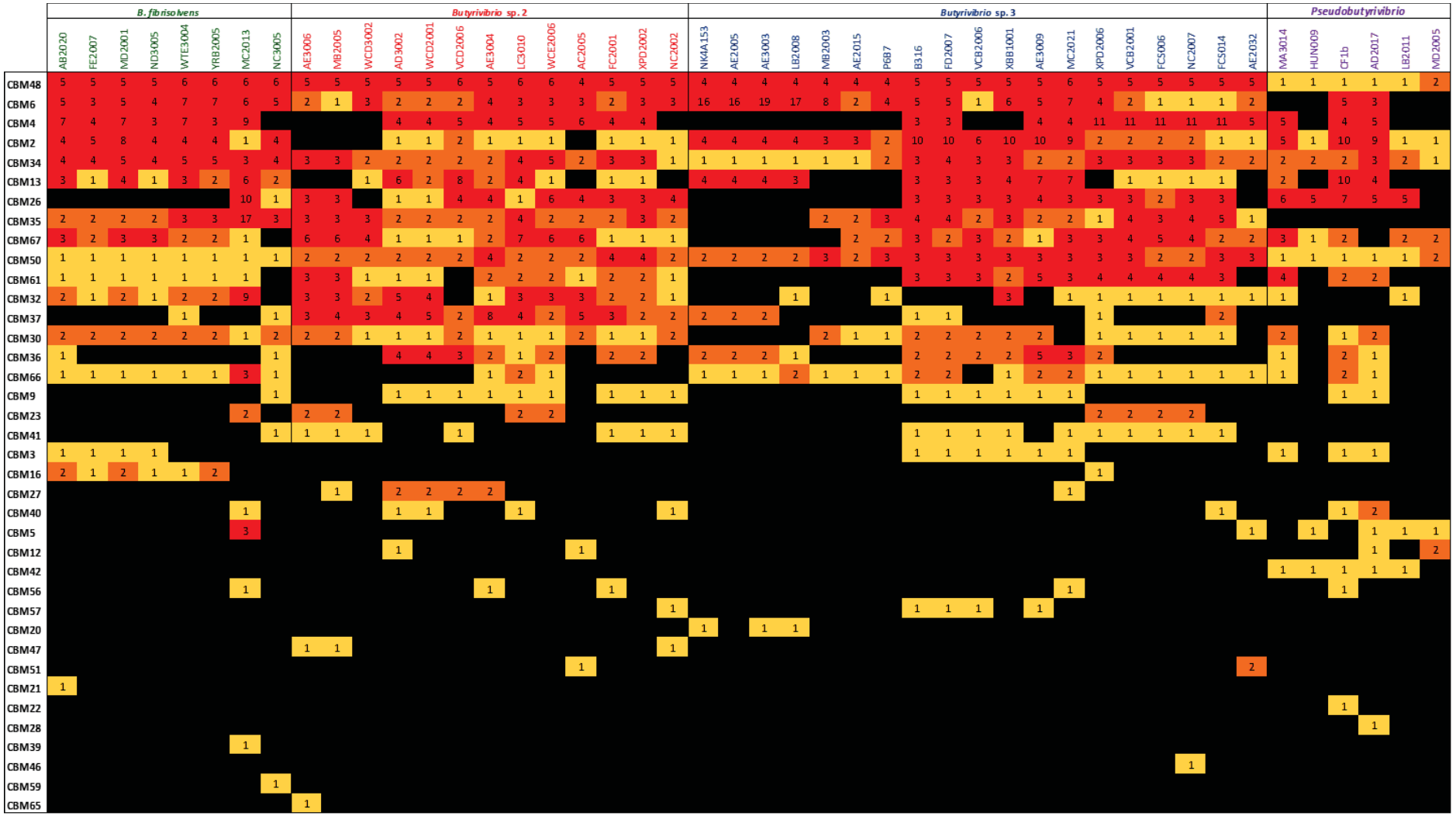


Figure 4.12. Heatmap showing the abundances of CBM families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes. The degree of redness indicates relative abundance and black indicates an absence of that CBM.

4.6.2 Glycoside Hydrolases (GHs)

GHs hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety, such as a protein, or a lipid (Cantarel *et al.*, 2009). To date, GHs are grouped into 135 families based on amino acid sequence in the CAZy database. Our analysis of the GH families showed distinct variation in relative abundance and distribution in both *Butyrivibrio* and *Pseudobutyrivibrio*. Among the GHs detected, the most abundant were GHs 13, 2, 43 and 3 (Figure 4.13). Only a few strains possessed GHs 55, 65, 76, 37, 117, 15, 85, 24 and 74, while strain MC2013 displayed high relative abundance of GHs 16 and 18, and was the only strain to possess GHs 128, 87, 81, 12, and 64.

The GHs that commonly contribute strongly to the degradation of polysaccharides, such as GHs 5, 51 and 67, and xylanases in GHs 8, 10 and 11, were examined in detail, based on their Pfam domains. The GH5 family endoglucanases were among the most abundant and heterogeneous of the CAZyme families for both *Butyrivibrio* and *Pseudobutyrivibrio* and they predominantly contain a single Pfam domain (Pfam00150) corresponding with the GH5 catalytic domain (Figure 4.14). However, the exact function or allocation to a specific GH5 subfamily of these CAZymes is unknown. Some GH5 endoglucanases of *B. fibrisolvens* strains, as well as *Butyrivibrio* sp. 3 strains B316 and MB2003 also contain a CBM2 domain at the N- or C- terminal end of the enzyme. Enzymes that contain a GT2 domain (Pfam00535) in addition to GH5 are deemed glycosyltransferases by the IMG functional annotation system. Strain MC2013 has the only predicted endo-β-mannanase, consisting of a GH5 and CBM11 domain (1,486 aa in size), and strain MC2021 has a unique endoglucanase consisting of GH5, GH9, CBM2 and CBM3 domains (1,294 aa in size).

The α-L-arabinofuranosidases of the GH51 family (Pfam06964) are abundant in all *Butyrivibrio* and *Pseudobutyrivibrio* except *B. fibrisolvens* MC2013 (Figure 4.15). The *B. fibrisolvens* YRB2005 and WTE3004 strains also encode α-L-arabinofuranosidases (containing Pfam13385; A-like lectin/glucanases superfamily) including a predicted protein corresponding to a LaminiG3, which is 1,266 aa in size and contains a signal peptide. *Butyrivibrio* sp. 3 strains MB2003 and AE2015 of both encode a 640 aa enzyme containing a GH51 and a CBM4 domain (Pfam02018). All of the α-L-arabinofurano-sidases examined contain a signal peptide suggesting they exert their activity outside the cell.

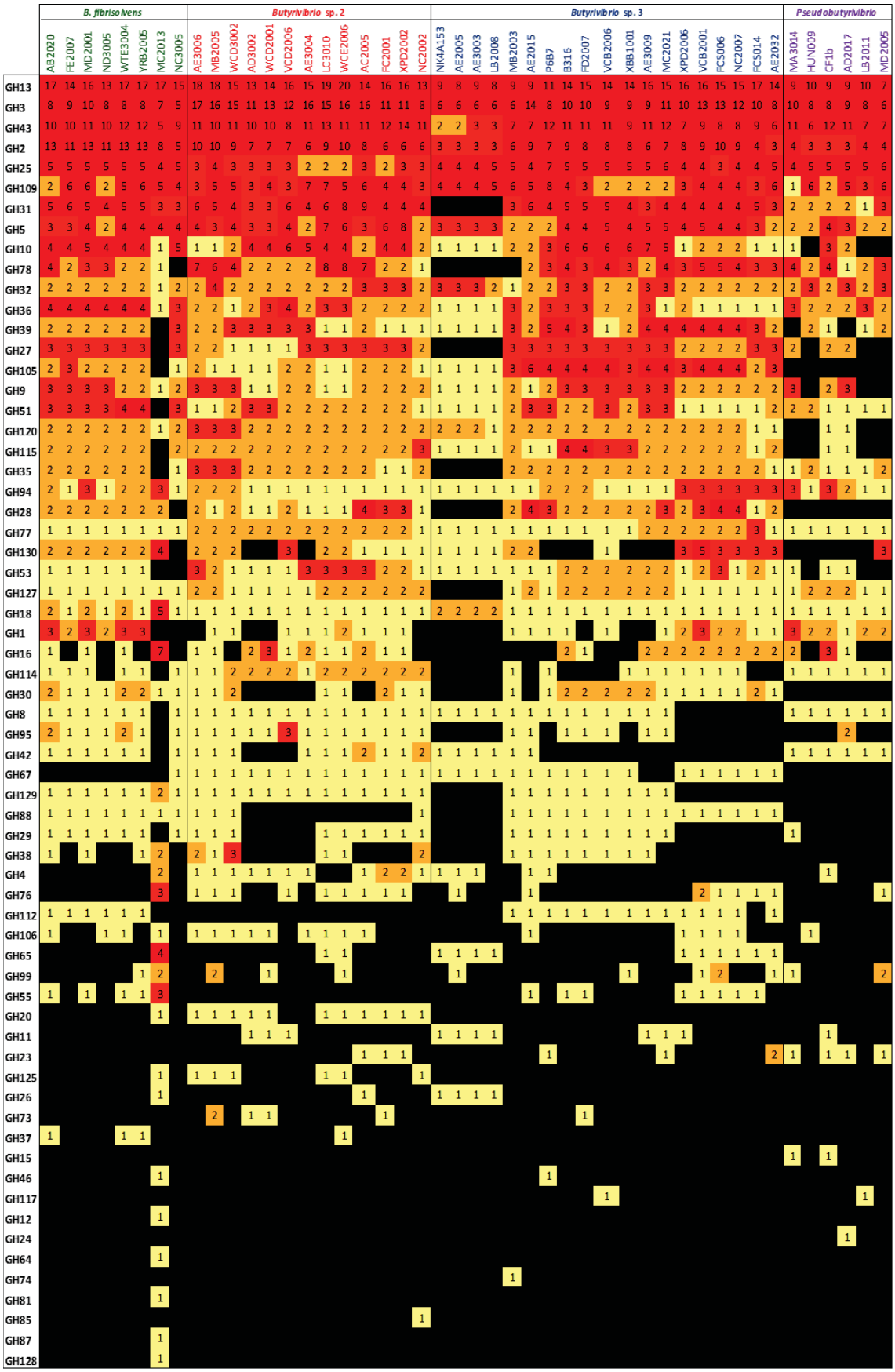
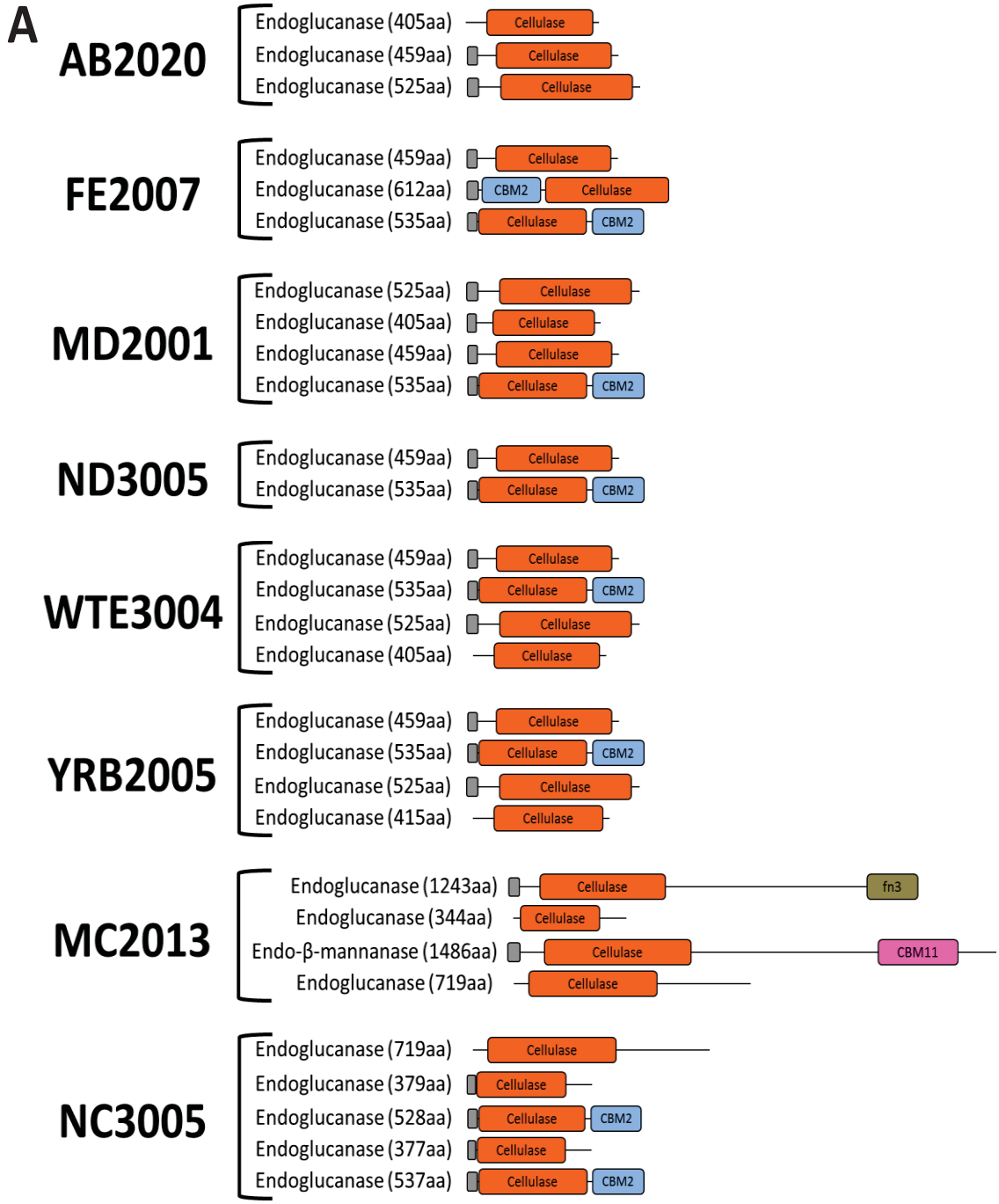


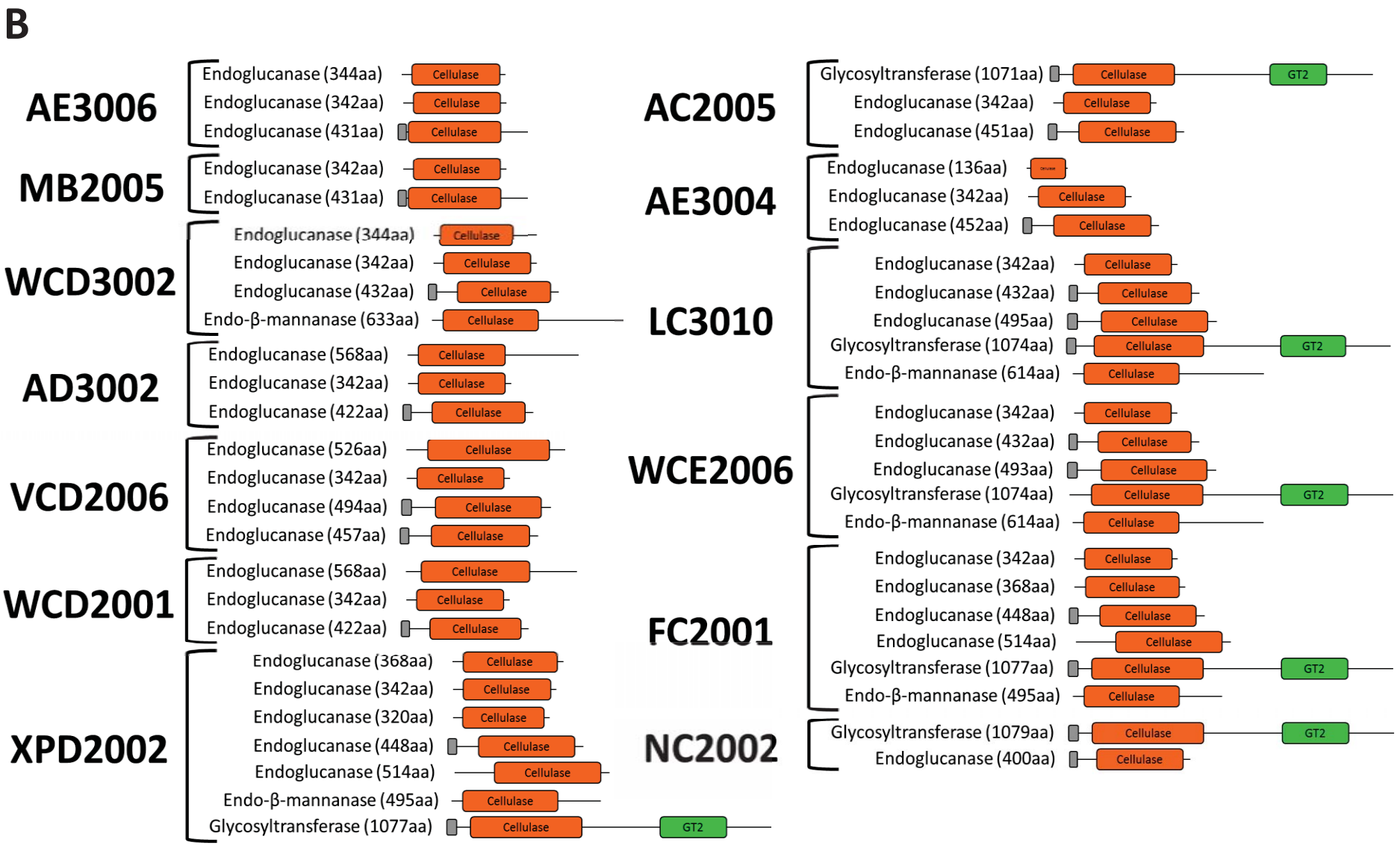
Figure 4.13. Heatmap showing the abundances of GH families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes. The degree of redness indicates relative abundance and black indicates an absence of that GH.

The α-glucuronidases of GH67 family are approximately 660 aa in size and consist of GH67N (Pfam03648), GH67M (Pfam07488) and GH67C (Pfam07477) domains (Figure A2.9). They have no signal peptides and are largely absent from *B. fibrisolvens* (except NC3005) and completely absent in all *Pseudobutyrivibrio* strains. The GH8 CAZymes that have activities associated with reducing end xylose-releasing exo-oligoxylanases or endoglucanase Y containing Pfam01270 domains (Figure A2.10), are approximately 380 aa in size, do not encode signal peptides and are present in all *Pseudobutyrivibrio* and 23 *Butyrivibrio* strains.

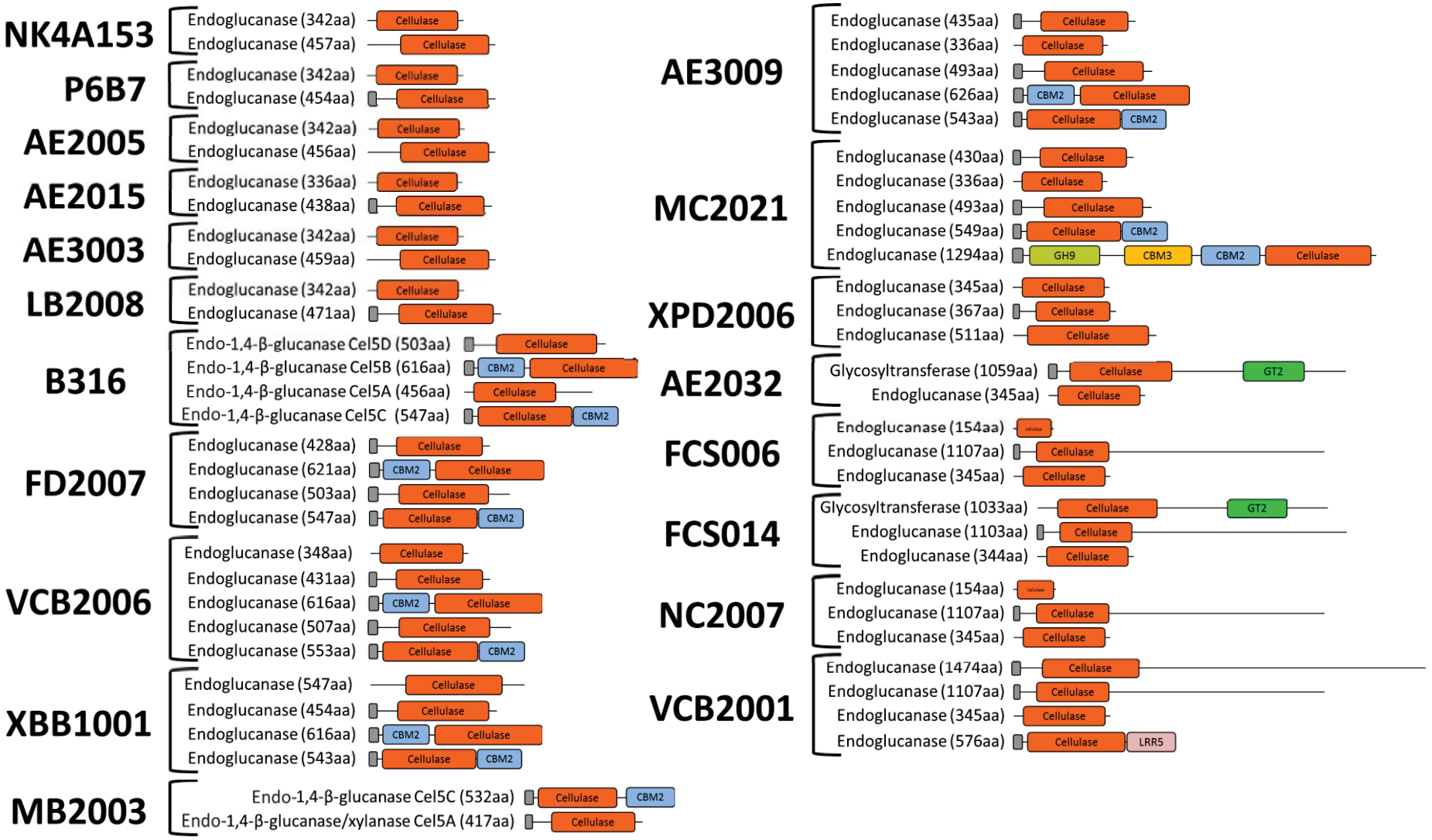
The GH10 are endo-1,4-β-xylanases that contain the Pfam00331 domain. GH10 CAZymes are present in all *Butyrivibrio* but are absent in *P. ruminis* HUN009, *Pseudobutyrivibrio* sp. strains LB2011 and MD2005 (Figure 4.16). Among *Butyrivibrio*, all *B. fibrisolvens* (except MC2013), six *Butyrivibrio* sp. 2 and 8 *Butyrivibrio* sp. 3 strains, contain four or more GH10 genes, where B316 and MB2003 have six and two, respectively. The GH10 enzymatic domains are found with many different domains, and not all these proteins contain signal peptides. Examples include GH10 domains combined with CBM13 domains (Pfam14200, 600 aa) which are named Xyn10A; CBM13 and CBM2 domains (Pfam00553, 680 aa); Xyn10D with an esterase domain (Pfam00756, 690 aa), Xyn10B with CBM9 domains (Pfam02018, 1,295 aa), with GDSL lipase (Pfam13472), and with CBM9 containing cell wall binding repeats (Pfam01473). The examples of unique GH10 enzymes are the Mxy10-43A (1,332 aa) found in *Butyrivibrio* sp. 2 strains AD3002, WCD2001, AE3004, FC2001 and XPD2002, and *Butyrivibrio* sp. 3 strains B316, MC2021, FD2007, VCB2006, and XBB1001. These enzymes are composed of multiple domains including: two CBM6 (Pfam03422), CBM2 (Pfam00553), CBM13 (Pfam14200) and GH43 (Pfam04616) domains.

The GH11 enzymes are endo-xylanases and are associated with a Pfam00457 domain (Figure 4.17). The CAZyme and Pfam analysis of these genes confirm their presence in three *Butyrivibrio* sp. 2, seven *Butyrivibrio* sp. 3 and one *Pseudobutyrivibrio* strains. Some of these enzymes contain CE2 domains (cd01831, 680 aa), or additional CBM6 domains (Pfam03422, 770 aa), and all have signal peptides.


A
AB2020
Endoglucanase (405aa)
Cellulase
Endoglucanase (459aa)
Cellulase
Endoglucanase (525aa)
Cellulase
FE2007
Endoglucanase (459aa)
Cellulase
Endoglucanase (612aa)
CBM2
Cellulase
Endoglucanase (535aa)
Cellulase
CBM2
MD2001
Endoglucanase (525aa)
Cellulase
Endoglucanase (405aa)
Cellulase
Endoglucanase (459aa)
Cellulase
Endoglucanase (535aa)
Cellulase
CBM2
ND3005
Endoglucanase (459aa)
Cellulase
Endoglucanase (535aa)
Cellulase
CBM2
WTE3004
Endoglucanase (459aa)
Cellulase
Endoglucanase (535aa)
Cellulase
CBM2
Endoglucanase (525aa)
Cellulase
Endoglucanase (405aa)
Cellulase
YRB2005
Endoglucanase (459aa)
Cellulase
Endoglucanase (535aa)
Cellulase
CBM2
Endoglucanase (525aa)
Cellulase
Endoglucanase (415aa)
Cellulase
MC2013
Endoglucanase (1243aa)
Cellulase
fn3
Endoglucanase (344aa)
Cellulase
Endo-β-mannanase (1486aa)
Cellulase
CBM11
Endoglucanase (719aa)
Cellulase
NC3005
Endoglucanase (719aa)
Cellulase
Endoglucanase (379aa)
Cellulase
Endoglucanase (528aa)
Cellulase
CBM2
Endoglucanase (377aa)
Cellulase
Endoglucanase (537aa)
Cellulase
CBM2


B
AE3006
Endoglucanase (344aa)
Endoglucanase (342aa)
Endoglucanase (431aa)
Cellulase
MB2005
Endoglucanase (342aa)
Endoglucanase (431aa)
WCD3002
Endoglucanase (344aa)
Endoglucanase (342aa)
Endoglucanase (432aa)
Endo-β-mannanase (633aa)
AD3002
Endoglucanase (568aa)
Endoglucanase (342aa)
Endoglucanase (422aa)
VCD2006
Endoglucanase (526aa)
Endoglucanase (342aa)
Endoglucanase (494aa)
Endoglucanase (457aa)
WCD2001
Endoglucanase (568aa)
Endoglucanase (342aa)
Endoglucanase (422aa)
XPD2002
Endoglucanase (368aa)
Endoglucanase (342aa)
Endoglucanase (320aa)
Endoglucanase (448aa)
Endoglucanase (514aa)
Endo-β-mannanase (495aa)
Glycosyltransferase (1077aa)
GT2
AC2005
Glycosyltransferase (1071aa)
Endoglucanase (342aa)
Endoglucanase (451aa)
AE3004
Endoglucanase (136aa)
Endoglucanase (342aa)
Endoglucanase (452aa)
LC3010
Endoglucanase (342aa)
Endoglucanase (432aa)
Endoglucanase (495aa)
Glycosyltransferase (1074aa)
Endo-β-mannanase (614aa)
WCE2006
Endoglucanase (342aa)
Endoglucanase (432aa)
Endoglucanase (493aa)
Glycosyltransferase (1074aa)
Endo-β-mannanase (614aa)
FC2001
Endoglucanase (342aa)
Endoglucanase (368aa)
Endoglucanase (448aa)
Endoglucanase (514aa)
Glycosyltransferase (1077aa)
Endo-β-mannanase (495aa)
NC2002
Glycosyltransferase (1079aa)
Endoglucanase (400aa)

C


NK4A153
Endoglucanase (342aa)
Cellulase
Endoglucanase (457aa)
Cellulase
P6B7
Endoglucanase (342aa)
Cellulase
Endoglucanase (454aa)
Cellulase
AE2005
Endoglucanase (342aa)
Cellulase
Endoglucanase (456aa)
Cellulase
AE2015
Endoglucanase (336aa)
Cellulase
Endoglucanase (438aa)
Cellulase
AE3003
Endoglucanase (342aa)
Cellulase
Endoglucanase (459aa)
Cellulase
LB2008
Endoglucanase (342aa)
Cellulase
Endoglucanase (471aa)
Cellulase
B316
Endo-1,4-β-glucanase Cel5D (503aa)
Cellulase
Endo-1,4-β-glucanase Cel5B (616aa)
CBM2
Cellulase
Endo-1,4-β-glucanase Cel5A (456aa)
Cellulase
Endo-1,4-β-glucanase Cel5C (547aa)
Cellulase
CBM2
FD2007
Endoglucanase (428aa)
Cellulase
Endoglucanase (621aa)
CBM2
Cellulase
Endoglucanase (503aa)
Cellulase
Endoglucanase (547aa)
Cellulase
CBM2
VCB2006
Endoglucanase (348aa)
Cellulase
Endoglucanase (431aa)
Cellulase
Endoglucanase (616aa)
CBM2
Cellulase
Endoglucanase (507aa)
Cellulase
Endoglucanase (553aa)
Cellulase
CBM2
XBB1001
Endoglucanase (547aa)
Cellulase
Endoglucanase (454aa)
Cellulase
Endoglucanase (616aa)
CBM2
Cellulase
Endoglucanase (543aa)
Cellulase
CBM2
MB2003
Endo-1,4-β-glucanase Cel5C (532aa)
Cellulase
CBM2
Endo-1,4-β-glucanase/xylanase Cel5A (417aa)
Cellulase
AE3009
Endoglucanase (435aa)
Cellulase
Endoglucanase (336aa)
Cellulase
Endoglucanase (493aa)
Cellulase
Endoglucanase (626aa)
CBM2
Cellulase
Endoglucanase (543aa)
Cellulase
CBM2
MC2021
Endoglucanase (430aa)
Cellulase
Endoglucanase (336aa)
Cellulase
Endoglucanase (493aa)
Cellulase
Endoglucanase (549aa)
Cellulase
CBM2
Endoglucanase (1294aa)
GH9
CBM3
CBM2
Cellulase
XPD2006
Endoglucanase (345aa)
Cellulase
Endoglucanase (367aa)
Cellulase
Endoglucanase (511aa)
Cellulase
AE2032
Glycosyltransferase (1059aa)
Cellulase
GT2
Endoglucanase (345aa)
Cellulase
FCS006
Endoglucanase (154aa)
Endoglucanase (1107aa)
Cellulase
Endoglucanase (345aa)
Cellulase
FCS014
Glycosyltransferase (1033aa)
Cellulase
GT2
Endoglucanase (1103aa)
Cellulase
Endoglucanase (344aa)
Cellulase
NC2007
Endoglucanase (154aa)
Endoglucanase (1107aa)
Cellulase
Endoglucanase (345aa)
Cellulase
VCB2001
Endoglucanase (1474aa)
Cellulase
Endoglucanase (1107aa)
Cellulase
Endoglucanase (345aa)
Cellulase
Endoglucanase (576aa)
Cellulase
LRR5

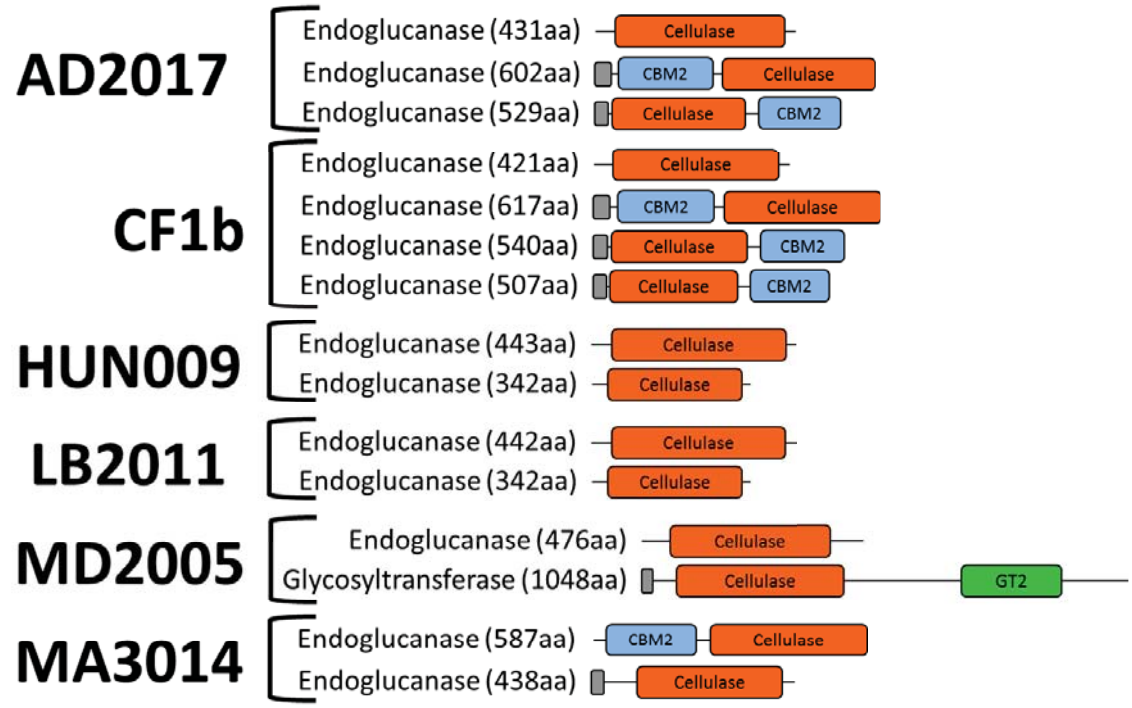
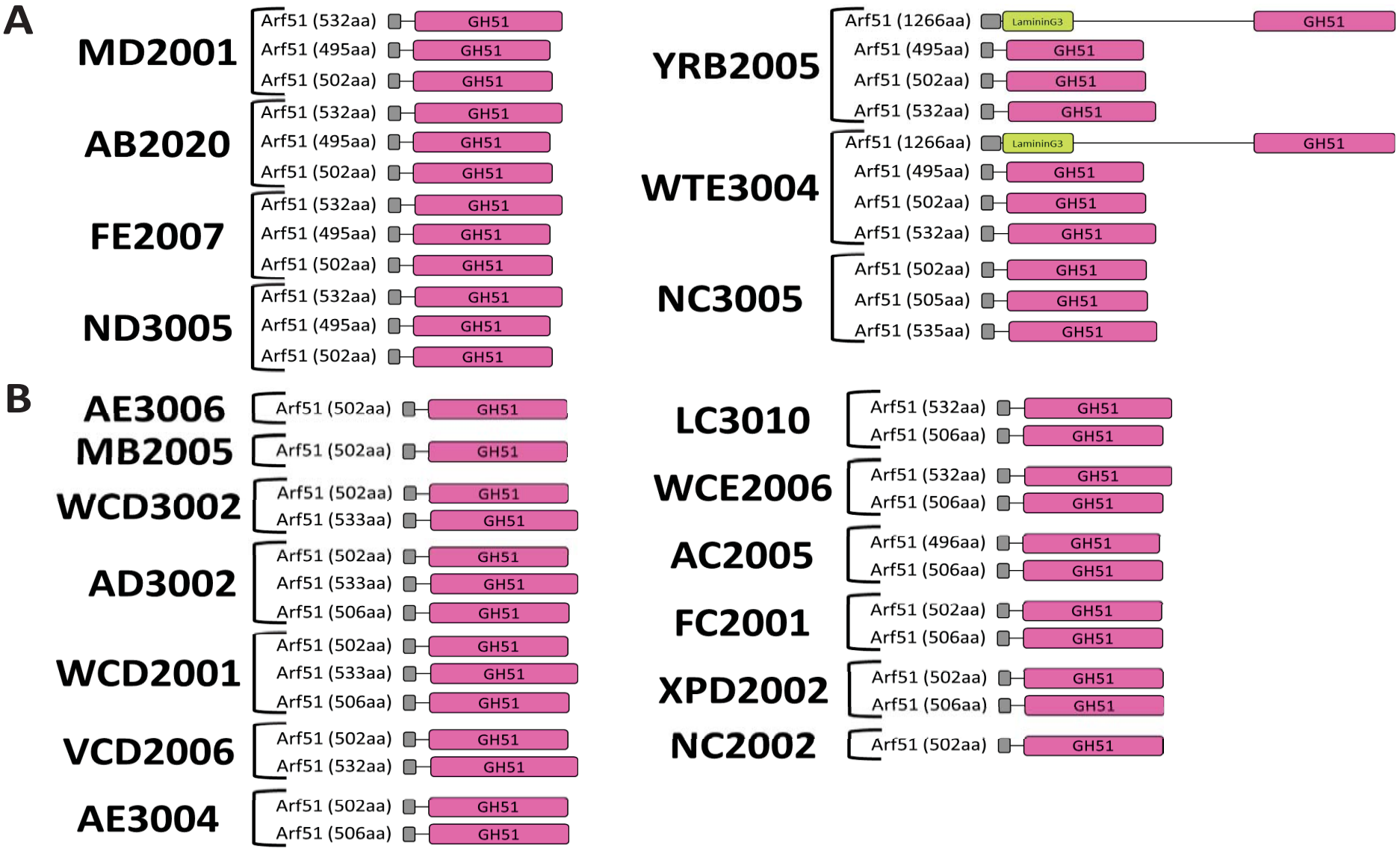


Figure 4.14. Predicted Pfam domains of GH5 endoglucanases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: Cellulase, glycoside hydrolase family 5 (Pfam00150); CBM2, carbohydrate-binding module 2 (Pfam00553); CBM3, carbohydrate-binding module 3 (Pfam00942); GH9, glycoside hydrolase family 9 (Pfam00759); fn3, fibronectin type III domain (Pfam00041); GT2, glycosyl transferase family 2 (Pfam13641); LRR5, leucine rich repeats (6 copies) (Pfam13306). ▮, signal peptide sequences.


A
MD2001
Arf51 (532aa)
GH51
Arf51 (495aa)
GH51
Arf51 (502aa)
GH51
AB2020
Arf51 (532aa)
GH51
Arf51 (495aa)
GH51
Arf51 (502aa)
GH51
FE2007
Arf51 (532aa)
GH51
Arf51 (495aa)
GH51
Arf51 (502aa)
GH51
ND3005
Arf51 (532aa)
GH51
Arf51 (495aa)
GH51
Arf51 (502aa)
GH51
YRB2005
Arf51 (1266aa)
LaminiG3
GH51
Arf51 (495aa)
GH51
Arf51 (502aa)
GH51
Arf51 (532aa)
GH51
WTE3004
Arf51 (1266aa)
LaminiG3
GH51
Arf51 (495aa)
GH51
Arf51 (502aa)
GH51
Arf51 (532aa)
GH51
NC3005
Arf51 (502aa)
GH51
Arf51 (505aa)
GH51
Arf51 (535aa)
GH51
B
AE3006
Arf51 (502aa)
GH51
MB2005
Arf51 (502aa)
GH51
WCD3002
Arf51 (502aa)
GH51
Arf51 (533aa)
GH51
AD3002
Arf51 (502aa)
GH51
Arf51 (533aa)
GH51
Arf51 (506aa)
GH51
WCD2001
Arf51 (502aa)
GH51
Arf51 (533aa)
GH51
Arf51 (506aa)
GH51
VCD2006
Arf51 (502aa)
GH51
Arf51 (532aa)
GH51
AE3004
Arf51 (502aa)
GH51
Arf51 (506aa)
GH51
LC3010
Arf51 (532aa)
GH51
Arf51 (506aa)
GH51
WCE2006
Arf51 (532aa)
GH51
Arf51 (506aa)
GH51
AC2005
Arf51 (496aa)
GH51
Arf51 (506aa)
GH51
FC2001
Arf51 (502aa)
GH51
Arf51 (506aa)
GH51
XPD2002
Arf51 (502aa)
GH51
Arf51 (506aa)
GH51
NC2002
Arf51 (502aa)
GH51

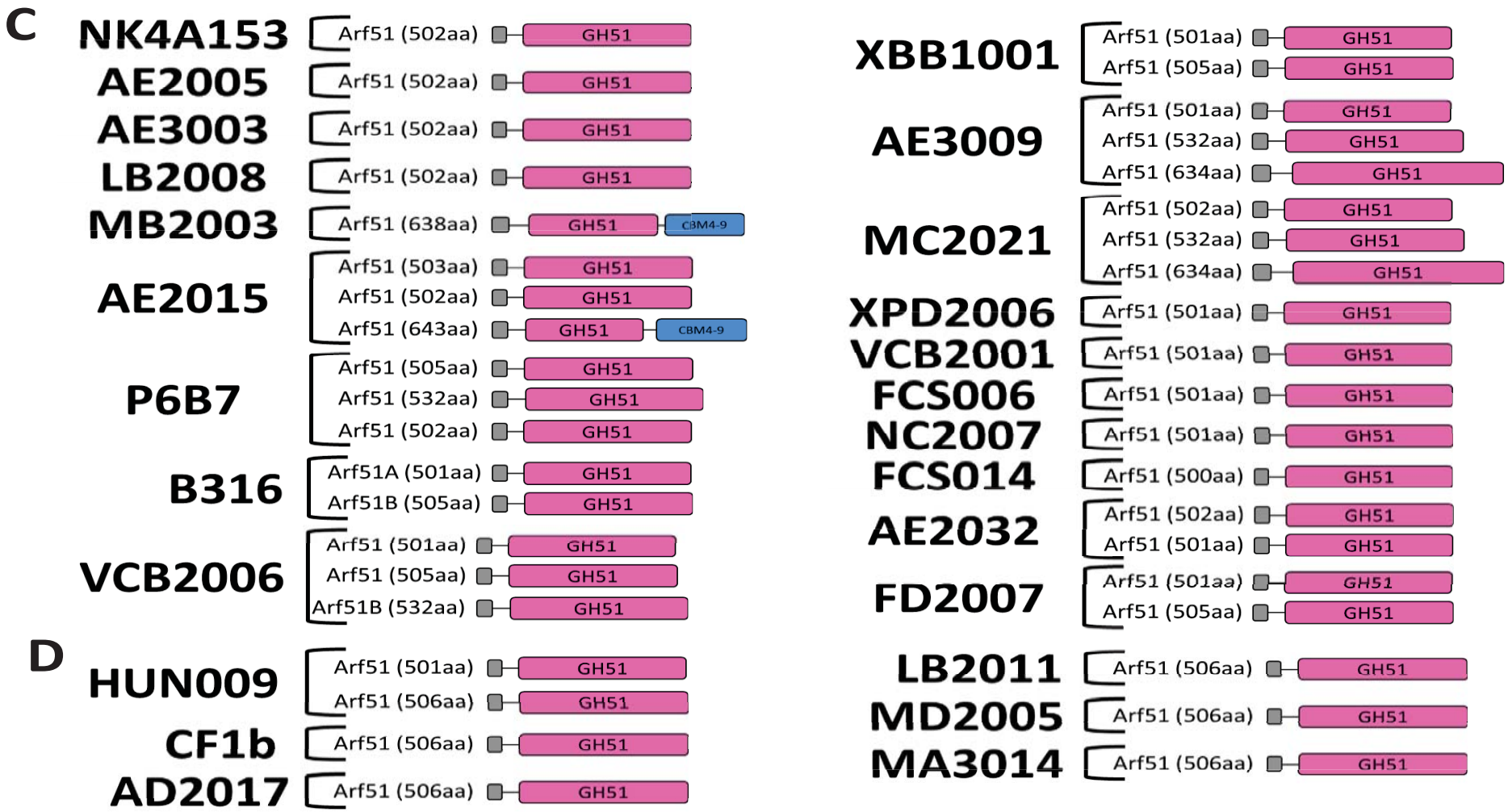
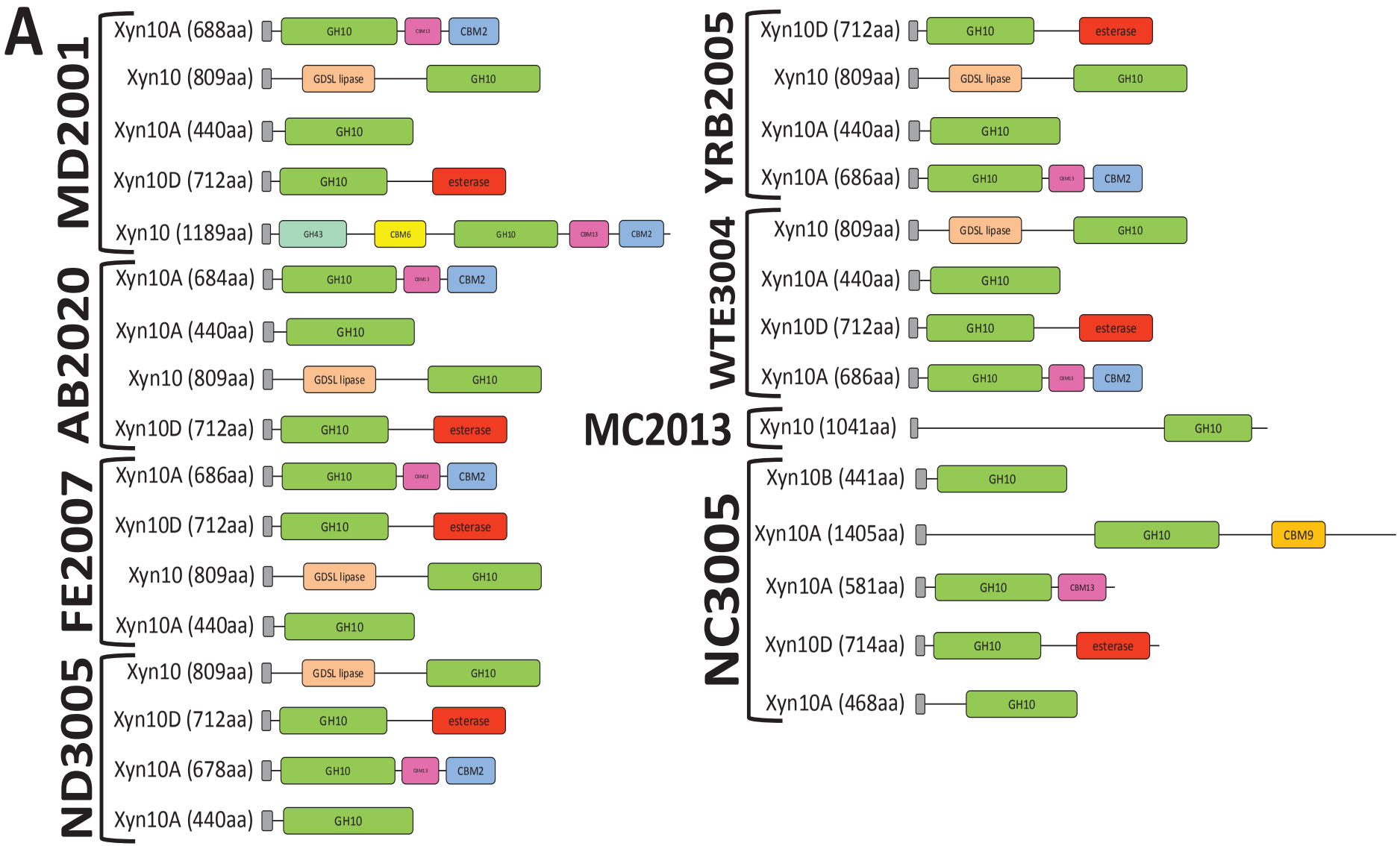
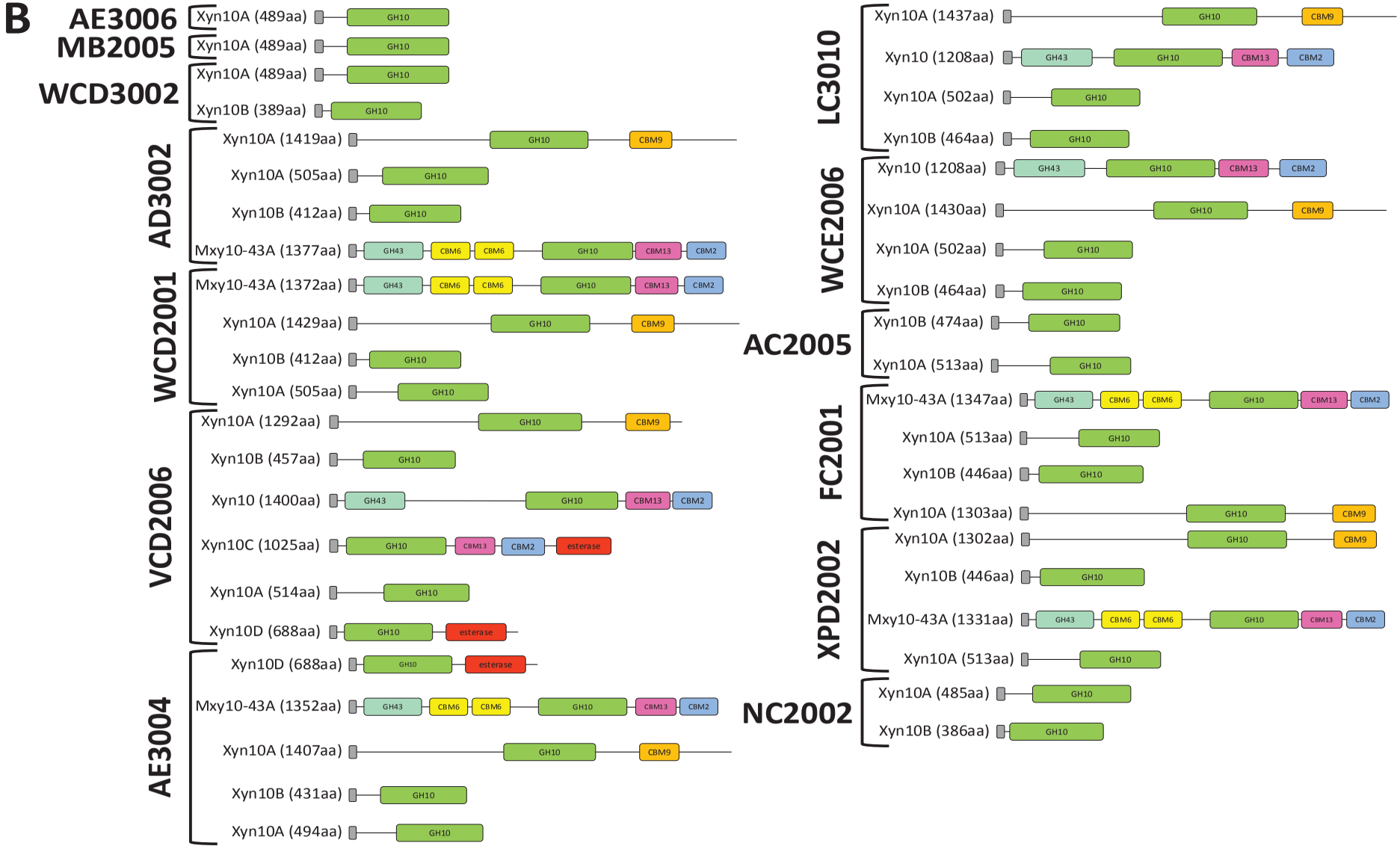
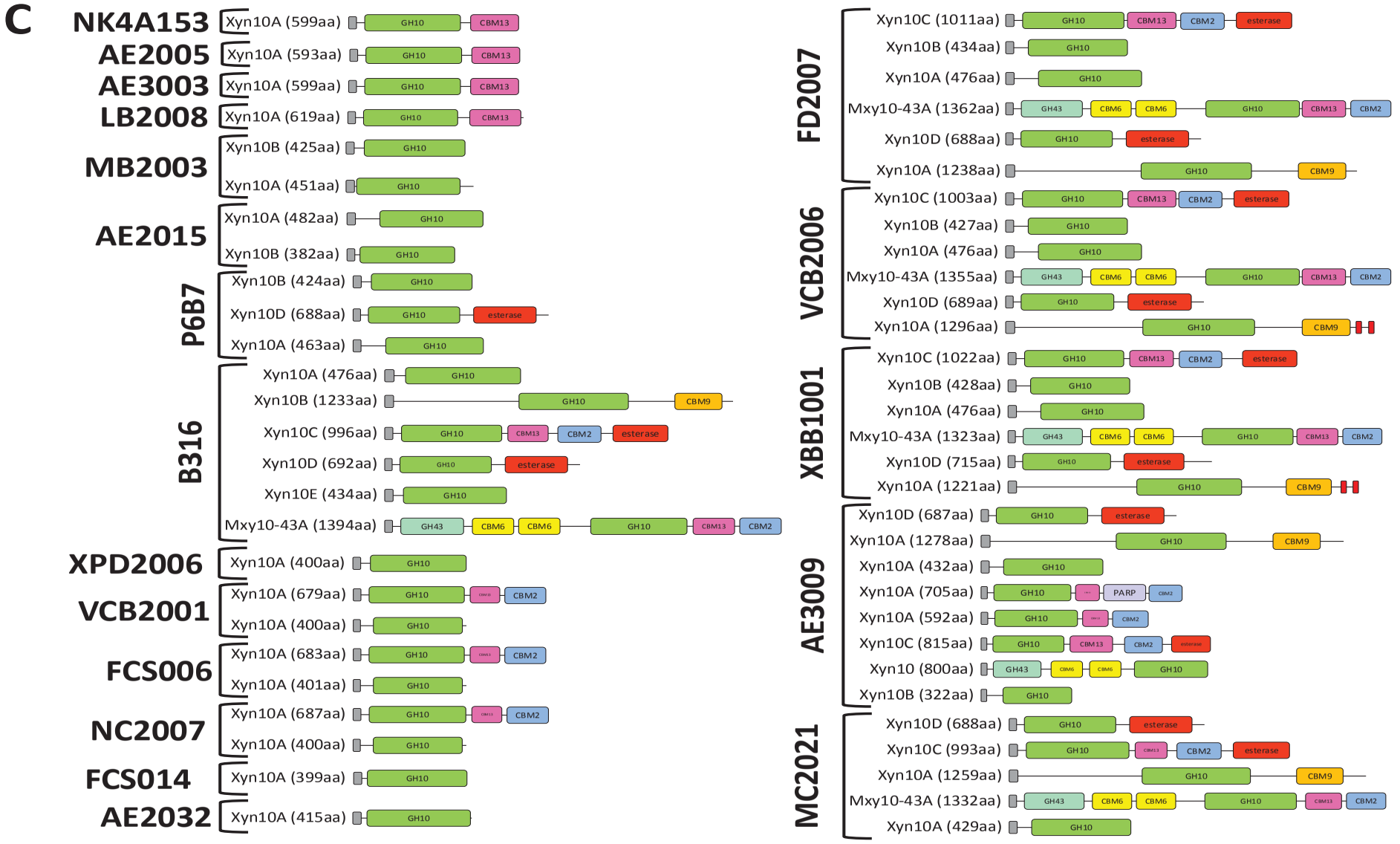


Figure 4.15. Predicted Pfam domains of GH51 α-L-arabinofuranosidases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: GH51, Arf51 - α-L-arabinofuranosidase C-terminus (Pfam06964); LaminiG3, concanavalin A-like lectin/glucanases superfamily (Pfam13385); CBM4-9, carbohydrate-binding module 4-9 (Pfam02018). ▯, signal peptide sequences.


A
MD2001
Xyn10A (688aa)
GH10
CBM13
CBM2
Xyn10 (809aa)
GDSL lipase
GH10
Xyn10A (440aa)
GH10
Xyn10D (712aa)
GH10
esterase
Xyn10 (1189aa)
GH43
CBM6
GH10
CBM13
CBM2
AB2020
Xyn10A (684aa)
GH10
CBM13
CBM2
Xyn10A (440aa)
GH10
Xyn10 (809aa)
GDSL lipase
GH10
Xyn10D (712aa)
GH10
esterase
FE2007
Xyn10A (686aa)
GH10
CBM13
CBM2
Xyn10D (712aa)
GH10
esterase
Xyn10 (809aa)
GDSL lipase
GH10
Xyn10A (440aa)
GH10
ND3005
Xyn10 (809aa)
GDSL lipase
GH10
Xyn10D (712aa)
GH10
esterase
Xyn10A (678aa)
GH10
CBM13
CBM2
Xyn10A (440aa)
GH10
YRB2005
Xyn10D (712aa)
GH10
esterase
Xyn10 (809aa)
GDSL lipase
GH10
Xyn10A (440aa)
GH10
Xyn10A (686aa)
GH10
CBM13
CBM2
WTE3004
Xyn10 (809aa)
GDSL lipase
GH10
Xyn10A (440aa)
GH10
Xyn10D (712aa)
GH10
esterase
Xyn10A (686aa)
GH10
CBM13
CBM2
MC2013
Xyn10 (1041aa)
GH10
NC3005
Xyn10B (441aa)
GH10
Xyn10A (1405aa)
GH10
CBM9
Xyn10A (581aa)
GH10
CBM13
Xyn10D (714aa)
GH10
esterase
Xyn10A (468aa)
GH10


B
AE3006
MB2005
WCD3002
AD3002
WCD2001
VCD2006
AE3004
LC3010
WCE2006
AC2005
FC2001
XPD2002
NC2002
Xyn10A (489aa)
Xyn10A (489aa)
Xyn10A (489aa)
Xyn10B (389aa)
Xyn10A (1419aa)
Xyn10A (505aa)
Xyn10B (412aa)
Mxy10-43A (1377aa)
Mxy10-43A (1372aa)
Xyn10A (1429aa)
Xyn10B (412aa)
Xyn10A (505aa)
Xyn10A (1292aa)
Xyn10B (457aa)
Xyn10 (1400aa)
Xyn10C (1025aa)
Xyn10A (514aa)
Xyn10D (688aa)
Xyn10D (688aa)
Mxy10-43A (1352aa)
Xyn10A (1407aa)
Xyn10B (431aa)
Xyn10A (494aa)
Xyn10A (1437aa)
Xyn10 (1208aa)
Xyn10A (502aa)
Xyn10B (464aa)
Xyn10 (1208aa)
Xyn10A (1430aa)
Xyn10A (502aa)
Xyn10B (464aa)
Xyn10B (474aa)
Xyn10A (513aa)
Mxy10-43A (1347aa)
Xyn10A (513aa)
Xyn10B (446aa)
Xyn10A (1303aa)
Xyn10A (1302aa)
Xyn10B (446aa)
Mxy10-43A (1331aa)
Xyn10A (513aa)
Xyn10A (485aa)
Xyn10B (386aa)
GH10
GH43
CBM6
CBM13
CBM2
CBM9
esterase


C
NK4A153
Xyn10A (599aa)
GH10
CBM13
AE2005
Xyn10A (593aa)
GH10
CBM13
AE3003
Xyn10A (599aa)
GH10
CBM13
LB2008
Xyn10A (619aa)
GH10
CBM13
MB2003
Xyn10B (425aa)
GH10
Xyn10A (451aa)
GH10
AE2015
Xyn10A (482aa)
GH10
Xyn10B (382aa)
GH10
P6B7
Xyn10B (424aa)
GH10
Xyn10D (688aa)
GH10
esterase
Xyn10A (463aa)
GH10
B316
Xyn10A (476aa)
GH10
Xyn10B (1233aa)
GH10
CBM9
Xyn10C (996aa)
GH10
CBM13
CBM2
esterase
Xyn10D (692aa)
GH10
esterase
Xyn10E (434aa)
GH10
Mxy10-43A (1394aa)
GH43
CBM6
CBM6
GH10
CBM13
CBM2
XPD2006
Xyn10A (400aa)
GH10
VCB2001
Xyn10A (679aa)
GH10
CBM2
Xyn10A (400aa)
GH10
FCS006
Xyn10A (683aa)
GH10
CBM2
Xyn10A (401aa)
GH10
NC2007
Xyn10A (687aa)
GH10
CBM2
Xyn10A (400aa)
GH10
FCS014
Xyn10A (399aa)
GH10
AE2032
Xyn10A (415aa)
GH10
FD2007
Xyn10C (1011aa)
GH10
CBM13
CBM2
esterase
Xyn10B (434aa)
GH10
Xyn10A (476aa)
GH10
Mxy10-43A (1362aa)
GH43
CBM6
CBM6
GH10
CBM13
CBM2
Xyn10D (688aa)
GH10
esterase
Xyn10A (1238aa)
GH10
CBM9
VCB2006
Xyn10C (1003aa)
GH10
CBM13
CBM2
esterase
Xyn10B (427aa)
GH10
Xyn10A (476aa)
GH10
Mxy10-43A (1355aa)
GH43
CBM6
CBM6
GH10
CBM13
CBM2
Xyn10D (689aa)
GH10
esterase
Xyn10A (1296aa)
GH10
CBM9
XBB1001
Xyn10C (1022aa)
GH10
CBM13
CBM2
esterase
Xyn10B (428aa)
GH10
Xyn10A (476aa)
GH10
Mxy10-43A (1323aa)
GH43
CBM6
CBM6
GH10
CBM13
CBM2
Xyn10D (715aa)
GH10
esterase
Xyn10A (1221aa)
GH10
CBM9
AE3009
Xyn10D (687aa)
GH10
esterase
Xyn10A (1278aa)
GH10
CBM9
Xyn10A (432aa)
GH10
Xyn10A (705aa)
GH10
PARP
CBM2
Xyn10A (592aa)
GH10
CBM2
Xyn10C (815aa)
GH10
CBM13
CBM2
esterase
Xyn10 (800aa)
GH43
CBM6
CBM6
GH10
Xyn10B (322aa)
GH10
MC2021
Xyn10D (688aa)
GH10
esterase
Xyn10C (993aa)
GH10
CBM13
CBM2
esterase
Xyn10A (1259aa)
GH10
CBM9
Mxy10-43A (1332aa)
GH43
CBM6
CBM6
GH10
CBM13
CBM2
Xyn10A (429aa)
GH10

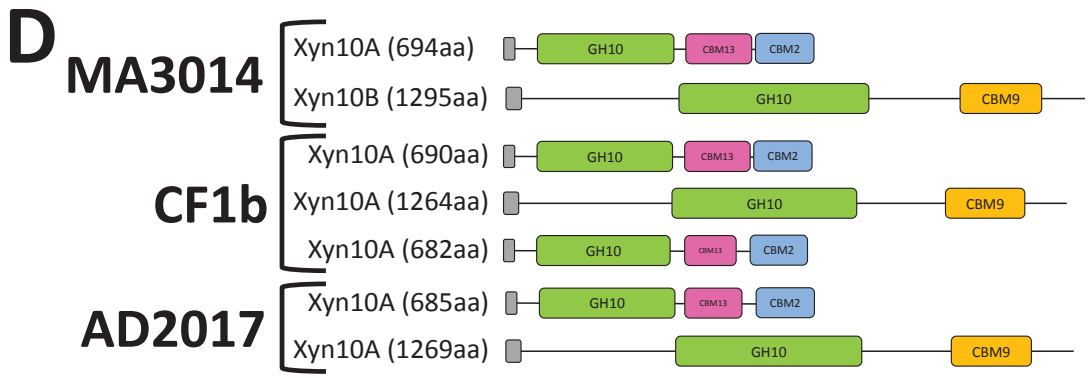


Figure 4.16. Predicted Pfam domains of GH10 xylanases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: Esterase, feruloyl esterase (CE1) (Pfam00756); CBM2, carbohydrate-binding module 2 (Pfam00553); CBM6, carbohydrate-binding module 6 (Pfam03422); CBM9, carbohydrate-binding module 9 (Pfam06452); CBM13, carbohydrate-binding module 13 (Pfam14200); GH10, glycoside hydrolase family 10 (Pfam00331); GH43, glycoside hydrolases family 43 (Pfam04616); GDSL lipase, GDSL-like Lipase/ Acylhydrolase family (Pfam13472). ▯, signal peptide sequences. ▮, cell wall-binding repeat domain.

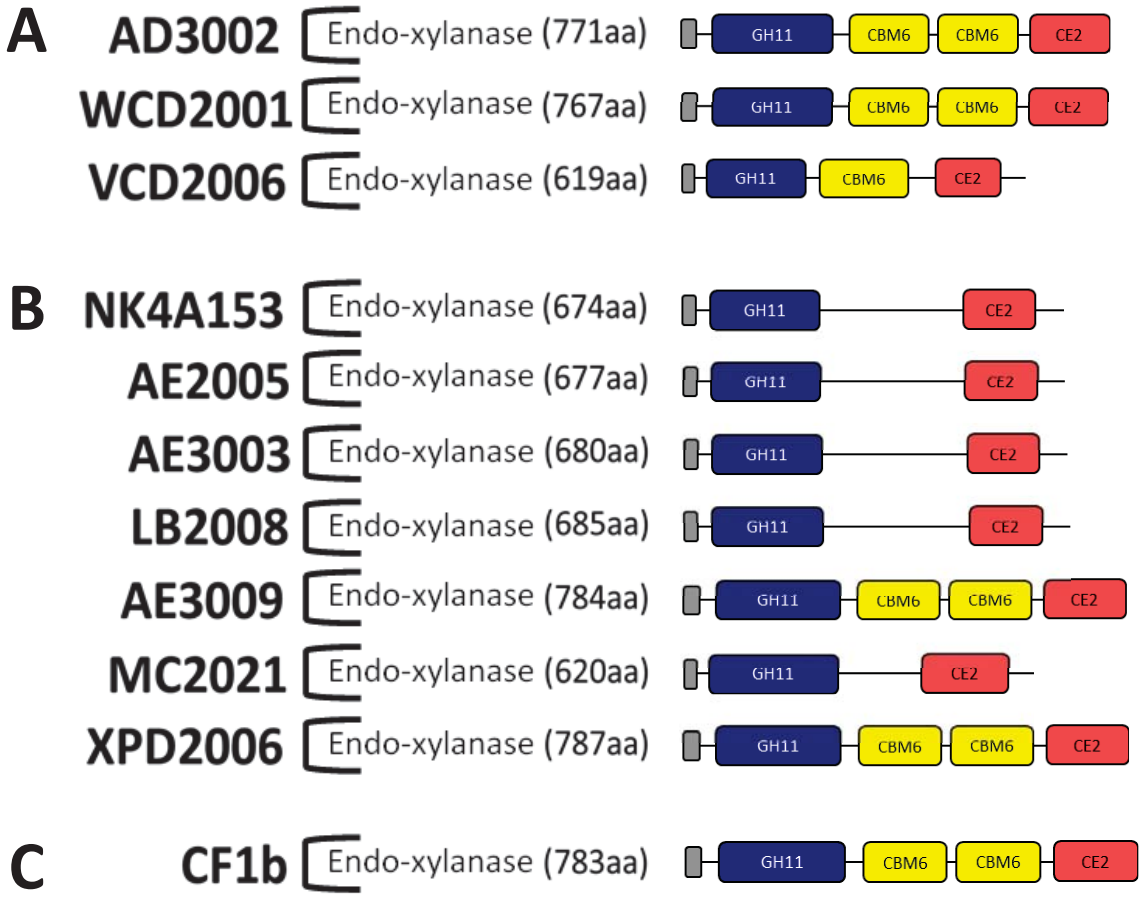


Figure 4.17. Predicted Pfam domains of GH11 endo-xylanases. A, *Butyrivibrio* sp. 2, B, *Butyrivibrio* sp. 3 and C, *Pseudobutyrivibrio*. Abbreviations: GH11, glycoside hydrolase family 11 (Pfam00457); CE2, carbohydrate esterase (CE2) (cd01831); CBM6, carbohydrate-binding module 6 (Pfam03422). ▯, signal peptide sequences.

4.6.3 Carbohydrate Esterases (CEs)

CEs catalyse the de-*O*- or de-*N*-acylation of esters or amides and other substituted saccharides in which sugars play a role of alcohol or amine (Biely, 2012). Our results showed that *Butyrivibrio* and *Pseudobutyrivibrio* encode 15 of the known 16 CE families, with family CE11 being the only missing CE. Families CE1, CE4 and CE10 associated with acetyl-xylan esterase and carboxyl esterase functions, were present in all *Butyrivibrio* and *Pseudobutyrivibrio* examined (Figure 4.18A). Strains WCD2001 and MD2001 were the only strains to encode a CE5 and a CE15, respectively. However, CE10 has subsequently been removed from the CAZy database because the vast majority of the members of this family are esterases acting on non-carbohydrate substrates, and therefore will not be further analysed. The most prevalent CE was the CE1. The CE1 (Pfam00756) domain corresponds to acetyl-xylan esterases and ferulic acid esterases (Figure 4.19). The strain WCD3002 encodes the most CE1 enzymes amongst the analysed genomes with 16, whereas B316, MB2003 and *P. xylanivorans* MA3014 have eight, four and 11, respectively. CE1 domains are heterogeneous and they are found with many different domains; some CE1 domain-containing proteins also have signal peptide sequences. The *B. fibrisolvens* strains all have multiple esterase genes, each composed of the CE1 domain only (mostly 260 to 380 aa), either with or without a signal peptide sequence. The *Butyrivibrio* sp. 2 is similar, except for strain VCD2006 that encodes a β-1,4-xylanase (1,022 aa) that has multiple domains, including a CBM2 (Pfam00553), CBM13 (Pfam14200) and GH10 (Pfam00331) along with the CE1.

Butyrivibrio sp. 3 is the most diverse group in regard to CEs, with examples of esterases that include CE1 combined with CBM6 (Pfam03422), CBM13 (Pfam14200) and GH43 (Pfam04616) domains (1,033 aa), CBM6 (Pfam03422), CBM13 (Pfam14200) and GH10 (Pfam00331) domains (1,010 aa), and a large 1,700 aa enzyme with a CE10 domain (Pfam07859). Strain MC2021 has the only gene found in this analysis that encodes for a CE1 and a GH114 domain (Pfam03537, 574 aa), where the GH114 domain has known activity as an endo-α-1,4-polygalactosaminidase. The *Pseudobutyrivibrio* strains *P. ruminis* AD2017 and CF1b encode a CE1 domain combined with CBM13 (Pfam14200) and CBM2 (Pfam00553) domains and has a signal peptide.

The CE4 are also members of the acetyl xylan esterase enzyme family, more specifically known as polysaccharide deacetylases (abbreviated to PD-Est4, Pfam01522). CE4 are present in all *Butyrivibrio* and *Pseudobutyrivibrio* genomes (Figure 4.20). The *Butyrivibrio* sp. 2 strains NK4A153, AE2005 and AE3003 encode 9 CE4s and LB2008

encodes 10, CE4 enzymes. These enzymes exist mainly in the form of a single PD-Est4 (Pfam01522) domain, without a signal peptide, and range in size from approximately 270 aa to 590 aa. The analysis has found more elaborate variations of these enzymes than reported before. For example, *Butyrivibrio* sp. 3 strains NK4A153, AE2005, AE3003 and LB2008 each have a large CE4 enzyme (1,038 aa) with a signal sequence and with three or more PD-Est4 domains.

4.6.4 Polysaccharide Lyases (PLs)

PLs were the only major CAZy class not present across all 46 strains of *Butyrivibrio* and *Pseudobutyrivibrio* (Figure 4.18B). There are currently no accurately curated Pfam domains for all the PL families. Among the 35 strains that contain PLs, only strains MC2021 and AE3009 from *Butyrivibrio* sp. 3 encode a PL10, the only PL known to be involved in lignocellulose breakdown. *Butyrivibrio* sp. 3 strains AE2005, AE3003, LB2008 and NK4A153 encode another PL family, PL6. The *B. proteoclasticus* B316 genome encodes PL1 (pectate lyase Pel1A), PL9 (pectate lyase Pel9A and pectin methylesterase Pme8B) and PL11 (carbohydrate esterase family 12 Est12B). The PL1 enzymes were otherwise rare among the *Butyrivibrio* and *Pseudobutyrivibrio*. In contrast, there was a clear trend for an abundance of PLs 9, 11 and 12 enzymes in *B. fibrisolvens* and *Butyrivibrio* sp. 3 groups.

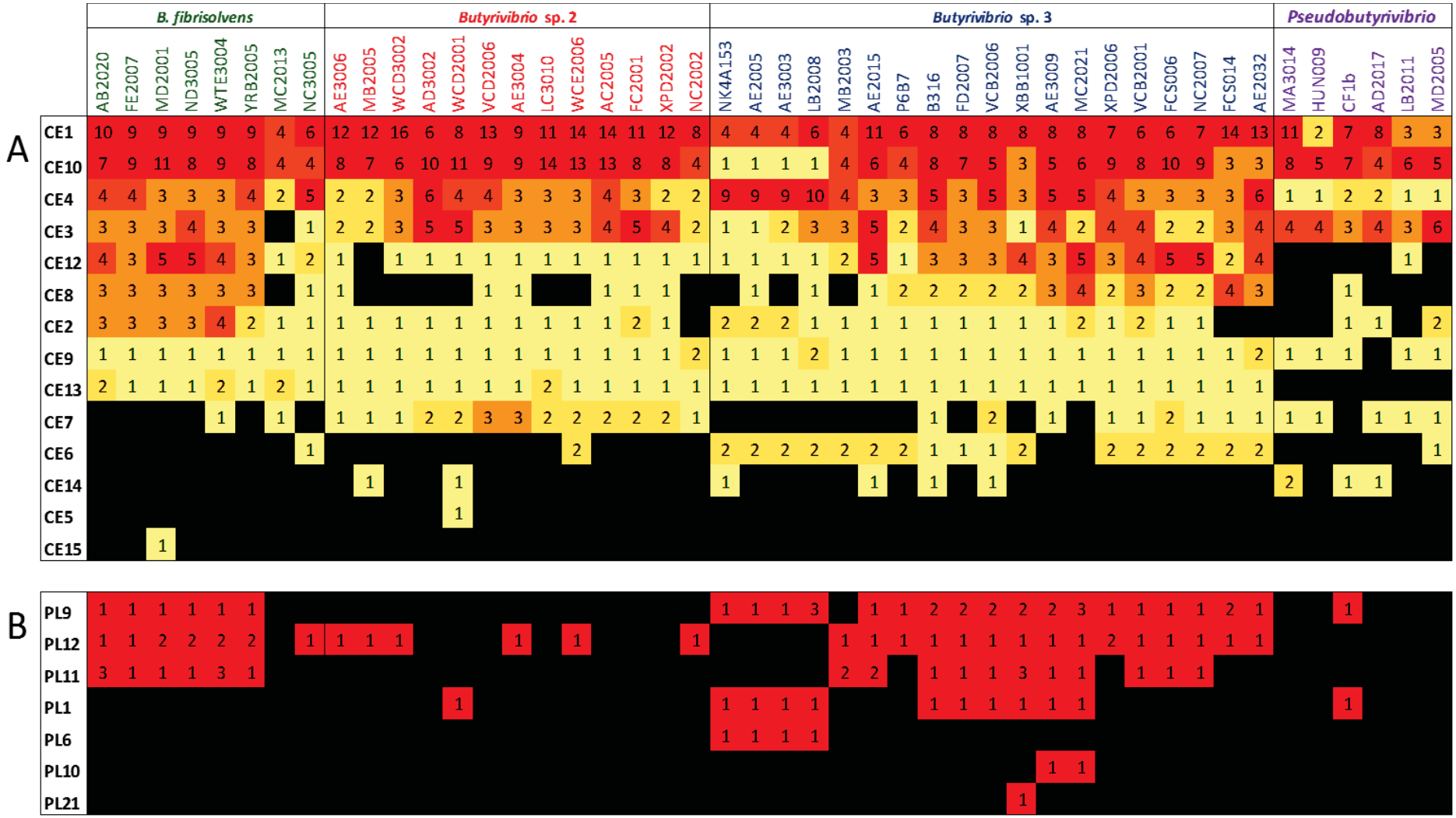
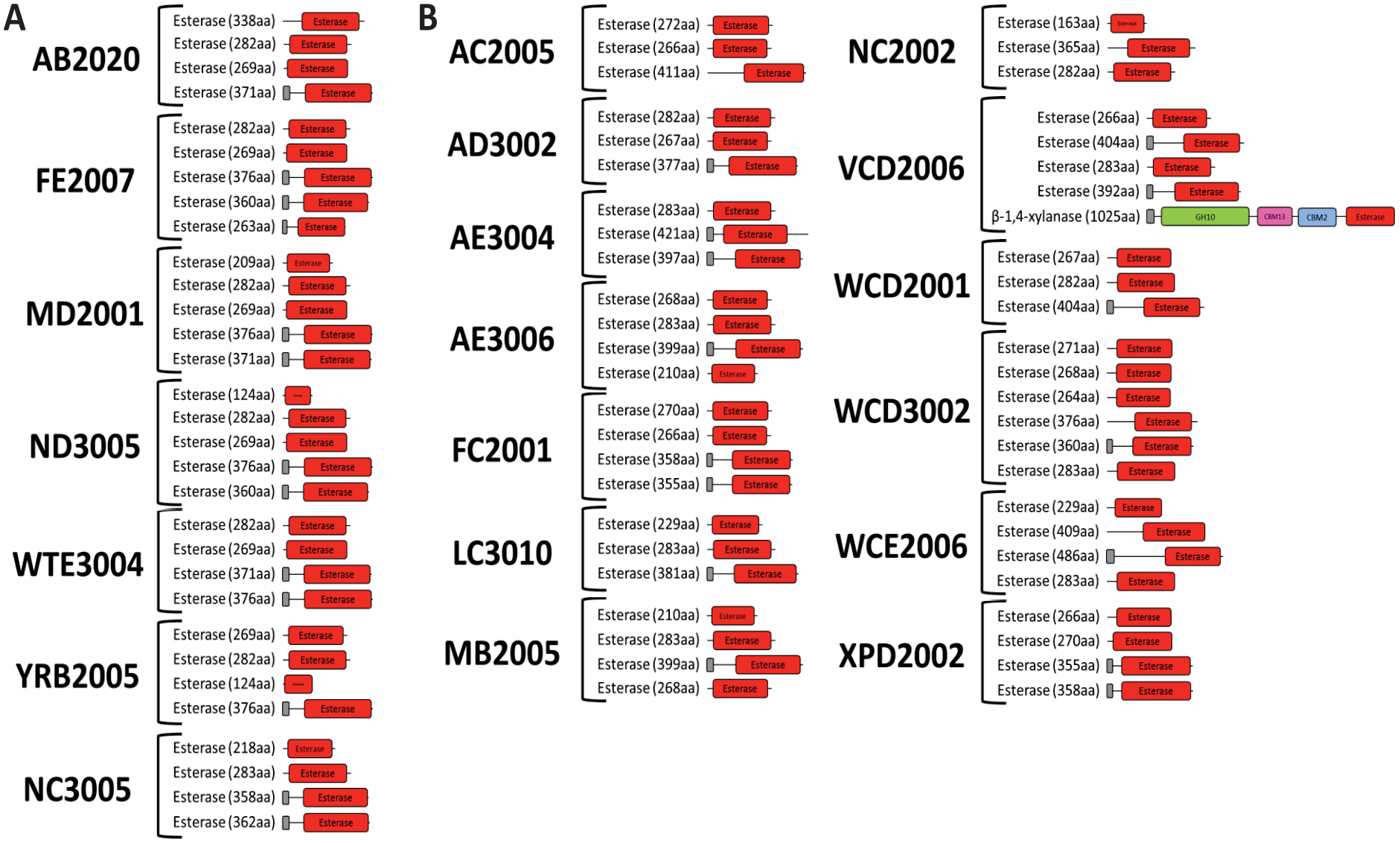
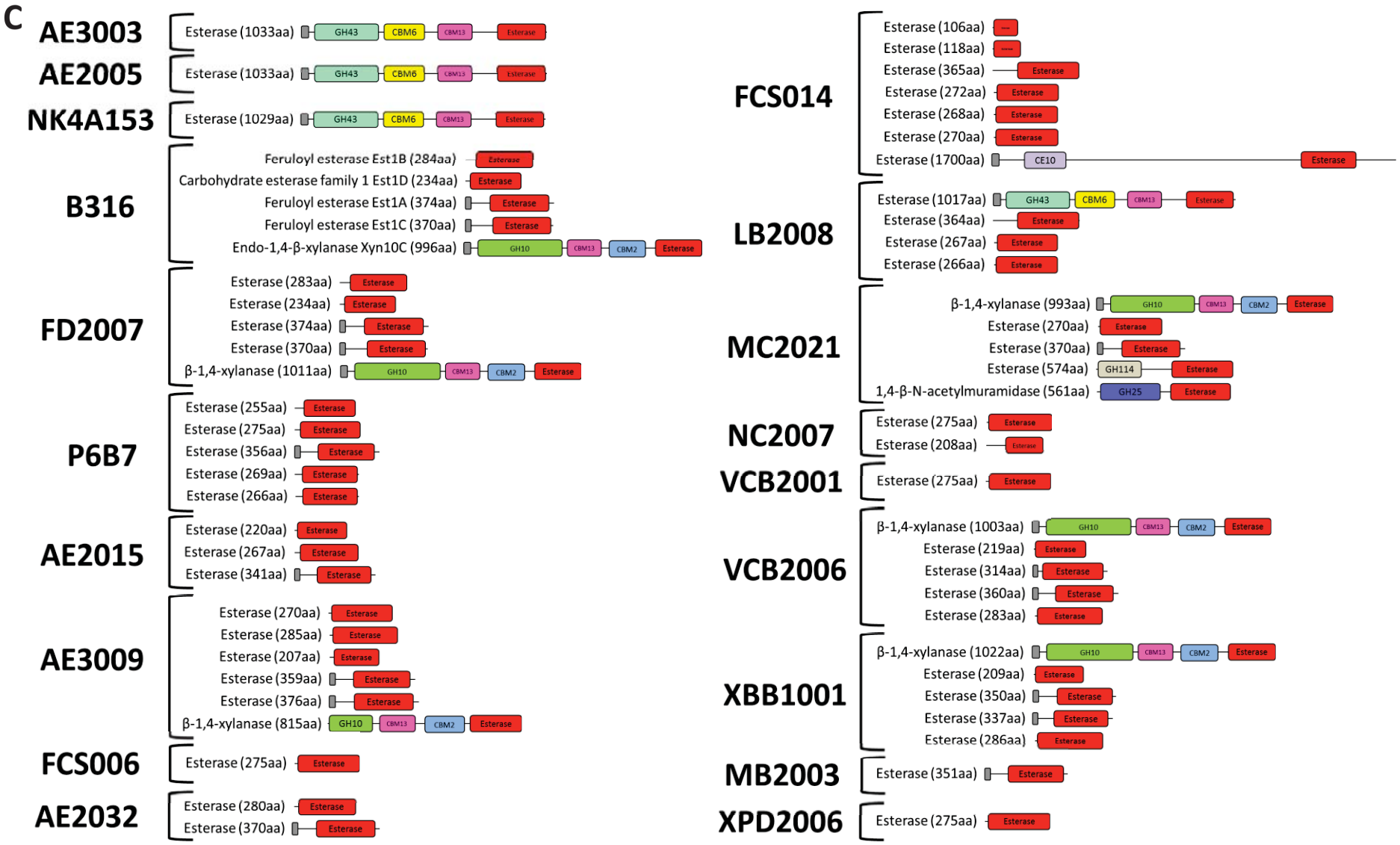


Figure 4.18. Heatmap showing the abundances of CE and PL families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes. A, CEs; B, PLs. The degree of redness indicates relative abundance and black indicates an absence of that CE or PL.


A
AB2020
Esterase (338aa)
Esterase (282aa)
Esterase (269aa)
Esterase (371aa)
FE2007
Esterase (282aa)
Esterase (269aa)
Esterase (376aa)
Esterase (360aa)
Esterase (263aa)
MD2001
Esterase (209aa)
Esterase (282aa)
Esterase (269aa)
Esterase (376aa)
Esterase (371aa)
ND3005
Esterase (124aa)
Esterase (282aa)
Esterase (269aa)
Esterase (376aa)
Esterase (360aa)
WTE3004
Esterase (282aa)
Esterase (269aa)
Esterase (371aa)
Esterase (376aa)
YRB2005
Esterase (269aa)
Esterase (282aa)
Esterase (124aa)
Esterase (376aa)
NC3005
Esterase (218aa)
Esterase (283aa)
Esterase (358aa)
Esterase (362aa)
B
AC2005
Esterase (272aa)
Esterase (266aa)
Esterase (411aa)
AD3002
Esterase (282aa)
Esterase (267aa)
Esterase (377aa)
AE3004
Esterase (283aa)
Esterase (421aa)
Esterase (397aa)
AE3006
Esterase (268aa)
Esterase (283aa)
Esterase (399aa)
Esterase (210aa)
FC2001
Esterase (270aa)
Esterase (266aa)
Esterase (358aa)
Esterase (355aa)
LC3010
Esterase (229aa)
Esterase (283aa)
Esterase (381aa)
MB2005
Esterase (210aa)
Esterase (283aa)
Esterase (399aa)
Esterase (268aa)
NC2002
Esterase (163aa)
Esterase (365aa)
Esterase (282aa)
VCD2006
Esterase (266aa)
Esterase (404aa)
Esterase (283aa)
Esterase (392aa)
β-1,4-xylanase (1025aa)
GH10
CBM13
CBM2
WCD2001
Esterase (267aa)
Esterase (282aa)
Esterase (404aa)
WCD3002
Esterase (271aa)
Esterase (268aa)
Esterase (264aa)
Esterase (376aa)
Esterase (360aa)
Esterase (283aa)
WCE2006
Esterase (229aa)
Esterase (409aa)
Esterase (486aa)
Esterase (283aa)
XPD2002
Esterase (266aa)
Esterase (270aa)
Esterase (355aa)
Esterase (358aa)


C
AE3003
Esterase (1033aa)
GH43
CBM6
CBM13
Esterase
AE2005
Esterase (1033aa)
GH43
CBM6
CBM13
Esterase
NK4A153
Esterase (1029aa)
GH43
CBM6
CBM13
Esterase
B316
Feruloyl esterase Est1B (284aa)
Esterase
Carbohydrate esterase family 1 Est1D (234aa)
Esterase
Feruloyl esterase Est1A (374aa)
Esterase
Feruloyl esterase Est1C (370aa)
Esterase
Endo-1,4-β-xylanase Xyn10C (996aa)
GH10
CBM13
CBM2
Esterase
FD2007
Esterase (283aa)
Esterase
Esterase (234aa)
Esterase
Esterase (374aa)
Esterase
Esterase (370aa)
Esterase
β-1,4-xylanase (1011aa)
GH10
CBM13
CBM2
Esterase
P6B7
Esterase (255aa)
Esterase
Esterase (275aa)
Esterase
Esterase (356aa)
Esterase
Esterase (269aa)
Esterase
Esterase (266aa)
Esterase
AE2015
Esterase (220aa)
Esterase
Esterase (267aa)
Esterase
Esterase (341aa)
Esterase
AE3009
Esterase (270aa)
Esterase
Esterase (285aa)
Esterase
Esterase (207aa)
Esterase
Esterase (359aa)
Esterase
Esterase (376aa)
Esterase
β-1,4-xylanase (815aa)
GH10
CBM13
CBM2
Esterase
FCS006
Esterase (275aa)
Esterase
AE2032
Esterase (280aa)
Esterase
Esterase (370aa)
Esterase
FCS014
Esterase (106aa)
Esterase (118aa)
Esterase (365aa)
Esterase
Esterase (272aa)
Esterase
Esterase (268aa)
Esterase
Esterase (270aa)
Esterase
Esterase (1700aa)
CE10
Esterase
LB2008
Esterase (1017aa)
GH43
CBM6
CBM13
Esterase
Esterase (364aa)
Esterase
Esterase (267aa)
Esterase
Esterase (266aa)
Esterase
MC2021
β-1,4-xylanase (993aa)
GH10
CBM13
CBM2
Esterase
Esterase (270aa)
Esterase
Esterase (370aa)
Esterase
Esterase (574aa)
GH114
Esterase
1,4-β-N-acetylmuramidase (561aa)
GH25
Esterase
NC2007
Esterase (275aa)
Esterase
Esterase (208aa)
Esterase
VCB2001
Esterase (275aa)
Esterase
VCB2006
β-1,4-xylanase (1003aa)
GH10
CBM13
CBM2
Esterase
Esterase (219aa)
Esterase
Esterase (314aa)
Esterase
Esterase (360aa)
Esterase
Esterase (283aa)
Esterase
XBB1001
β-1,4-xylanase (1022aa)
GH10
CBM13
CBM2
Esterase
Esterase (209aa)
Esterase
Esterase (350aa)
Esterase
Esterase (337aa)
Esterase
Esterase (286aa)
Esterase
MB2003
Esterase (351aa)
Esterase
XPD2006
Esterase (275aa)
Esterase

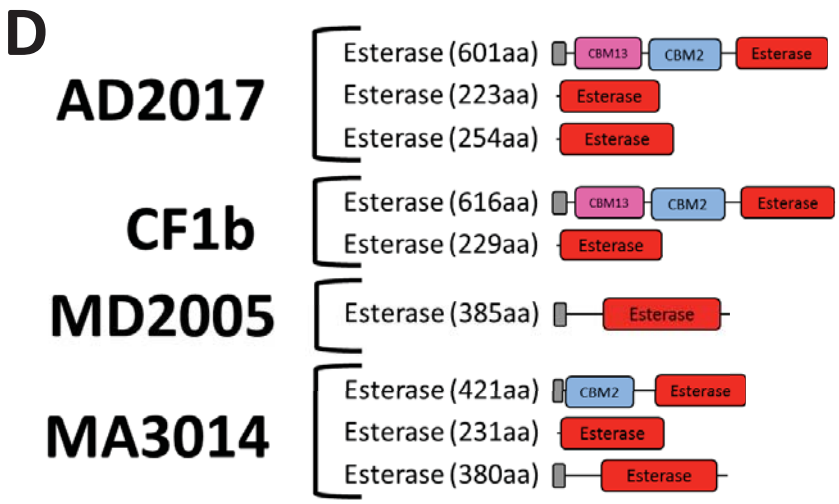
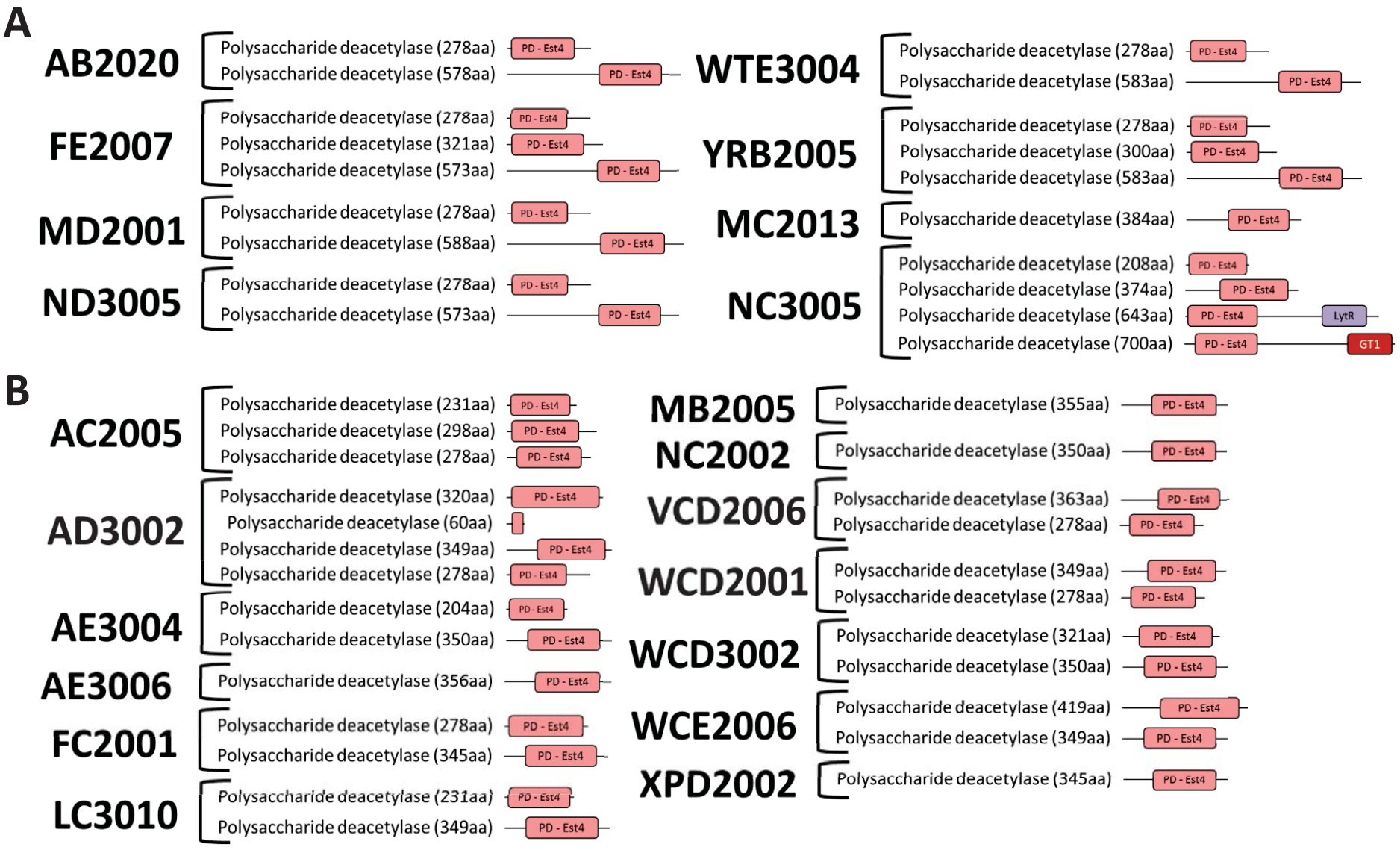
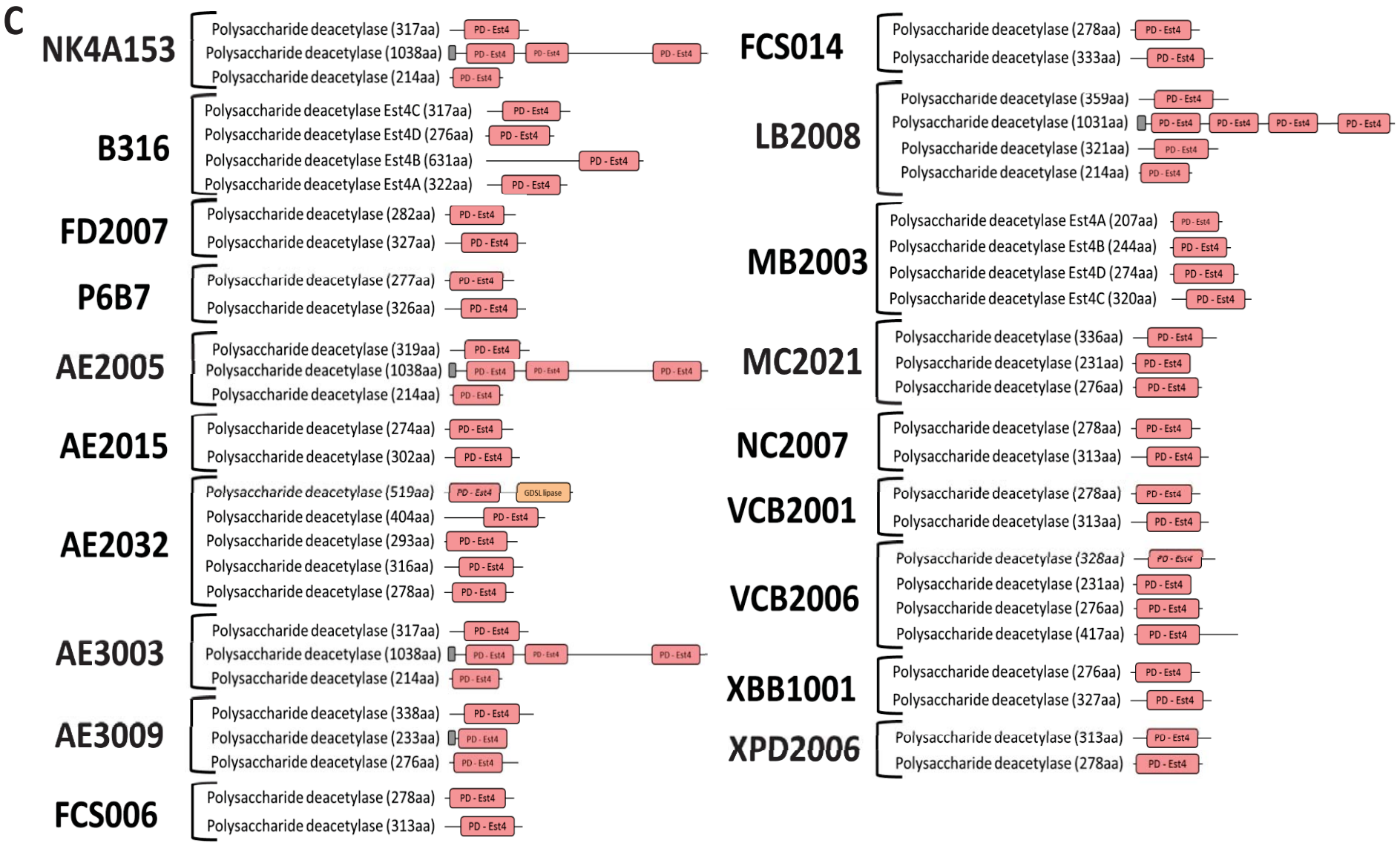


Figure 4.19. Predicted Pfam domains of CE1 acetyl-xylan and ferulic acid esterases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: Esterase, feruloyl esterase (CE1) (Pfam00756); CE10, acetyl xylan esterase - abhydrolase (Pfam07859); CBM2, carbohydrate-binding module 2 (Pfam00553); CBM6, carbohydrate-binding module 6 (Pfam03422); CBM13, carbohydrate-binding module 13 (Pfam14200); GH10, glycoside hydrolases family 10 (Pfam00331); GH25, glycoside hydrolases family 25 (Pfam01183); GH43, glycoside hydrolases family 43 (Pfam04616); GH114, glycoside hydrolases family 114 (Pfam03537). ▯, signal peptide sequences.


A
AB2020
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (578aa)
PD - Est4
FE2007
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (321aa)
PD - Est4
Polysaccharide deacetylase (573aa)
PD - Est4
MD2001
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (588aa)
PD - Est4
ND3005
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (573aa)
PD - Est4
WTE3004
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (583aa)
PD - Est4
YRB2005
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (300aa)
PD - Est4
Polysaccharide deacetylase (583aa)
PD - Est4
MC2013
Polysaccharide deacetylase (384aa)
PD - Est4
NC3005
Polysaccharide deacetylase (208aa)
PD - Est4
Polysaccharide deacetylase (374aa)
PD - Est4
Polysaccharide deacetylase (643aa)
PD - Est4
LytR
Polysaccharide deacetylase (700aa)
PD - Est4
GT1
B
AC2005
Polysaccharide deacetylase (231aa)
PD - Est4
Polysaccharide deacetylase (298aa)
PD - Est4
Polysaccharide deacetylase (278aa)
PD - Est4
AD3002
Polysaccharide deacetylase (320aa)
PD - Est4
Polysaccharide deacetylase (60aa)
Polysaccharide deacetylase (349aa)
PD - Est4
Polysaccharide deacetylase (278aa)
PD - Est4
AE3004
Polysaccharide deacetylase (204aa)
PD - Est4
Polysaccharide deacetylase (350aa)
PD - Est4
AE3006
Polysaccharide deacetylase (356aa)
PD - Est4
FC2001
Polysaccharide deacetylase (278aa)
PD - Est4
Polysaccharide deacetylase (345aa)
PD - Est4
LC3010
Polysaccharide deacetylase (231aa)
PD - Est4
Polysaccharide deacetylase (349aa)
PD - Est4
MB2005
Polysaccharide deacetylase (355aa)
PD - Est4
NC2002
Polysaccharide deacetylase (350aa)
PD - Est4
VCD2006
Polysaccharide deacetylase (363aa)
PD - Est4
Polysaccharide deacetylase (278aa)
PD - Est4
WCD2001
Polysaccharide deacetylase (349aa)
PD - Est4
Polysaccharide deacetylase (278aa)
PD - Est4
WCD3002
Polysaccharide deacetylase (321aa)
PD - Est4
Polysaccharide deacetylase (350aa)
PD - Est4
WCE2006
Polysaccharide deacetylase (419aa)
PD - Est4
Polysaccharide deacetylase (349aa)
PD - Est4
XPD2002
Polysaccharide deacetylase (345aa)
PD - Est4


C
NK4A153
Polysaccharide deacetylase (317aa)
Polysaccharide deacetylase (1038aa)
Polysaccharide deacetylase (214aa)
PD - Est4
B316
Polysaccharide deacetylase Est4C (317aa)
Polysaccharide deacetylase Est4D (276aa)
Polysaccharide deacetylase Est4B (631aa)
Polysaccharide deacetylase Est4A (322aa)
FD2007
Polysaccharide deacetylase (282aa)
Polysaccharide deacetylase (327aa)
P6B7
Polysaccharide deacetylase (277aa)
Polysaccharide deacetylase (326aa)
AE2005
Polysaccharide deacetylase (319aa)
Polysaccharide deacetylase (1038aa)
Polysaccharide deacetylase (214aa)
AE2015
Polysaccharide deacetylase (274aa)
Polysaccharide deacetylase (302aa)
AE2032
Polysaccharide deacetylase (519aa)
GDSL lipase
Polysaccharide deacetylase (404aa)
Polysaccharide deacetylase (293aa)
Polysaccharide deacetylase (316aa)
Polysaccharide deacetylase (278aa)
AE3003
Polysaccharide deacetylase (317aa)
Polysaccharide deacetylase (1038aa)
Polysaccharide deacetylase (214aa)
AE3009
Polysaccharide deacetylase (338aa)
Polysaccharide deacetylase (233aa)
Polysaccharide deacetylase (276aa)
FCS006
Polysaccharide deacetylase (278aa)
Polysaccharide deacetylase (313aa)
FCS014
Polysaccharide deacetylase (278aa)
Polysaccharide deacetylase (333aa)
LB2008
Polysaccharide deacetylase (359aa)
Polysaccharide deacetylase (1031aa)
Polysaccharide deacetylase (321aa)
Polysaccharide deacetylase (214aa)
MB2003
Polysaccharide deacetylase Est4A (207aa)
Polysaccharide deacetylase Est4B (244aa)
Polysaccharide deacetylase Est4D (274aa)
Polysaccharide deacetylase Est4C (320aa)
MC2021
Polysaccharide deacetylase (336aa)
Polysaccharide deacetylase (231aa)
Polysaccharide deacetylase (276aa)
NC2007
Polysaccharide deacetylase (278aa)
Polysaccharide deacetylase (313aa)
VCB2001
Polysaccharide deacetylase (278aa)
Polysaccharide deacetylase (313aa)
VCB2006
Polysaccharide deacetylase (328aa)
Polysaccharide deacetylase (231aa)
Polysaccharide deacetylase (276aa)
Polysaccharide deacetylase (417aa)
XBB1001
Polysaccharide deacetylase (276aa)
Polysaccharide deacetylase (327aa)
XPD2006
Polysaccharide deacetylase (313aa)
Polysaccharide deacetylase (278aa)

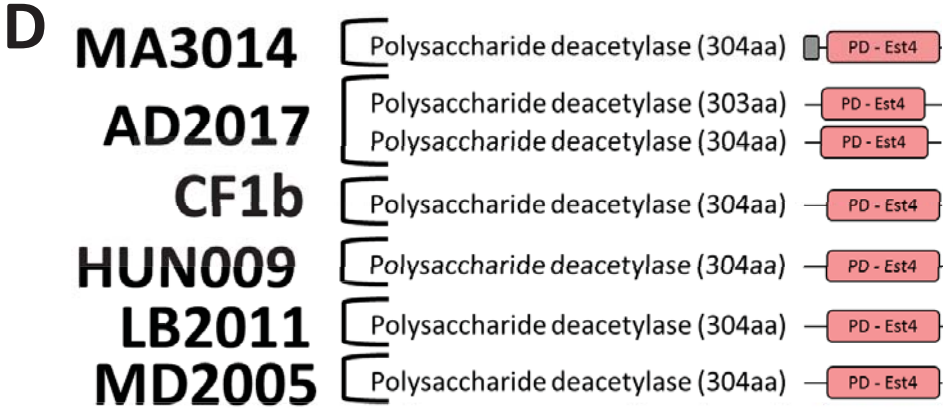


Figure 4.20. Predicted Pfam domains of CE4 polysaccharide deacetylases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: PD Est4, polysaccharide deacetylase (Pfam01522); LytR, cell envelope-related transcriptional attenuator domain (Pfam03816); GT1, glycosyl transferase family 1 (Pfam00534); GDSL lipase, GDSL-like Lipase/Acylhydrolase family (Pfam13472). ▯, signal peptide sequences.

4.6.5 Glycosyl Transferases (GTs)

GTs catalyse the formation of glycosidic linkages for the biosynthesis of mono-, di-, oligo- or poly-saccharides (Coutinho *et al.*, 2003; Lairson *et al.*, 2008). The most abundant and heterogeneous GTs among all *Butyrivibrio* and *Pseudobutyrivibrio* genomes were GT2 and GT4 (Figure 4.21). Our results showed that *Butyrivibrio* and *Pseudobutyrivibrio* encode 36 of the 98 GT families (Figure 4.8A). Some of the other relatively abundant GTs detected in *Butyrivibrio* and *Pseudobutyrivibrio* genomes were GTs 28, 35, and 51, associated with galactosyltransferase (Pfam04101), glycogen or starch phosphorylase (Pfam00343) and murein polymerase (Pfam00912) functions, respectively. Analysis of the Pfam domains of the major GTs are shown in Figure 4.22 and indicate that all *Butyrivibrio* and *Pseudobutyrivibrio* encode two GT35 (Pfam00343) glycogen or starch phosphorylases, that do not contain signal peptides. The GTs that were present in all *Butyrivibrio* but were completely absent in *Pseudobutyrivibrio* were GTs 11, 26 and 17, putatively associated with fucosyltransferase (Pfam01531), mannos-aminuronyltransferase (Pfam03808) and acetylglucosaminyltransferase (Pfam04724) functions, respectively. The rare GTs only detected in a few *Butyrivibrio* and *Pseudobutyrivibrio* genomes were GTs 9, 87, 19, 76, 46, 7, 56, 45, 94, 92 and 73. These GTs have reported functions as lipid di- and poly-saccharide synthases, manno-syltransferases, lactose synthases, acetylglucosaminyltransferases and galactosyl-transferases. Strains NC3005, LB2008, AC2005 and *P. ruminis* CF1b were the only ones to possess a GT90 (glucuronoxylomannan/galactoxylomannan β-1,2-xylosyltransferase), GT60 (polypeptide α-*N*-acetylglucosaminyltransferase), GT22 (α-1,2-mannosyl-transferase) and GT30 (α-3-deoxy-D-manno-octulosonic-acid (KDO) transferase), respectively.

Of all the genomes examined, *Butyrivibrio* sp. 2 FC2001 and *Butyrivibrio* sp. 3 AE2032 encoded the most GTs (Figure 4.22), suggesting that these bacteria are able to catalyse formation of a wide range of glycosidic bonds. Both strains encode a ~400 aa enzyme containing a GT1 and a GT4 domain (Pfam00534) containing a signal peptide. Strain FC2001 has the only gene found in this analysis that encodes for a GT2 (Pfam00535) and a GT4 domain (Pfam13692, 716 aa), and two genes that encode for a GT4 (Pfam13579) and a GT1 domain (Pfam00534) both containing signal peptides. The only GTs consisting of a GT1 (Pfam13524) and DUF3380 (Pfam12996) and signal peptides were found in strain AE2032.

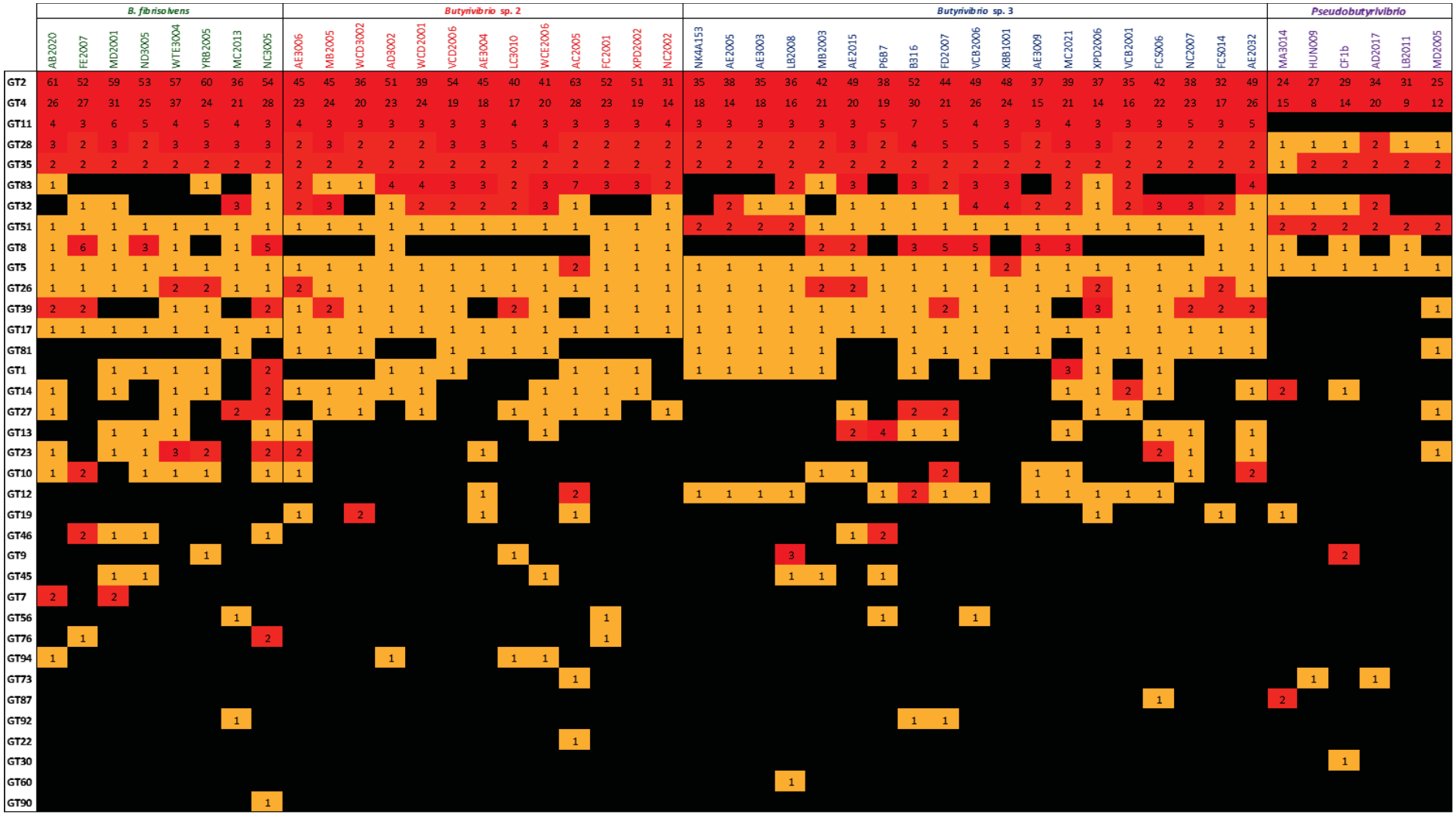
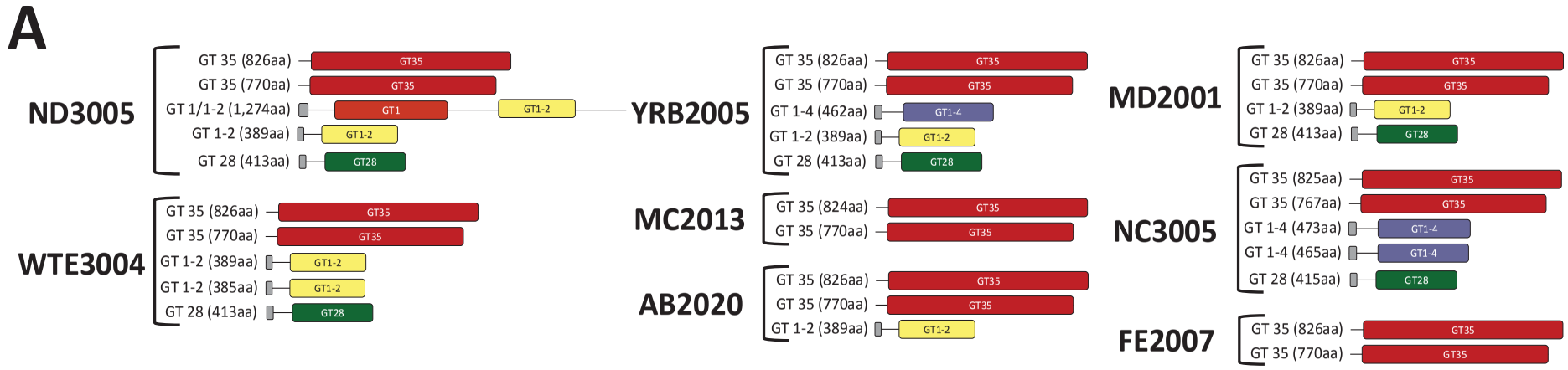
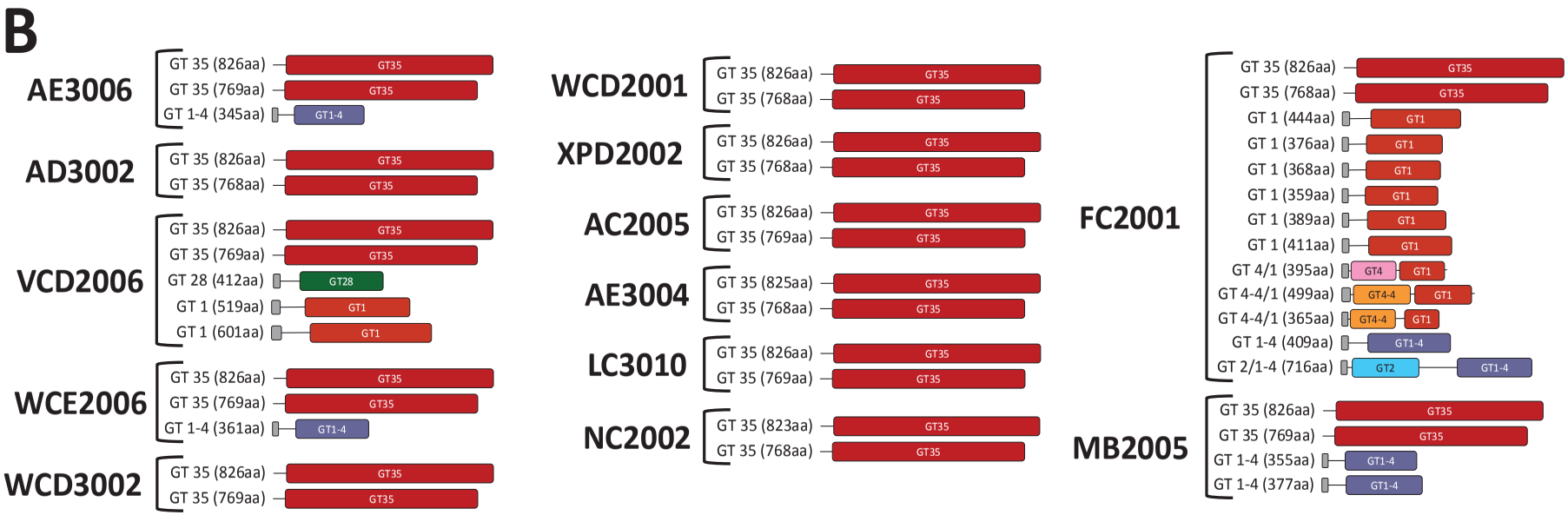


Figure 4.21. Heatmap showing the abundances of GT families in *Butyrivibrio* and *Pseudobutyrivibrio* genomes. The degree of redness indicates relative abundance and black indicates an absence of that GT.


A
ND3005
GT 35 (826aa)
GT35
GT 35 (770aa)
GT35
GT 1/1-2 (1,274aa)
GT1
GT1-2
GT 1-2 (389aa)
GT1-2
GT 28 (413aa)
GT28
WTE3004
GT 35 (826aa)
GT35
GT 35 (770aa)
GT35
GT 1-2 (389aa)
GT1-2
GT 1-2 (385aa)
GT1-2
GT 28 (413aa)
GT28
YRB2005
GT 35 (826aa)
GT35
GT 35 (770aa)
GT35
GT 1-4 (462aa)
GT1-4
GT 1-2 (389aa)
GT1-2
GT 28 (413aa)
GT28
MC2013
GT 35 (824aa)
GT35
GT 35 (770aa)
GT35
AB2020
GT 35 (826aa)
GT35
GT 35 (770aa)
GT35
GT 1-2 (389aa)
GT1-2
MD2001
GT 35 (826aa)
GT35
GT 35 (770aa)
GT35
GT 1-2 (389aa)
GT1-2
GT 28 (413aa)
GT28
NC3005
GT 35 (825aa)
GT35
GT 35 (767aa)
GT35
GT 1-4 (473aa)
GT1-4
GT 1-4 (465aa)
GT1-4
GT 28 (415aa)
GT28
FE2007
GT 35 (826aa)
GT35
GT 35 (770aa)
GT35


B
AE3006
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
GT 1-4 (345aa)
GT1-4
AD3002
GT 35 (826aa)
GT35
GT 35 (768aa)
GT35
VCD2006
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
GT 28 (412aa)
GT28
GT 1 (519aa)
GT1
GT 1 (601aa)
GT1
WCE2006
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
GT 1-4 (361aa)
GT1-4
WCD3002
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
WCD2001
GT 35 (826aa)
GT35
GT 35 (768aa)
GT35
XPD2002
GT 35 (826aa)
GT35
GT 35 (768aa)
GT35
AC2005
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
AE3004
GT 35 (825aa)
GT35
GT 35 (768aa)
GT35
LC3010
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
NC2002
GT 35 (823aa)
GT35
GT 35 (768aa)
GT35
FC2001
GT 35 (826aa)
GT35
GT 35 (768aa)
GT35
GT 1 (444aa)
GT1
GT 1 (376aa)
GT1
GT 1 (368aa)
GT1
GT 1 (359aa)
GT1
GT 1 (389aa)
GT1
GT 1 (411aa)
GT1
GT 4/1 (395aa)
GT4
GT1
GT 4-4/1 (499aa)
GT4-4
GT1
GT 4-4/1 (365aa)
GT4-4
GT1
GT 1-4 (409aa)
GT1-4
GT 2/1-4 (716aa)
GT2
GT1-4
MB2005
GT 35 (826aa)
GT35
GT 35 (769aa)
GT35
GT 1-4 (355aa)
GT1-4
GT 1-4 (377aa)
GT1-4

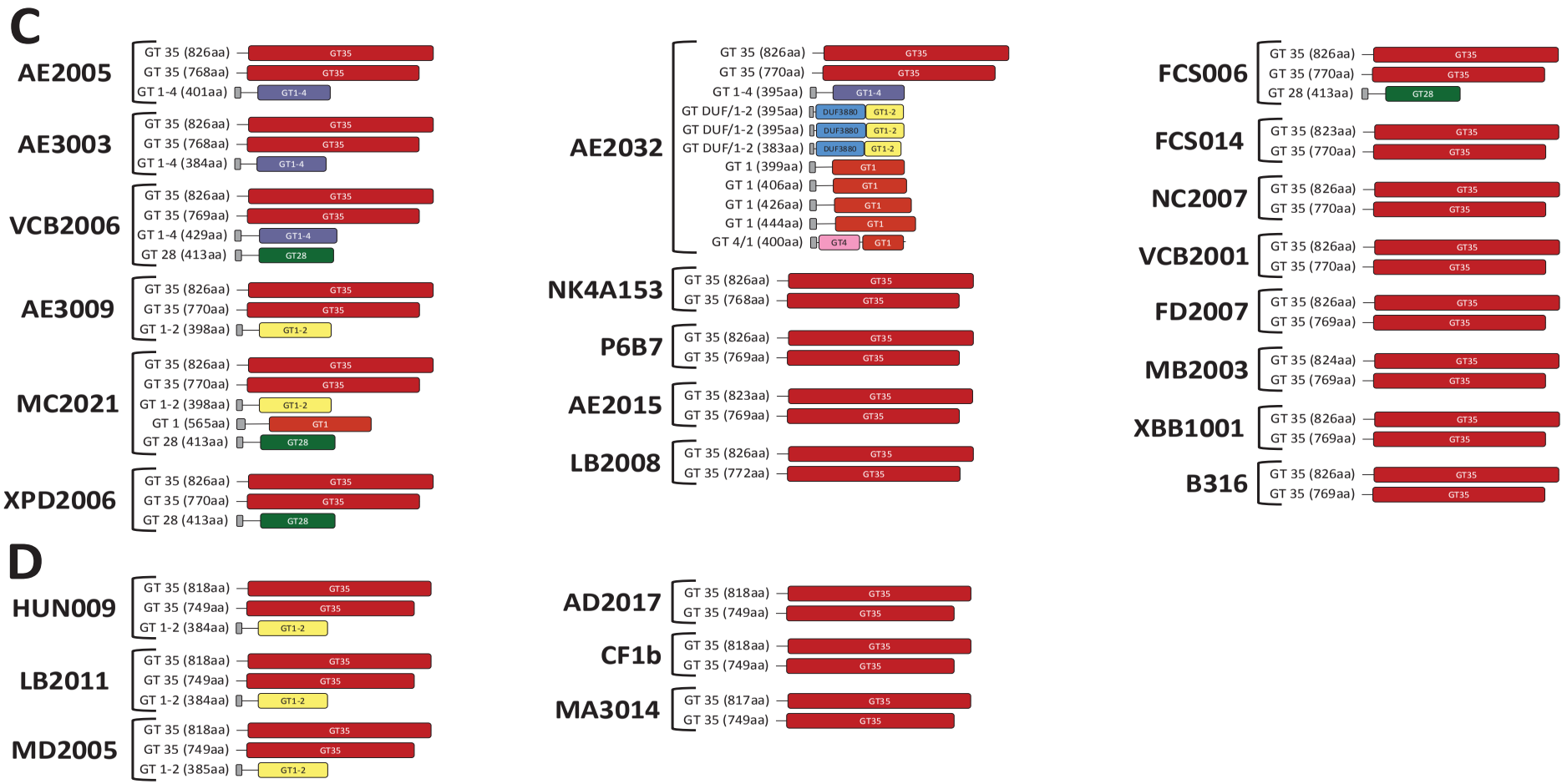


Figure 4.22. Predicted Pfam domains of GTs. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: GT1, glycosyl transferase family 1 (Pfam00534); GT2, glycosyl transferase family 2 (Pfam00535); GT1-2, glycosyl transferase family 1 (Pfam13524); GT1-4, glycosyl transferase family 1-4 (Pfam13692); GT4, glycosyl transferase family 4 (Pfam00534); GT4-4, glycosyl transferase 4-like domain (Pfam13579); GT28, glycosyl transferase family 28 N-terminal domain (Pfam04101); GT35, glycosyl transferase family 35 (Pfam00343); DUF3380, domain of unknown function 3380 (Pfam12996). ▯, signal peptide sequences.

4.7 Genotypic variation associated with selected phenotypes

4.7.1 Cell motility and flagellum-encoding loci

The availability of draft *Butyrivibrio* genome sequences allowed for an analysis of their genes in the main flagella biosynthesis locus in relation to their observed motility (Chapter 3, Figures 3.8-3.11). The genes located in the main flagellum biosynthesis operon have been identified for all *B. fibrisolvens*, *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3 strains. Two reference organisms, the non-motile *Streptococcus bovis* 2B, and the motile *P. xylanivorans* MA3014 were used as controls (Figure 4.23). Previous motility assays (Palevich, 2011), showed that two non-motile *Butyrivibrio* strains, *B. proteoclasticus* B316 and *B. hungatei* MB2003, contained genes apparently inserted within the flagellum operon, between genes *flgE* and *motA*. The genes *est*1B and *flb*D were inserted in the B316 locus, whereas a gene encoding an ion channel and *flb*D were inserted in the MB2003 locus. The *flb*D genes are the same size in both MB2003 and B316 and the *est*1B and gene encoding a putative ion channel were approximately the same size in their respective genomes. In contrast, no extra genes were found between genes *flgE* and *motA* in the genome of the motile control *P. xylanivorans* MA3014. Based on these findings, a hypothesis was formulated that the additional genes inserted between *flgE* and *motA* disrupt the flagellar function (Palevich, 2011).

The variable regions of the flagellum biosynthesis operon were compared in 29 *Butyrivibrio* strains and correlated with their motility in order to determine the relationship between the genotype and phenotype. The motility assays (Figures 4.24-4.26), showed that among the 29 strains, only the *B. fibrisolvens* MC2013 and *Butyrivibrio* sp. 2 FC2001 strains were motile, while all remaining strains were non-motile. The flagellum locus of MC2013 had no insertions, but FC2001 had an insert between the *flgE* and *motA* genes, yet it was motile. Conversely, a number of strains that were not motile (specifically five out of six *B. proteoclasticus* strains) did not contain an insert in the flagellum locus. Sequence analysis and annotation of the flagellum operon region encompassing *flgE* and *motA*, identified variation among the three groupings of *Butyrivibrio*. The *flb*D gene, was present in most of the *Butyrivibrio* sp. 2 and 3 strains, but absent from all *B. fibrisolvens* strains. The presence and nature of the second gene located between the *flg*E and *flb*D genes, varies greatly amongst the *Butyrivibrio* sp. 2 and 3 strains.

These genes were identified by BLASTn as mainly hypothetical proteins or as esterases that are homologous to the feruloyl esterase *est1B* (bpr_I1368) of B316. One gene, encoding protein PLDc2, was only found in strain AE2015. The *motA* and *motB* genes were similar in size and present in all analysed genomes. Unique insertions were found downstream of their respective *flgE* genes in two strains, AE3004 (Figure 4.25) and AE3009 (Figure 4.26), containing three and at least six genes, respectively. These findings refuted the original hypothesis that insertions between *flgE* and *motA* genes inactivate an otherwise active flagellum in *Butyrivibrio* strains. The cause for motility in this cluster therefore likely depends on other, unidentified mutations or genes that regulate expression of this flagellum operon or other flagellum-associated genes.

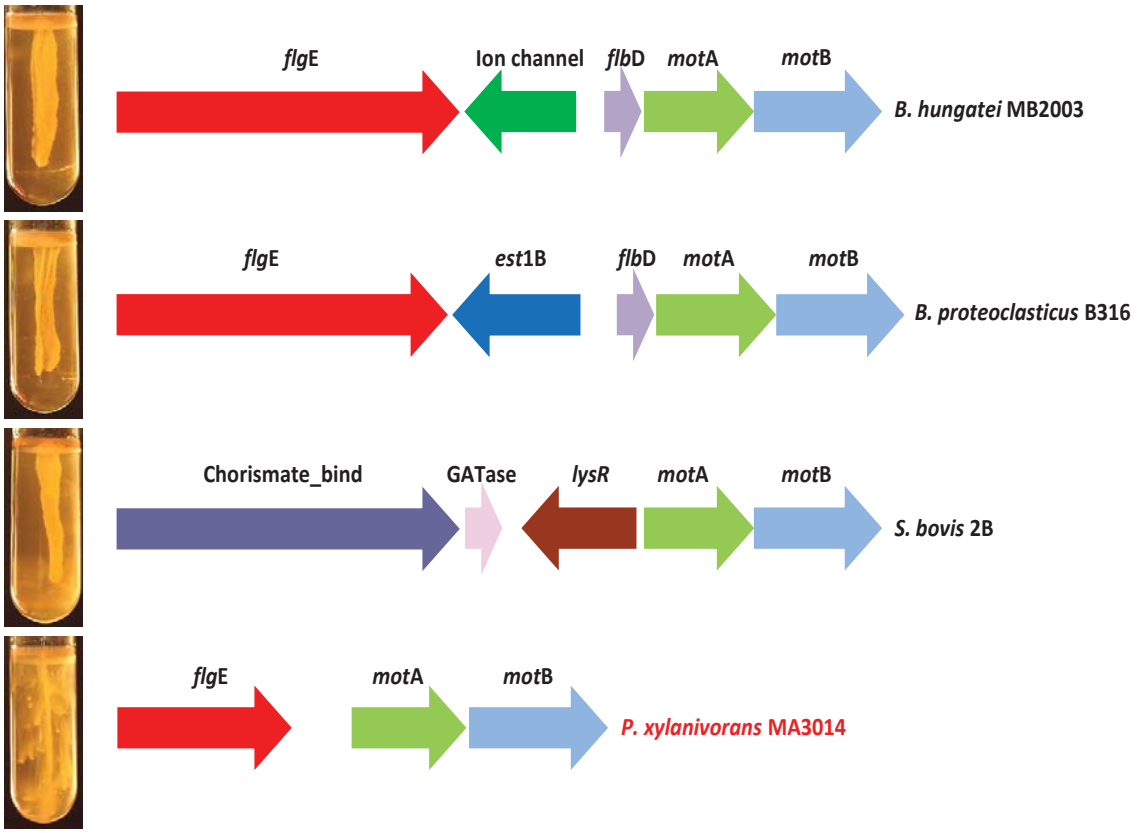


Figure 4.23. Motility assays and flagellar biosynthesis operons of control strains. Non-motile control strains are *B. hungatei* MB2003, *B. proteoclasticus* B316 and *S. bovis* 2B, while *P. xylanivorans* MA3014 was the motile control strain. Arrow lengths indicate the relative size of each gene.

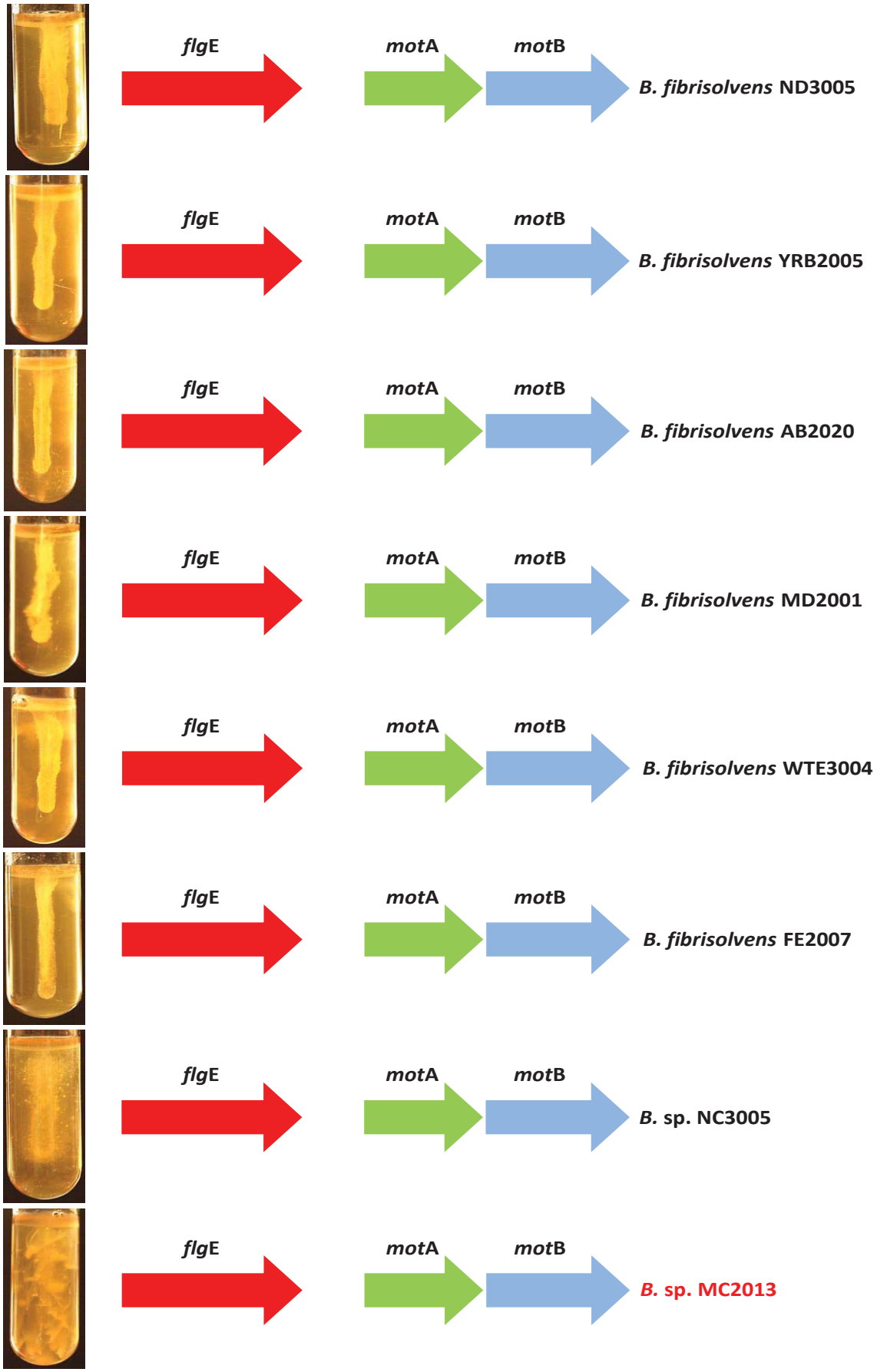
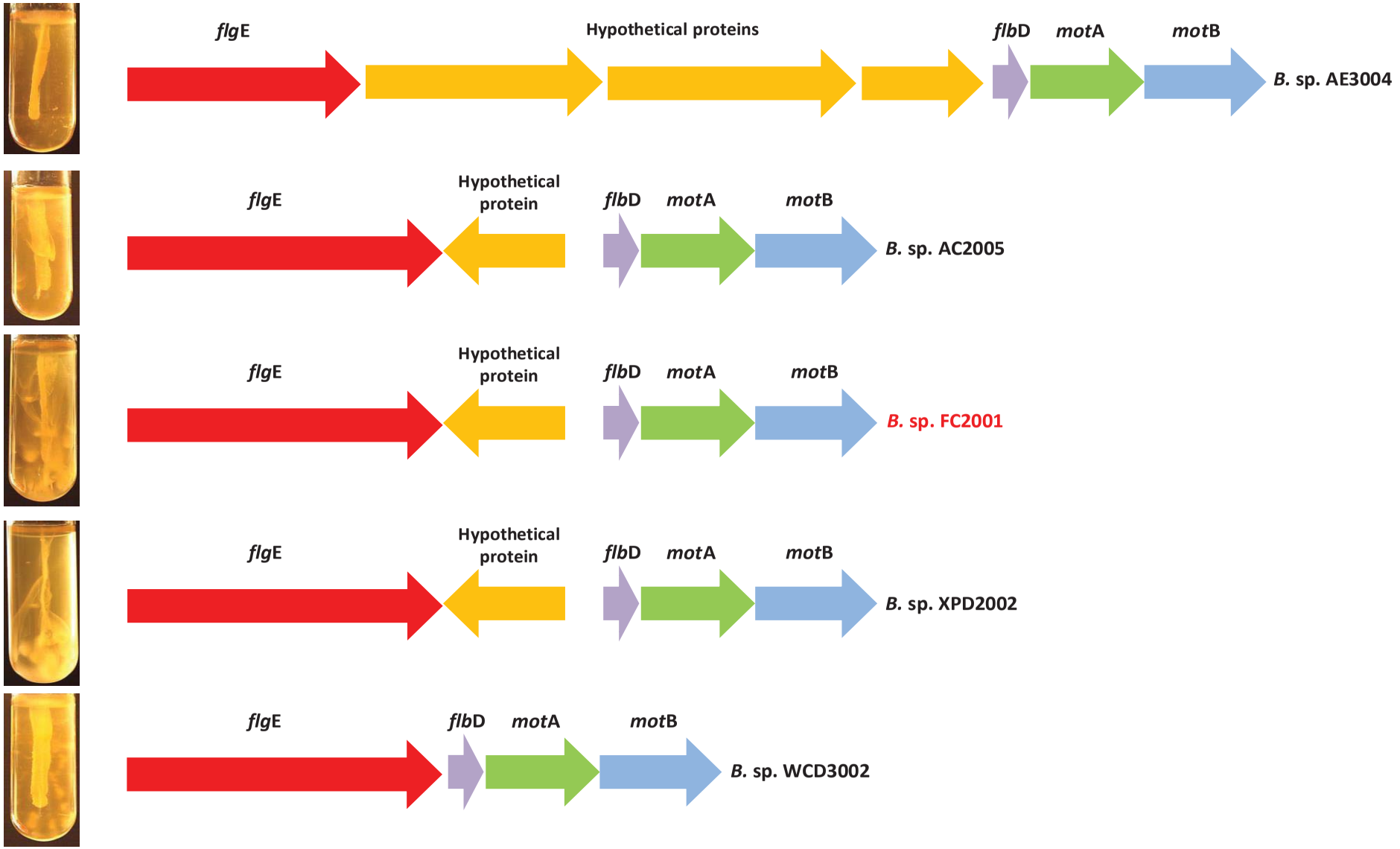


Figure 4.24. Motility assays and flagellar biosynthesis operons of *B. fibrisolvens* strains. Arrow lengths indicate the relative size of each gene.


flgE
Hypothetical proteins
flbD
motA
motB
B. sp. AE3004
flgE
Hypothetical protein
flbD
motA
motB
B. sp. AC2005
flgE
Hypothetical protein
flbD
motA
motB
B. sp. FC2001
flgE
Hypothetical protein
flbD
motA
motB
B. sp. XPD2002
flgE
flbD
motA
motB
B. sp. WCD3002

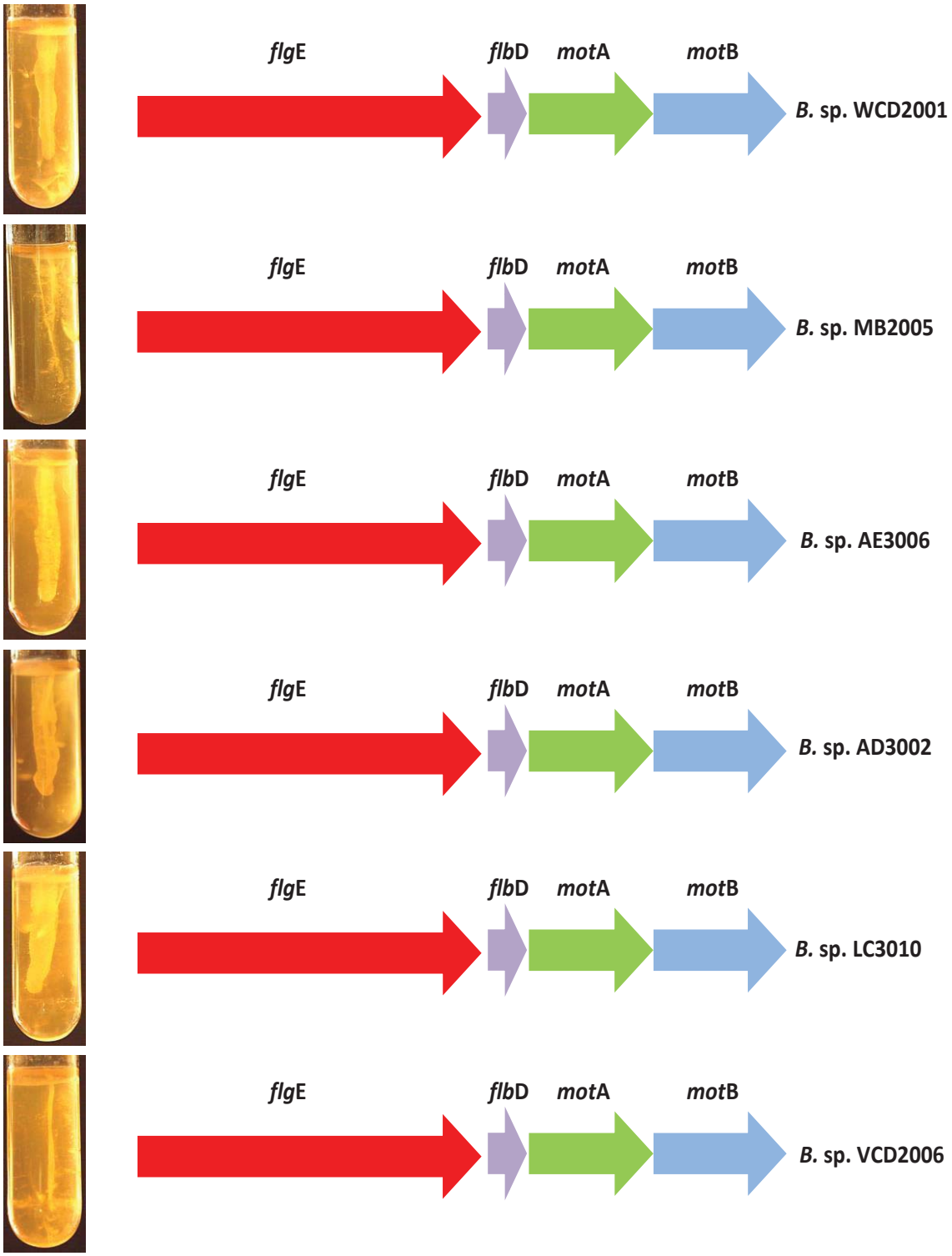
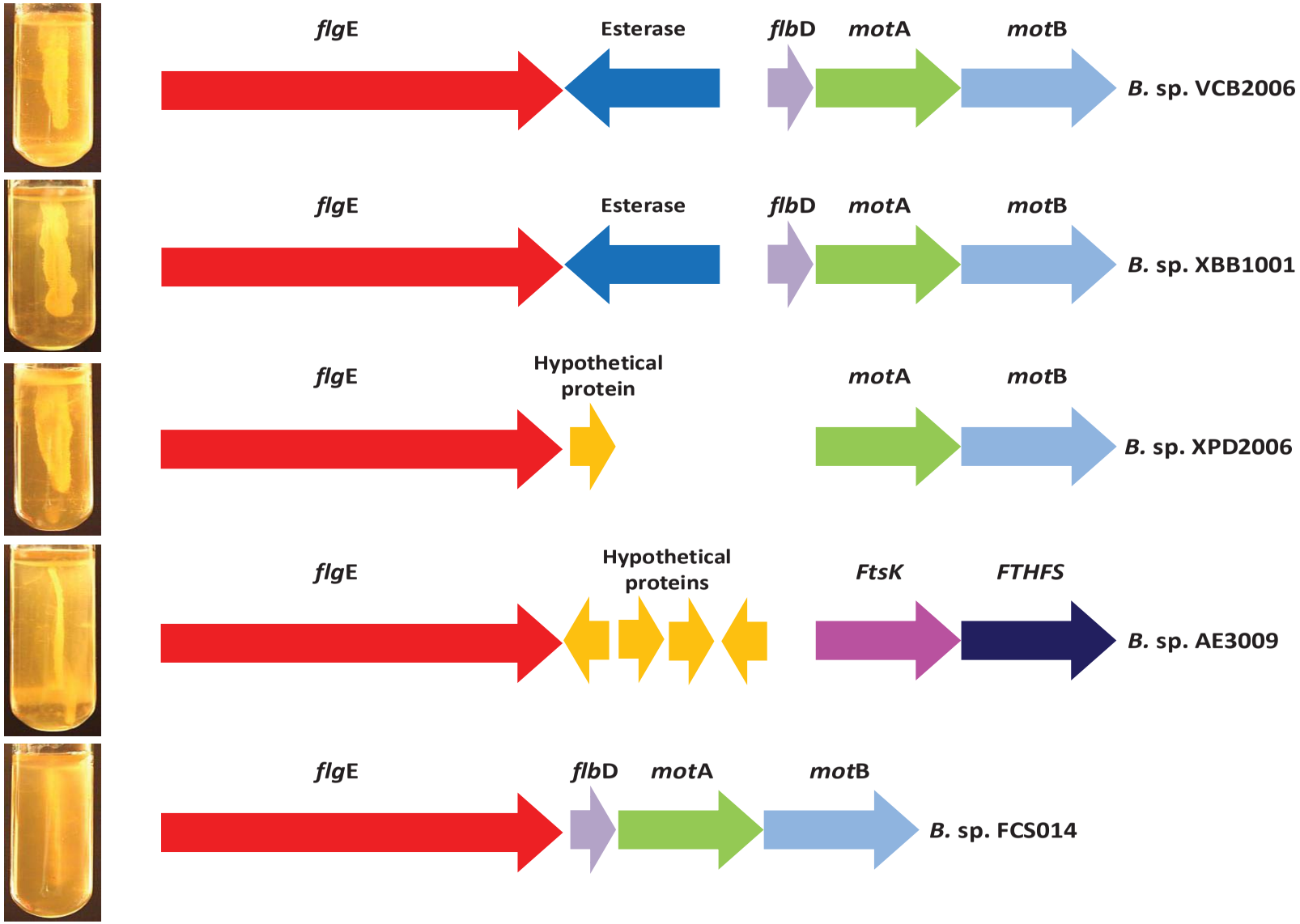


Figure 4.25. Motility assays and flagellar biosynthesis operons of *Butyrivibrio* sp. 2 strains. Arrow lengths indicate the relative size of each gene.


flgE
Esterase
flbD
motA
motB
B. sp. VCB2006
flgE
Esterase
flbD
motA
motB
B. sp. XBB1001
flgE
Hypothetical protein
motA
motB
B. sp. XPD2006
flgE
Hypothetical proteins
FtsK
FTHFS
B. sp. AE3009
flgE
flbD
motA
motB
B. sp. FCS014

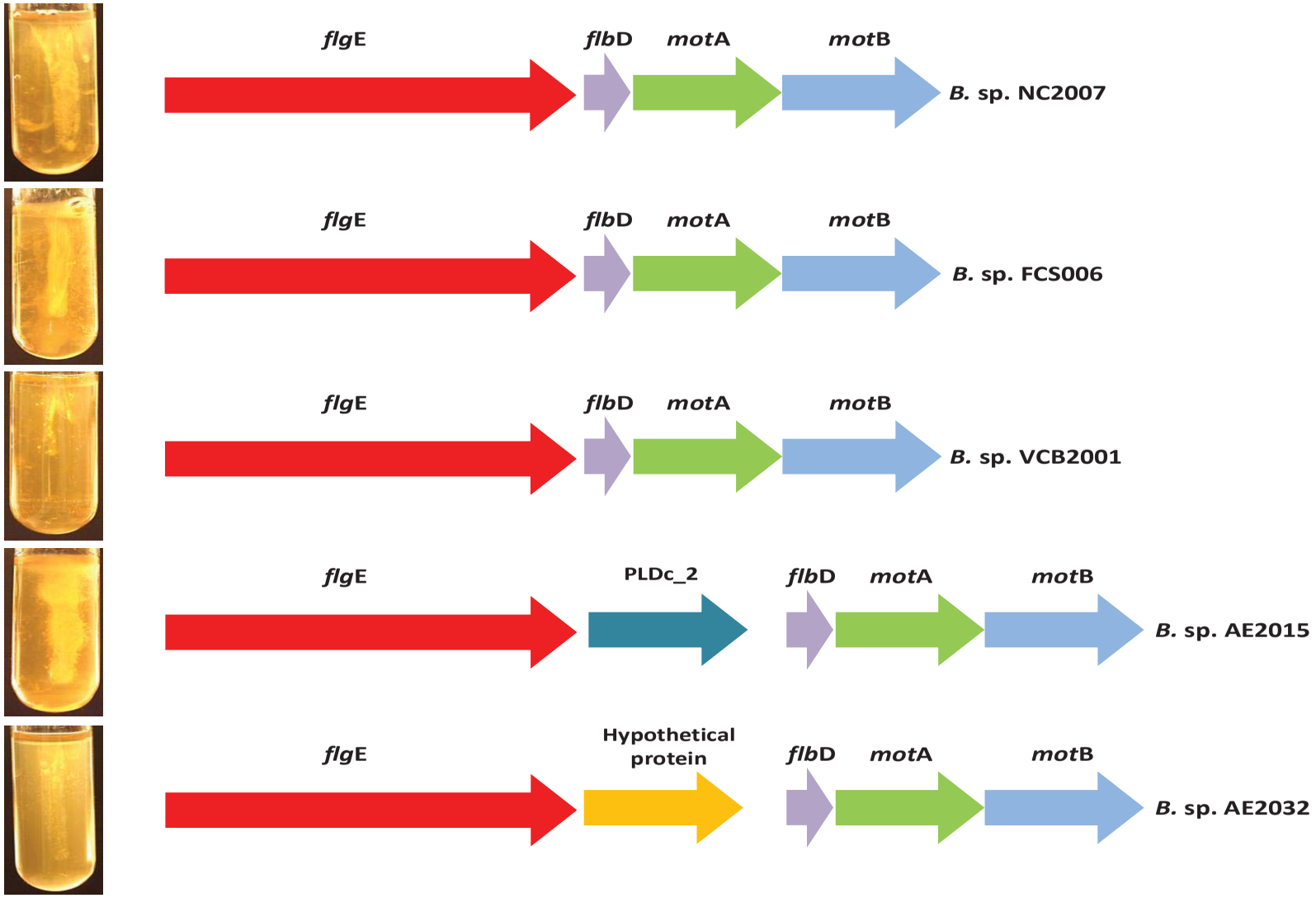


Figure 4.26. Motility assays and flagellar biosynthesis operons of *Butyrivibrio* sp. 3 strains. Arrow lengths indicate the relative size of each gene.

4.7.2 Lactate production

Analysis of fermentation end products in Chapter 3 identified production of lactate (≥3 mM) in 11 *Butyrivibrio* strains. Genes encoding lactate dehydrogenase (*ldh*), have been identified and compared in the draft genome sequences of 29 *Butyrivibrio* strains (Figure 4.27). There are two main types of *ldh* genes encoded by the lactate-producing strains, containing very different Pfam domain combinations: i) a L-isomer specific L-*ldh* containing two Pfam domains named Ldh_1_N (NAD binding) and Ldh_1_C (α/β-C-terminal); ii) a D-isomer specific *ldhD* or 2-hydroxyacid dehydrogenase (2-Hacid-dh) composed of a 2-Hacid-dh catalytic domain and a 2-Hacid-dhC (NAD binding domain) domain. 2-Hacid-dh is annotated with dual functions, as an *ldh* and as a glyoxylate/hydroxypyruvate reductase (2-Hacid-dh superfamily). However, the *ldhD* containing 2-Hac-dh/C Pfam domains is not proven to be a D-isomer specific lactate dehydrogenase.

L-*ldh* plays a key role in the production of L-lactate from pyruvate. Genomes of all 18 *Butyrivibrio* strains that produced very little or no lactate, encode an *ldhD* containing 2-Hac-dh/C Pfam domains, however not all encode the L-*ldh* containing Ldh_1_N/C domains, that is required for production of L-lactate. Also all strains, except FCS006 and NC2007 (*Butyrivibrio* sp. 2) and AE3006, MB2005, WCD3002 and AE3004 (*Butyrivibrio* sp. 3), possess both types of *ldh*. Strains VCB2001, FCS006 and NC2007 (*Butyrivibrio* sp. 3) and all strains from *Butyrivibrio* sp. 2 except AD3002 and WCD2001, contain an L-*ldh* with the FMN-dh (flavin mononucleotide-dependent dehydrogenase) Pfam domain and their amino acid sequence varies greatly in length. Furthermore, only MD2001 strain from *B. fibrisolvens* encodes a FMN-dh Pfam domain-containing *ldhD*. The predicted function of the FMN-dh domain is to interact with the haem-binding domain (cytochrome b_2) to catalyse the conversion of lactate into pyruvate (Xia and Mathews, 1990), however the exact function of this domain requires further investigation.

The amino acid sequence-based Maximum Likelihood (ML) analysis of *ldhs* from draft *Butyrivibrio* genome sequences identified a clear difference in the two main types of *ldhs* and also the grouping of *Butyrivibrio* that are either lactate or non-lactate producers (Figure 4.28). All lactate producing strains encode both types of *ldh* genes, and most of the *B. fibrisolvens* contain two main *ldh* types. The L-*ldh* form two highly conserved clusters based on their amino acid sequences alignment, while the *ldhD* genes encoding 2-Hac-dh Pfam domains have a much larger variation in their amino acid sequences, and cluster into numerous sub-groups.

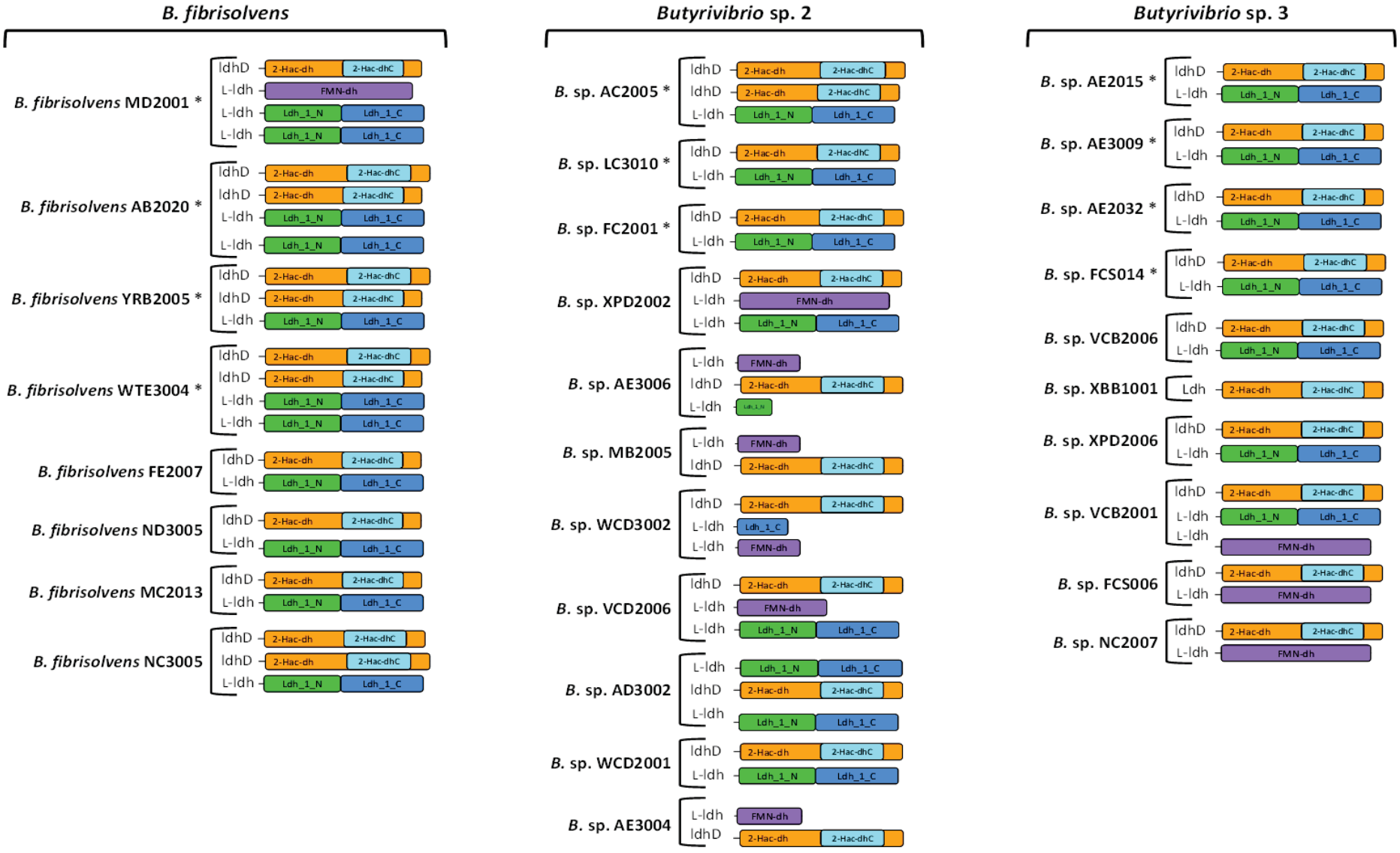

B. fibrisolvens
B. fibrisolvens MD2001 *
IdhD
2-Hac-dh
2-Hac-dhC
L-Idh
FMN-dh
Ldh_1_N
Ldh_1_C
B. fibrisolvens AB2020 *
B. fibrisolvens YRB2005 *
B. fibrisolvens WTE3004 *
B. fibrisolvens FE2007
B. fibrisolvens ND3005
B. fibrisolvens MC2013
B. fibrisolvens NC3005
Butyrivibrio sp. 2
B. sp. AC2005 *
B. sp. LC3010 *
B. sp. FC2001 *
B. sp. XPD2002
B. sp. AE3006
B. sp. MB2005
B. sp. WCD3002
B. sp. VCD2006
B. sp. AD3002
B. sp. WCD2001
B. sp. AE3004
Butyrivibrio sp. 3
B. sp. AE2015 *
B. sp. AE3009 *
B. sp. AE2032 *
B. sp. FCS014 *
B. sp. VCB2006
B. sp. XBB1001
Ldh
B. sp. XPD2006
B. sp. VCB2001
B. sp. FCS006
B. sp. NC2007

Figure 4.27. Comparison of *ldh* Pfam domains in *Butyrivibrio* genome sequences. ***** indicates lactate-producing (≥3 mM) strains. Abbreviations: *ldhD,* D-lactate dehydrogenase; L-*ldh*, L-lactate dehydrogenase; Ldh_1_N, lactate/malate dehydrogenase NAD binding domain (Pfam00056); Ldh_1_C, lactate/malate dehydrogenase α/β C-terminal domain (Pfam02866); 2-Hacid-dh, D-isomer specific 2-hydroxyacid dehydrogenase catalytic domain (Pfam00389); 2-Hacid-dhC, D-isomer specific 2-hydroxyacid dehydrogenase NAD binding domain (Pfam02828); FMN-dh, FMN-dependent dehydrogenase (Pfam01070).

The *B. fibrisolvens* MD2001 *ldh* is very different from other variants within the analysed group. This organism is the only example of a lactate-producing *Butyrivibrio* encoding a FMN-dependent α-hydroxy acid dehydrogenase. Contrary to findings amongst the lactate-producing *B. fibrisolvens* strains that encode multiple *ldh* enzymes, those *B. fibrisolvens* strains that do not produce large amounts of lactate only have single copies of the two main *ldh* genes. The *ldhD* genes containing 2-Hac-dh Pfam domains displayed large variation in their amino acid sequence, thus numerous sub-groups are identified when aligned using a ML cluster analysis.

Overall, the *in silico* analysis has identified a correlation between possession of the D-lactate dehydrogenase containing a D-isomer specific 2-hydroxyacid dehydrogenase domain (2-Hacid-dh) in all eleven lactate-producing *Butyrivibrio* strains. However, as the isomer of lactate produced was not determined, these associations are not definite and require further experimental validation.

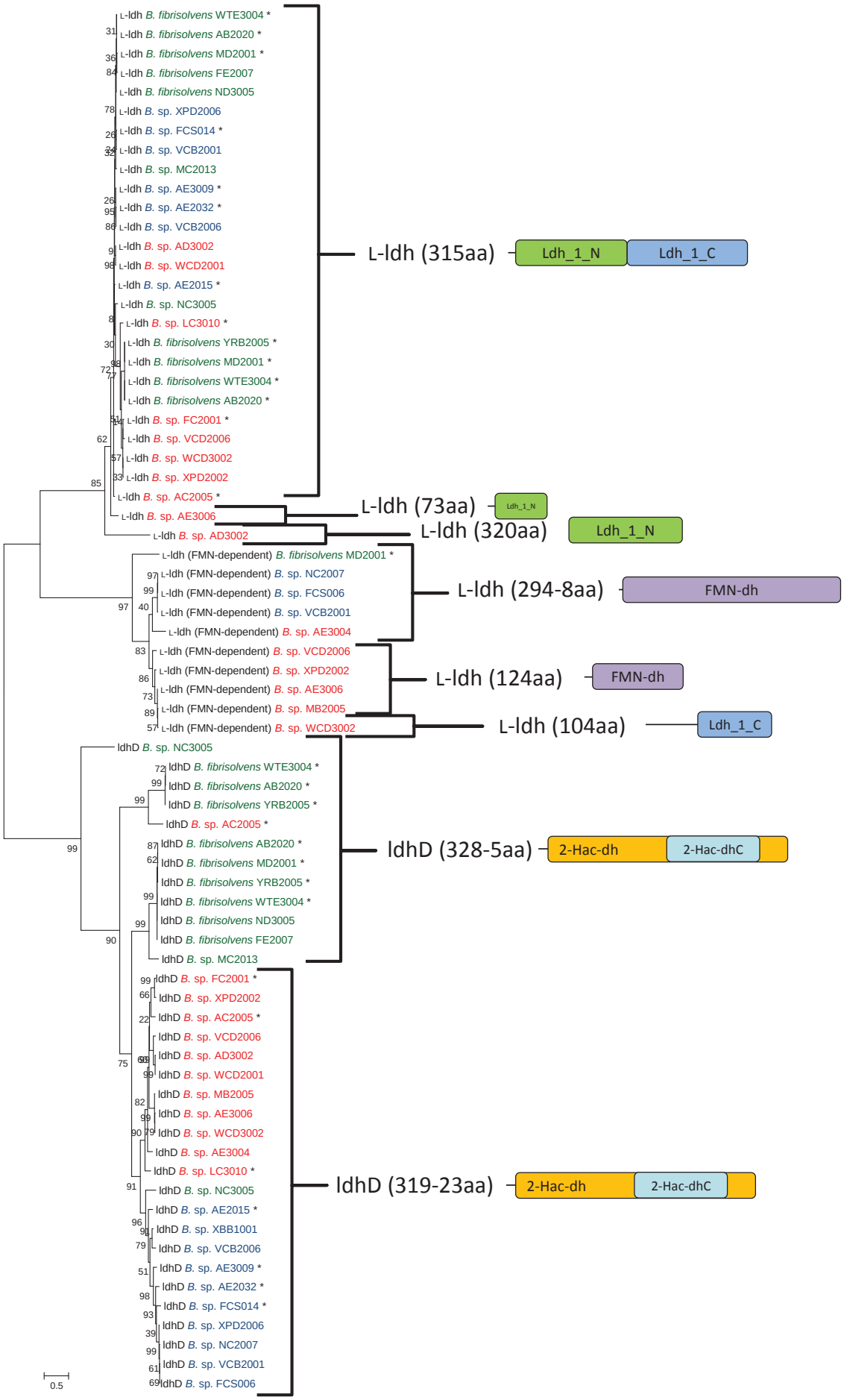

L-ldh B. fibrisolvens WTE3004 *
L-ldh B. fibrisolvens AB2020 *
L-ldh B. fibrisolvens MD2001 *
L-ldh B. fibrisolvens FE2007
L-ldh B. fibrisolvens ND3005
L-ldh B. sp. XPD2006
L-ldh B. sp. FCS014 *
L-ldh B. sp. VCB2001
L-ldh B. sp. MC2013
L-ldh B. sp. AE3009 *
L-ldh B. sp. AE2032 *
L-ldh B. sp. VCB2006
L-ldh B. sp. AD3002
L-ldh B. sp. WCD2001
L-ldh B. sp. AE2015 *
L-ldh B. sp. NC3005
L-ldh B. sp. LC3010 *
L-ldh B. fibrisolvens YRB2005 *
L-ldh B. fibrisolvens MD2001 *
L-ldh B. fibrisolvens WTE3004 *
L-ldh B. fibrisolvens AB2020 *
L-ldh B. sp. FC2001 *
L-ldh B. sp. VCD2006
L-ldh B. sp. WCD3002
L-ldh B. sp. XPD2002
L-ldh B. sp. AC2005 *
L-ldh B. sp. AE3006
L-ldh B. sp. AD3002
L-ldh (FMN-dependent) B. fibrisolvens MD2001 *
L-ldh (FMN-dependent) B. sp. NC2007
L-ldh (FMN-dependent) B. sp. FCS006
L-ldh (FMN-dependent) B. sp. VCB2001
L-ldh (FMN-dependent) B. sp. AE3004
L-ldh (FMN-dependent) B. sp. VCD2006
L-ldh (FMN-dependent) B. sp. XPD2002
L-ldh (FMN-dependent) B. sp. AE3006
L-ldh (FMN-dependent) B. sp. MB2005
L-ldh (FMN-dependent) B. sp. WCD3002
ldhD B. sp. NC3005
ldhD B. fibrisolvens WTE3004 *
ldhD B. fibrisolvens AB2020 *
ldhD B. fibrisolvens YRB2005 *
ldhD B. sp. AC2005 *
ldhD B. fibrisolvens AB2020 *
ldhD B. fibrisolvens MD2001 *
ldhD B. fibrisolvens YRB2005 *
ldhD B. fibrisolvens WTE3004 *
ldhD B. fibrisolvens ND3005
ldhD B. fibrisolvens FE2007
ldhD B. sp. MC2013
ldhD B. sp. FC2001 *
ldhD B. sp. XPD2002
ldhD B. sp. AC2005 *
ldhD B. sp. VCD2006
ldhD B. sp. AD3002
ldhD B. sp. WCD2001
ldhD B. sp. MB2005
ldhD B. sp. AE3006
ldhD B. sp. WCD3002
ldhD B. sp. AE3004
ldhD B. sp. LC3010 *
ldhD B. sp. NC3005
ldhD B. sp. AE2015 *
ldhD B. sp. XBB1001
ldhD B. sp. VCB2006
ldhD B. sp. AE3009 *
ldhD B. sp. AE2032 *
ldhD B. sp. FCS014 *
ldhD B. sp. XPD2006
ldhD B. sp. NC2007
ldhD B. sp. VCB2001
ldhD B. sp. FCS006
0.5
L-ldh (315aa)
Ldh_1_N
Ldh_1_C
L-ldh (73aa)
Ldh_1_N
L-ldh (320aa)
Ldh_1_N
L-ldh (294-8aa)
FMN-dh
L-ldh (124aa)
FMN-dh
L-ldh (104aa)
Ldh_1_C
ldhD (328-5aa)
2-Hac-dh
2-Hac-dhC
ldhD (319-23aa)
2-Hac-dh
2-Hac-dhC

Figure 4.28. Phylogenetic tree of *ldhs* encoded in *Butyrivibrio* genomes. The scale indicates the number of aa substitutions per site. The Maximum Likelihood analysis of lactate dehydrogenase proteins involved 72 amino acid sequences and there were a total of 357 positions in the final dataset. Strains are coloured to represent *B. fibrisolvens* in green, *Butyrivibrio* sp. 2 in red and *Butyrivibrio* sp. 3 in blue. * indicates lactate-producing (≥3 mM) strains. Abbreviations: *ldhD,* D-lactate dehydrogenase; L-*ldh*, L-lactate dehydrogenase; Ldh_1_N, lactate/ malate dehydrogenase NAD binding domain (Pfam00056); Ldh_1_C, lactate/malate de-hydrogenase α/β C-terminal domain (Pfam02866); 2-Hacid-dh, D-isomer specific 2-hydroxyacid dehydrogenase catalytic domain (Pfam00389); 2-Hacid-dhC, D-isomer specific 2-hydroxyacid dehydrogenase NAD binding domain (Pfam02828); FMN-dh, FMN-dependent dehydrogenase (Pfam01070).

4.7.3 Enolase genes

An unusual feature of the *B. proteoclasticus* B316 genome sequence previously reported is that it lacks an enolase gene (*eno*), which encodes an essential enzyme of glycolysis, catalysing conversion of 2-phospho-D-glycerate to phsophoenol-pyruvate (EC 4.2.1.11). Analysis of all draft *Butyrivibrio* genomes identified 13 strains containing *eno* genes (Table 4.4). The *eno* genes were annotated based on the identification of enolase Pfams (C-terminal Pfam00113 and N-terminal Pfam03952), COG (COG0148), KOG (KOG2670) and KO (KO1689) domains (Figure 4.29A). The *Butyrivibrio* strains were also screened for enolase genes by PCR amplification using primers developed previously (Kelly *et al.*, 2010). Of the 13 *Butyrivibrio* strains encoding enolases, five strains (NC3005, WCD2001, VCD2006, AC2005 and AE2015) did not produce a PCR product (Figure A2.11 in Appendix 2). To investigate the possibility that the enolase primers were not detecting all enolase genes, the enolase primer sequences were aligned with the draft genomes of the 13 *Butyrivibrio* strains. A strong alignment of both forward and reverse degenerative primers was achieved for all eight *Butyrivibrio* strains that screened positive for presence of *eno*, while the five *Butyrivibrio* strains that tested negative on the PCR screens did not produce strong alignments, explaining the negative PCR results for these strains.

ML alignment of *Butyrivibrio eno* genes identified little difference amongst their amino acid sequences (Figure 4.29B). *B. fibrisolvens* strains and *Butyrivibrio* sp. 2 VCD2006 grouped together while *Butyrivibrio* sp. 2 strains AC2005, WCD2001, *Butyrivibrio* sp. 2 strain AE2015 and *B. fibrisolvens* strain NC3005 grouped separately. The two groups are correlated with the positive versus negative *eno* PCR.

Table 4.4. Enolase gene occurrence in *Butyrivibrio* and *Pseudobutyrivibrio* genomes.

	Strain	*eno* (N-/C- domains)	*eno* PCR Test	F & R *eno* primer alignment
B. fibrisolvens	MD2001 *	+	+	+
	AB2020 *	+	+	+
	FE2007	+	+	+
	ND3005	+	+	+
	YRB2005 *	+	+	+
	WTE3004 *	+	+	+
	MC2013	-	-	-
	NC3005	+	-	-
Butyrivibrio. sp. 2	AE3006	-	-	-
	MB2005	-	-	-
	WCD3002	-	-	-
	AD3002	-	-	-
	WCD2001	+	-	-
	VCD2006	+	-	-
	AE3004	-	-	-
	LC3010 *	-	-	-
	AC2005 *	+	-	-
	FC2001 *	+	+	+
	XPD2002	+	+	+
Butyrivibrio sp. 3	AE2015 *	+	-	-
	VCB2006	-	-	-
	XBB1001	-	-	-
	AE3009 *	-	-	-
	XPD2006	-	-	-
	VCB2001	-	-	-
	FCS006	-	-	-
	NC2007	-	-	-
	FCS014 *	-	-	-
	AE2032 *	-	-	-
	MB2003	-	-	-
	B316	-	-	-
Pseudobutyrivibrio				
	MA3014 *	-	-	-

+, presence of enolase gene or positive screen for *eno* PCR. -, absence of enolase gene or negative screen for *eno* PCR. * lactate producing strains. Functional identifiers of *eno* include: EC 4.2.1.11, COG0148, KOG2670, and K01689, N-terminus domain (Pfam03952) and C-terminus domain (Pfam00113).

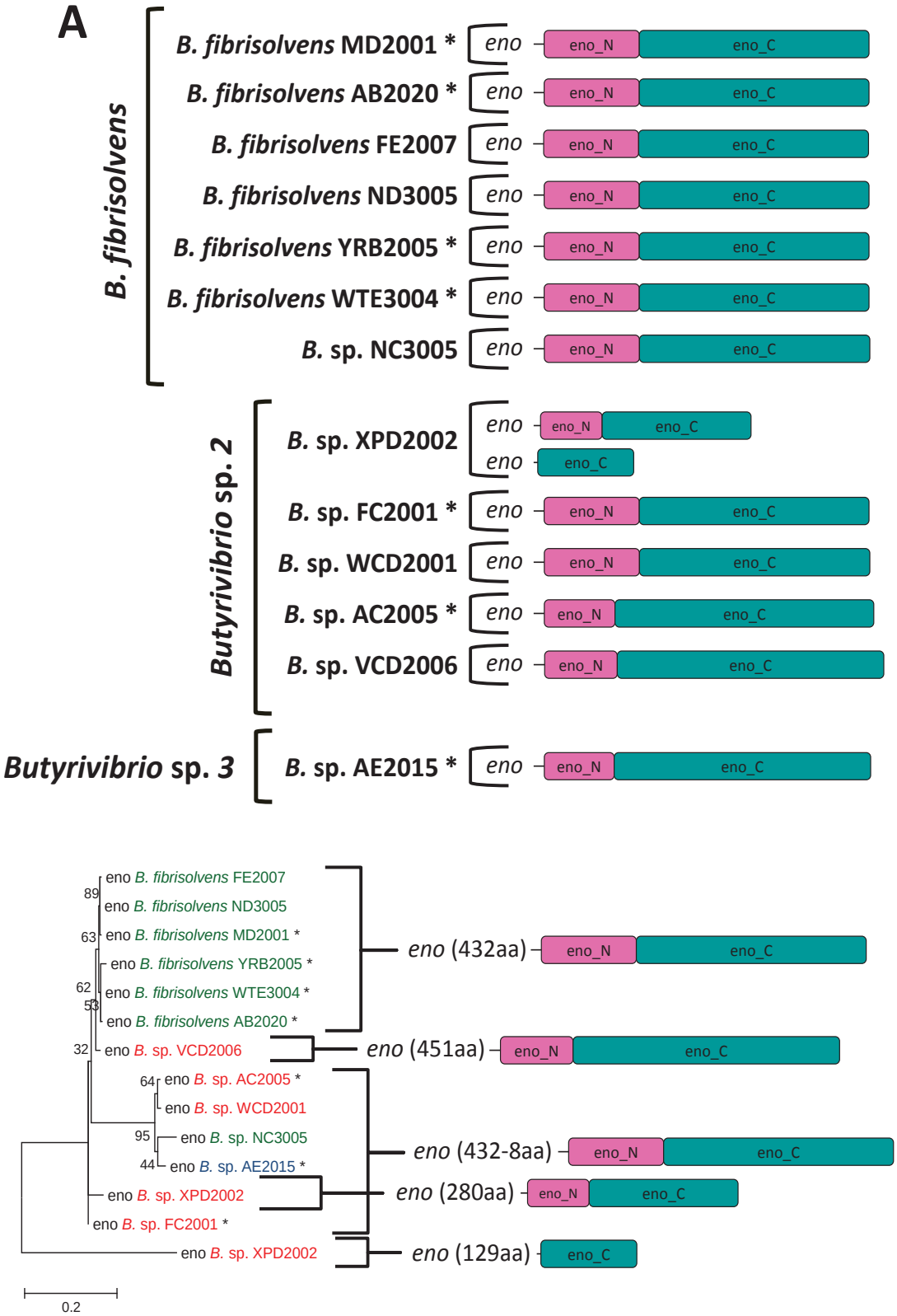


Figure 4.29. Enolases from *Butyrivibrio* genome sequences. A, comparison of Pfam domains for *eno* genes. B, Maximum Likelihood analysis of enolase proteins. The scale indicates the number of aa substitutions per site. The analysis involved 14 amino acid sequences and there were a total of 454 positions in the final dataset. Strains are coloured to represent *B. fibrisolvens* in green, *Butyrivibrio* sp. 2 in red and *Butyrivibrio* sp. 3 in blue. * lactate producing strains. Abbreviations: eno_N, enolase N-terminal domain (Pfam03952); eno_C, enolase, C-terminal TIM barrel domain (Pfam00113).

The XPD2002 strain draft genome appears to encode two *eno* genes, which occur adjacent to each other in the draft genome sequence. Alignments of the two genes with their homologues in other strains demonstrated that one of the *eno* genes was truncated. Due to the positions of the *eno* primers, the truncated *eno* of XPD2002 produced a PCR product, whereas the second full-length *eno* gene did not match the primers and was not amplified. The full-length copy of the XPD2002 *eno* gene is most similar to the *eno* of *Butyrivibrio* sp. 2 strain FC2001, hence these two enzymes group separately from those of other strains.

Previous work on B316 (Kelly *et al.*, 2010), proposed the Methylglyoxal Shunt is used by this organism to transform dihydroxyacetone phosphate (DHAP) to pyruvate, via methylglyoxal and D-lactate (Cooper, 1984). B316 has two methylglyoxal synthases (*mgsA1*, bpr_I1481 and *mgsA2*, bpr_I2940), together with enzymes that show homology to *gloA* (glyoxalase I, glyoxalase family protein, bpr_I2676), *gloB* (glyoxalase II, bpr_I1347, metallo-β-lactamase family protein) and *ldhD* (D-lactate dehydrogenase, bpr_I0186), also identified as D-isomer-specific 2-hydroxyacid dehydrogenase or glycerate dehydrogenase in IMG. These key genes of the Methylglyoxal Shunt were investigated in the draft *Butyrivibrio* genomes and were compared between the strains that did or did not encode an *eno* gene. Of particular interest were those *eno*-negative strains that produced lactate as a fermentation product (*B.* sp. AE3009 and LC3010). These findings have been summarised in Table 4.5.

The *mgsA*, *gloB* and *ldhD* genes were found in all *Butyrivibrio* genomes analysed, although the strains differed in the number of *mgsA* and *gloB* gene variants (Table 4.5). Seven strains (*B. fibrisolvens* strains FE2007, YRB2005, MC2013 and NC3005; *Butyrivibrio* sp. 2 strains WCD3002 and LC3010; *Butyrivibrio* sp. 3 strain FCS014) contained an incomplete set of Methylglyoxal Shunt pathway genes, in particular the glyoxalase I or lactoglutathione lyase (*gloA*) gene. Four of these strains (MC2013, WCD3002, LC3010 and FCS014) do not encode an *eno* gene, and two (LC3010 and FCS014) are lactate producers. These four strains may use alternative enzymes in the Methylglyoxal Shunt, or the relevant enzyme is missing from the draft genome sequences.

Table 4.5. Summary of genes involved in the Methylglyoxyl Shunt pathway.

	Strain	Methylglyoxal synthase	Glyoxalase I	Glyoxalase II	D-Lactate dehydrogenase
B. fibrisolvens	MD2001	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	AB2020	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	FE2007	*mgsA1*, *mgsA2*	-	*gloB1*, *gloB2*	*ldhD*
	ND3005	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	YRB2005	*mgsA1*, *mgsA2*	-	*gloB1*, *gloB2*	*ldhD*
	WTE3004	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	MC2013	*mgsA1*	-	*gloB*	*ldhD*
	NC3005	*mgsA1*, *mgsA2*	-	*gloB*	*ldhD*
Butyrivibrio sp. 2	AE3006	*mgsA1*, *mgsA2*, *mgsA3*	*gloA*	*gloB*	*ldhD*
	MB2005	*mgsA1*, *mgsA2*, *mgsA3*	*gloA*	*gloB*	*ldhD*
	WCD3002	*mgsA1*, *mgsA2*	-	*gloB*	*ldhD*
	AD3002	*mgsA1*, *mgsA2*, *mgsA3*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	WCD2001	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	VCD2006	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	AE3004	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
	LC3010	*mgsA1*, *mgsA2*	-	*gloB1*, *gloB2*, *gloB3*	*ldhD*
	AC2005	*mgsA1*, *mgsA2*, *mgsA3*	*gloA*	*gloB1*, *gloB2*, *gloB3*, *gloB4*	*ldhD*
	FC2001	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	XPD2002	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
Butyrivibrio sp. 3	AE2015	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
	VCB2006	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
	XBB1001	*mgsA1*, *mgsA2*	*gloA*	*gloB1*, *gloB2*	*ldhD*
	AE3009	*mgsA1*, *mgsA2*	*gloA1*, *gloA2*	*gloB1*, *gloB2*	*ldhD*
	XPD2006	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
	VCB2001	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
	FCS006	*mgsA1*, *mgsA2*, *mgsA3*	*gloA1*, *gloA2*	*gloB*	*ldhD*
	NC2007	*mgsA1*, *mgsA2*	*gloA1*, *gloA2*	*gloB*	*ldhD*
	FCS014	*mgsA1*, *mgsA2*	-	*gloB*	*ldhD*
	AE2032	*mgsA1*	*gloA1*, *gloA2*	*gloB*	*ldhD*
	MB2003	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
	B316	*mgsA1*, *mgsA2*	*gloA*	*gloB*	*ldhD*
Pseudobutyrivibrio					
MA3014		*mgsA1*	*gloA1*, *gloA2*	*gloB*	*ldhD*

Functional domains used to identify each gene of the Methylglyoxyl shunt pathway include: Methylglyoxal synthase, *mgsA* (EC: 4.2.3.3, Pfam02142, COG1803); Glyoxalase I/Lactoylglutathione lyase, *gloA* (EC: 4.4.1.5, Pfam00903, COG0346, KOG2944); Glyoxalase II/metallo-β-lactamase family protein, *gloB* (EC: 4.4.1.5, Pfam00753, COG0491, KOG0814); D-Lactate dehydrogenase/D-isomer specific 2-hydroxyacid dehydrogenase *ldhD* (EC: 1.1.1.29, Pfam00389, COG1052, KOG0069).

4.8 Discussion

In this chapter, FGD and ANI were used to examine whole-genome sequence similarity for 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains in an effort to estimate genomic diversity among this group and infer relationships between genomic groupings. Clustering microbes in this way has provided a more detailed depiction of their evolutionary history and will contribute towards a more accurate taxonomy in the future.

The FGD and ANI analyses support a clear difference between *Butyrivibrio* and *Pseudobutyrivibrio* at the genome level. Despite the distinct demarcation of these two genera, the boundaries between species were less distinct. The genome-wide analysis divides the 40 *Butyrivibrio* genomes into three main clusters; *B. fibrisolvens*, *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3, where the two species groups, *Butyrivibrio* sp. 2 and 3, are clearly distinct from one another at the genome level. It is important to bear in mind that the number of orthologous genes shared between species depends on the size of the genome and phylogenetic relationship (Bansal and Meyer, 2002; Mrázek, 2009). The *Butyrivibrio* genomes have a wide variation in genome size (Table 4.2), and this fact may influence the interpretation of the FGD and ANI analyses.

The pan-genome (29,105 genes) and the core-genome (602 genes) of all *Butyrivibrio* and *Pseudobutyrivibrio* reflect a large reservoir of genetic diversity within this group (Figure 4.6). The genes of the core-genome represent the proposed minimum set of genes that allow survival of these organisms in the rumen. The core-genome consists of genes from categories such as protein processing, folding and secretion (predominantly translation including ribosome function, maturation, modification, protein turnover and RNA degradation), cellular processes (cell division and transport), energy and metabolism (lipid metabolism, biosynthesis of nucleotides and cofactors), and also numerous poorly characterised genes (conserved hypothetical proteins, etc.). The pan-genome of these organisms, reflects their ability to occupy different niches within the rumen environment. The number of genes in the pan-genome increased along with the number of genomes analysed due to unique strain-specific genes which are found in each new strain analysed (Figure 4.6).

4.8.1 Proteins and domains involved in xylan and pectin breakdown

Given the important role of *Butyrivibrio* and *Pseudobutyrivibrio* strains in the rumen for the plant fibre degradation, genome sequence information was used to analyse their collective and individual polysaccharide-degrading potential. The analyses of the CAZymes involved in the breakdown of complex carbohydrates showed distinct variation in their distribution and abundance in both *Butyrivibrio* and *Pseudobutyrivibrio*. The analysis focused on hemicellulases and pectinases, because they are the two major plant cell wall degrading enzymes in *Butyrivibrio* and *Pseudobutyrivibrio*.

CBMs are non-catalytic carbohydrate-binding domains that are appended to many of the secreted carbohydrate active enzymes whose role is to degrade insoluble polysaccharides; these domains therefore play an important role in mediating the enzyme-substrate interaction. Thirty eight of the currently established 71 CBM families were detected in the *Butyrivibrio* and *Pseudobutyrivibrio* genomes examined, the most prevalent families being CBM48, CBM6, CBM4 and CBM2 (Figures 4.10 and 4.12). These CBM domains have predicted glycogen, β-1,4-xylan, numerous β-glucans and various xylan-binding functions, respectively (Boraston *et al.*, 2004).

Xylan can be broken down by a wide variety of enzymes including, xylanases, α-L-arabinofuranosidases, α-glucuronidases, β-xylosidases and acetyl xylan esterase (Shallom and Shoham, 2003; Lagaert *et al.*, 2009). These enzymes mainly fall into thirteen CAZyme families (Table 1.3), including GH10, GH11, GH43, GH51, GH115, GH67, GH3 and CE1 to CE7 (Cantarel *et al.*, 2009).

Families GH51, GH10 and GH115 are highly abundant and predominant GHs in *Butyrivibrio* (Figures 4.10 and 4.13). CAZymes of families GH10 and GH11 both display endo-xylanase activities (Cantarel *et al.*, 2009), but GH10 enzymes have higher substrate specificity than those of family GH11 (Ducros *et al.*, 2000). Results presented in this chapter showed that all 40 *Butyrivibrio* genomes encode GH10 enzymes, where *Butyrivibrio* sp. 3 strain AE3009 has seven GH10 genes (Figure 4.16). This family was less abundant in the *Pseudobutyrivibrio* strains; out of six analysed strains, three did not encode a GH10 (*Pseudobutyrivibrio* sp. LB2011, MD2005 and *P. ruminis* HUN009). Based on the presence of GH10 xylanase genes, the currently named *P. rumins* CF1b and AD2017 strains warrant reclassification as *P. xylanivorans*. In addition, the absence of GH10 xylanase genes is consistent with the non-xylanolytic phenotype of *P. ruminis* (Van Gylswyk *et al.*, 1996), and therefore *Pseudobutyrivibrio* sp. LB2011 and MD2005 strains should also be reclassified as *P. ruminis*. In contrast, only ten *Butyrivibrio* and one

Pseudobutyrivibrio, (*P. ruminis* CF1b) encoded a GH11 (Figure 4.17). These findings correlate with previous descriptions of a lack of xylanase activity of *P. ruminis* (Kopecny *et al*., 2003). In addition, despite the presence of GH10 xylanases, growth of *Butyrivibrio* on xylan varied considerably, however all strains with a GH11 endo-xylanase were able to utilise xylan (Figure 3.12).

It is well established that decreasing the amount of substituted arabinosyl side chains within hemicellulose polymers improves digestibility of the feed in ruminants (Brice and Morrison, 1982; Greve *et al.*, 1984). L-arabinose side chains in particular obstruct the complete hydrolysis of xylans. The arabinofuranosidases and α-arabinofuranosidases are mainly found in the GH43 (Pfam04616) and the GH51 (Pfam06964) families, and have a predominant role in the breakdown of lignocelluloses in that they hydrolyse the glycosidic bonds between L-arabinofuranoside side chains of hemicelluloses such as arabinoxylan, arabinogalactan, and L-arabinan (Paës *et al.*, 2008; Cantarel *et al.*, 2009). The GH43 family was highly abundant in all *Butyrivibrio* and *Pseudobutyrivibrio* (Figure 4.13), and the GH51 family was also highly abundant in all *Butyrivibrio* (except *B. fibrisolvens* strain MC2013) and *Pseudobutyrivibrio* (Figure 4.15). However, the VFA data showed that only the *Butyrivibrio* sp. 3 strain XPD2006 was able to utilise arabinose for growth. Moreover, it is curious that among all *Butyrivibrio* sp. 3 strains tested only AE2015 was able to utilise xylose despite it was the only strain to lack a GH30 β-xylosidase. Previous studies (Campbell and Bedford, 1992), have demonstrated that α-L-arabinofuranosidases function in concert with endo-xylanases in the hydrolysis of arabinoxylan, because xylanase activity is enhanced when presented with unsubstituted regions of xylan (Cotta, 1993; Mathlouthi *et al.*, 2002).

A xylan α-1,2-glucuronidase that belongs to the CAZyme family GH115 (Pfam15979), was demonstrated to have a key role in the degradation of xylan, liberating glucuronic acid from xylooligosaccharides and xylans (Kolenová *et al.*, 2010). The GH115 family was highly prevalent among the *Butyrivibrio* and *Pseudobutyrivibrio* strains analysed in this chapter (Figure 4.13). The *B. fibrisolvens* strain MC2013 was the only *Butyrivibrio* without a GH115, GH8 reducing end xylose-releasing exo-oligoxylanase and GH39 β-xylosidase that did not display growth on either xylan or xylose (Table 3.5 and Figure 3.12). The GH67 (Pfam07488) family CAZymes display α-glucuronidase activity and are important for the degradation of hemicellulose. This enzyme cleaves at the non-reducing α-1,2 bond between glucuronic acid side chain substitutions releasing glucuronic acid from xylo-oligosaccharides and xyloglucans, but

does not remove glucuronic acid from internal regions of xylan (Nurizzo *et al.*, 2002; Golan *et al.*, 2004). These enzymes do not possess a signal peptide and are generally intracellular, suggesting that they play a role in debranching decorated xylo-oligosaccharides inside the cell, making these molecules available for downstream degradation via β-xylosidases produced by the host (Shallom and Shoham, 2003). A single GH67 family CAZyme is present among all *Butyrivibrio* sp. 2 and 3 strains (except in strains MC2021 and AE3009) and absent in all *Pseudobutyrivibrio* and *B. fibrisolvens* strains (except in strain NC3005). Enzymes of the GH3 family are classified based on substrate specificity into β-D-glucosidases, α-L-arabinofuranosidases, β-D-xylopyranosidases, and *N*-acetyl-β-D-glucosaminidases (Harvey *et al.*, 2000), where the most common form is β-D-glucosidase (Binate *et al.*, 2008; Zhou *et al.*, 2012). Pan-genome analysis presented in this chapter showed that GH3 enzymes were abundant in all *Butyrivibrio* and *Pseudobutyrivibrio* examined. A total of 465 GH3s were found; 40% of those were found in *Butyrivibrio* sp. 2 (Figure 4.13). These findings call for further experimental validation to determine if the presence of GH3 CAZymes is associated with β-D-glucosidase activity by initially investigating *Butyrivibrio* and *Pseudobutyrivibrio* growth on glucuronic acid.

The CEs 1-7 CAZyme families represent key enzymes for xylan degradation and have known activities as acetyl xylan esterases (AcXEs) and also ferulic acid esterases (Kosugi *et al.*, 2002; Shallom and Shoham, 2003). The CE1 cinnamoyl and feruloyl esterases act on acetylated carbohydrates, including hexosides and also on cellulose acetate (Biely and Côté, 2005). The families CE1 and CE4 were present in all 46 *Butyrivibrio* and *Pseudobutyrivibrio* genomes examined (Figure 4.18A), where *Butyrivibrio* sp. 2 group had on average the largest amount of these CAZymes with *Butyrivibrio* sp. 2 strain WCD3002 having 16. This was a curious finding as most *Butyrivibrio* sp. 2 strains including WCD3002 were unable to utilise xylan for growth. The CE2 AcXE acts on position 6 in aldohexoses of hexopyranosyl residues of poly- and oligosaccharides (Topakas *et al.*, 2010), while CE3 have a wide substrate specificity, usually as domains within bi-functional enzymes; they act on chitin (Correia *et al.*, 2008). The CE3 family was found in all *Butyrivibrio* (except *B. fibrisolvens* strain MC2013) and *Pseudobutyrivibrio*. The CE6 enzymes were reported to have broad substrate specificity (López-Cortés *et al.*, 2007), and CE7 enzymes act as intracellular enzymes deacetylating oligosaccharides transported into the bacterial cytoplasm (Levisson *et al.*, 2012). The CE6

and CE7 families were much less common as the *B. fibrisolvens* group lacks CE6 and CE7 altogether while all *Butyrivibrio* sp. 2 members contain CE7 genes.

Pectin can be broken down by pectin lyases, pectate lyases, pectin esterases, polygalacturonases (PGA) and other enzymes (Alghisi and Favaron, 1995; Lagaert *et al.*, 2009). These enzymes mainly fall into twelve CAZyme families (Table 1.3), including CEs 8, 12, 13; PLs 1, 2, 3, 9, 10; GHs 28, 78, 88 and 105 (Cantarel *et al.*, 2009). Polygalacturonases of the family GH28 play a critical role in pectin degradation along with the families GH78 (α-L-rhamnosidase), GH88 (d-4, 5 unsaturated β-glucuronyl hydrolase) and GH105 (unsaturated rhamno-galacturonyl hydrolase). All four enzymes are encoded by the majority of *Butyrivibrio*, however the GH28, GH88 and GH105 families are completely absent from *Pseudobutyrivibrio*, suggesting that *Butyrivibrio* have high capacity for pectin degradation. In contrast, the GH78 family enzymes are more widespread, being present in 40 of the 46 strains.

In general, *Butyrivibrio* strains are rich in CEs, whereas *Pseudobutyrivibrio* have less CEs; *Pseudobutyrivibrio* completely lack CEs 5, 8, 12, 13 and 15, proposed to be involved in pectin breakdown. The family CE13 (pectin acetylesterase) was present in all *Butyrivibrio* examined and families CE5 and CE15 were present in only *Butyrivibrio* sp. 2 and *B. fibrisolvens*. Pectin methylesterase of the family CE8 catalyses the de-esterification of pectin to pectate and methanol and may play an auxiliary role in pectin degradation. Strains analysed here contain up to eight pectin esterases. All *Butyrivibrio* encode at least a single pectin acetyl esterase CE12 gene (except *Butyrivibrio* sp. 2 MB2005). The prevalence of CE8 varied amongst *Butyrivibrio*, except for the six *B. fibrisolvens* strains that each encode three CE8 genes and could all (except strain YRB2005) utilise pectin for growth. In fact all *Butyrivibrio* strains found lacking CE8 genes were unable to utilise pectin for growth (Figure 3.12), such as *B. hungatei* MB2003 and *B.* sp. MC2013, *Butyrivibrio* sp. 2 strains MB2005, WCD3002, AD3002, WCD2001, AE3004, LC3010. Overall, analysis of the pectin-active CE-domain-containing CAZymes suggests that *Butyrivibrio*, and notably *Butyrivibrio* sp. 3 and *B. fibrisolvens* groups, are specialists in the breakdown and utilisation of pectin, whereas *Pseudobutyrivibrio* appear to lack the necessary enzymatic repertoire to break down this complex substrate.

Analysis of the PL-domain-containing CAZymes showed that *Butyrivibrio* and *Pseudobutyrivibrio* lack PL2 and PL3 pectate lyases (Figure 4.18B). Among the 46 strains examined, 24% (11/46) lack enzymes belonging to the PL CAZyme families. In

terms of patterns between cluster groups, PLs were generally most abundant in *Butyrivibrio* sp. 3 group where all strains in this group had at least one PL. The PL content in *B. fibrisolvens* group was similar overall, except that no strains in this group encode a PL1 enzyme. The groups *Butyrivibrio* sp. 2 (with exception of strain WCD2001) and *Pseudobutyrivibrio* (except *P. ruminis* CF1b) groups, lacked many of the PLs. This, in addition to the CE distribution analysis, further confirms that the *B. fibrisolvens* and *Butyrivibrio* sp. 3 groups specialise in the breakdown and utilisation of pectin. These findings also call for further experimental validation and investigation of *Butyrivibrio* and *Pseudobutyrivibrio* utilisation of both acetyl and methyl groups and growth on galacturonic acid.

The family PL11 rhamnogalacturonans are important for the degradation of RG-I, and are highly abundant in *Butyrivibrio*, but are completely absent from *Pseudobutyrivibrio* strains. However, even though none of the strains tested could utilise rhamnose (Table 3.5), all *Butyrivibrio* strains (except for XPD2006) that tested positive for growth on pectin also possess a PL11 rhamnogalacturonan lyase (Figure 4.18). The PLs 6, 10, and 21 are *Butyrivibrio* sp. 3-specific although they are present only in a few *Butyrivibrio* strains. These PLs have known functions as alginate lyases, pectate lyases and heparin lyases, respectively (Appendix Table A2.2). The *P. ruminis* CF1b strain contains the only PL1 among all *Pseudobutyrivibrio* analysed. The analysis of the PL domain-containing CAZymes indicates that pectin is a preferred substrate for *Butyrivibrio*, especially for members of *B. fibrisolvens* and *Butyrivibrio* sp. 3 and that *Pseudobutyrivibrio* may not be able to degrade and metabolise pectin for growth. This is consistent with findings in Chapter 3 (Table 3.5), where all analysed *B. fibrisolvens* strains were able to grow on pectin, as well as 9 out of 12 *Butyrivibrio* sp. 2 strains and 8 out of 10 *Butyrivibrio* sp. 3 strains.

The two predominant plant types found in NZ pastures are perennial ryegrass (*Lolium perenne*) and the legume, white clover (*Trifolium repens*) (Waghorn and Clark, 2004). Based on the predominant pectin utilisation in *Butyrivibrio* but not the *Pseudobutyrivibrio*, it is likely that the representation of these two groups will significantly differ based on diet. In particular, ruminants given feed that is high in pectin and xyloglucans such as silage from clover, lucerne and other legumes, and also chicory, sugar beet or citrus pulp feed.

There are examples of contrasting differences in the abundance of some GH families in *Butyrivibrio* versus *Pseudobutyrivibrio*. For instance, *Butyrivibrio* strains have

many members of GHs 28, 30, 38, 65, 67, 88, 105, 112 and 129 (Table 4.3 and Figure 4.13), with predicted activities such as polygalacturonases, β-xylosidases, α-mannosidases, α-trehalases, α-glucuronidases, β-glucuronyl hydrolases, rhamnogalacturonyl hydrolases, lacto-*N*-biose phosphorylases and α-*N*-acetylgalactos-aminidases, respectively (Appendix Table A2.2). In contrast, these are completely absent from the six *Pseudobutyrivibrio* strains. Analysis of the CAZy profiles of 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* presented here suggests that, due to their extensive repertoire of GH domain-containing CAZymes, *Butyrivibrio* and in particular *B. fibrisolvens* are the predominant degraders of xylan and pectin when compared to *Pseudobutyrivibrio*.

The carbon source utilisation and VFA data combined with genome similarity and CAZyme analyses showed that *Butyrivibrio* utilisation of polysaccharides and their ability to assimilate the degradation products of these polysaccharides was variable with clear differences identified between species groups based on their initial phylogenetic placements. For example, the *B. fibrisolvens* MD2001, AB2020, FE2007, ND3005, YRB2005 and WTE3004 strains displayed similar growth patterns, in particular their ability to utilise pectin, xylan and xylose, whereas *B. fibrisolvens* sp. MC2013 and NC3005 expressed different growth patterns that coincide with their phylogenetic and genomic divergence within this cluster (Figure 4.3), along with their unique CAZyme profiles (Figure 4.9). All *Butyrivibrio* sp. 2 strains could utilise xylose and strains AE3004, XPD2002 and VCD2006 displayed similar growth patterns on all insoluble substrates analysed that is also reflected by their phylogenetic and genomic similarities. The *Butyrivibrio* sp. 3 strains B316, VCB2006, XBB1001 and AE3009 that grouped closely at the genome and phylogenetic levels, as well as having similar and highly developed CAZyme profiles, also displayed similar growth patterns across all insoluble substrates and monosaccharides analysed. These findings suggest that different *Butyrivibrio* strains can be grouped based on both genotypic and phenotypic attributes in relation to plant fibre degradation and carbohydrate assimilation and therefore may occupy different niches within the rumen in order to coexist.

4.8.2 Enolase and the glycolytic pathway

Butyrivibrio strains break down complex substrates such as xylan and pectin and use a wide range of released sugars to grow (Attwood *et al.*, 1996). The breakdown and degradation of plant material is a complex process that results in the release of a variety of sugars available to the cell and leads to production of the major fermentation end-products, butyrate, formate, acetate and lactate (Table 3.6). The predicted fermentation pathways in *Butyrivibrio* associated with the major end-products butyrate, formate and others are shown in Figure 4.30. Cellobiose is imported into the cell *via* the ABC and MFS sugar transport systems identified in all 30 *Butyrivibrio* strains. The specificity of these transporters, especially the ABC sugar transport systems, has not been determined, although there is evidence that some are involved in xylo-oligosaccharide uptake (Dunne *et al.*, 2011). Once in the cytoplasm, the cellobiose undergoes phosphorolytic cleavage via cellobiose phosphorylase (CBP) and β-glucosidase, generating glucose which then enters glycolysis. The end-products of cellobiose fermentation varied among the *Butyrivibrio* strains, however the majority produced large amounts of formate, acetate and *n*-butyrate from cellobiose. Furthermore, eleven of the *Butyrivibrio* strains also produced lactate. The production of lactate in *Butyrivibrio* has previously been reported (Diez-Gonzalez *et al.*, 1999). *B. proteoclasticus* B316, also produces lactate and was proposed to possess a D-lactate dehydrogenase (*ldhD*) gene containing the D-isomer specific 2-hydroxyacid dehydrogenase domain, and the Methylglyoxal Shunt has been proposed as a strategy by which D-lactate may be formed in this organism (Kelly *et al*., 2010). In this mechanism D-lactate is not an end product of fermentation, but an intermediate as the *ldhD* gene is responsible for conversion of D-lactate to pyruvate. However, the isomer of lactate produced was not determined in this study and is a point for future investigation. The twelve lactate-producing *Butyrivibrio* strains (including B316) reported in this work may also apply this alternative strategy for polysaccharide utilisation.

Enolase (EC4.2.1.11) (shown in red in Figure 4.30), converts 2-phospho-D-glycerate to phsophoenolpyruvate in the next to last step of the Embden-Meyerhof-Parnas (EM) glycolytic pathway. Previous work has shown that *B. proteoclasticus* B316 lacks a detectable enolase (Kelly *et al.*, 2010), and the Methylglyoxal Shunt (shown in orange in Figure 4.30) was proposed as a possible alternative to the EM pathway. In this pathway the dihydroxyacetone phosphate (DHAP) is transformed to pyruvate via methylglyoxal and D-lactate (Cooper, 1984).

Figure 4.30. Overview of possible carbohydrate metabolic pathways in *Butyrivibrio* leading to the formation of butyrate, formate, acetate and lactate from cellobiose. All metabolic pathways were compiled using information from the MetaCyc (Caspi *et al.*, 2012) and KEGG (Kanehisa and Goto, 2000) databases. The enolase catalysed reaction is shown in red as the gene was absent from a number of *Butyrivibrio* strains. Colour schemes represent formation of formate in blue, acetate in green, butyrate in purple, L-lactate in red and D-lactate proposed Methyglyoxal Shunt in orange (Cooper, 1984), metabolism pathways. Abbreviations for sugar transport systems: ABC, ATP Binding Cassette; MFS, major facilitator superfamily.

To date, the formation and utilisation of D-isoform of lactate has only been inferred in previous work on B316, as a possible alternative to glycolysis or the EM pathway. Previous to this thesis work, genes encoding L-lactate dehydrogenase enzymes had not been found in any of the analysed *Butyrivibrio* genomes. Moreover, *B. proteoclasticus* B316 encodes a D-lactate dehydrogenase containing a D-isomer specific 2-hydroxyacid dehydrogenase domain (Bpr_I0186); the same domain was found in the *ldhD* genes of the eleven lactate-producing *Butyrivibrio* strains (Figures 4.27 and 4.28).

Thirteen of the *Butyrivibrio* strains examined here were found to encode an intact enolase gene, and the enzyme *ldhD* was found in the draft genomes of all eleven *Butyrivibrio* strains that were found to produce substantial amounts of lactate. Five *Butyrivibrio* strains were found to encode L-*ldh* involved in lactate production from pyruvate, in combination with the absence of the enolase gene and incomplete set of genes for the Methylglyoxal Shunt pathway proposed as alternative pathway for production of pyruvate in enolase-negative strains. These *Butyrivibrio* strains are, *B. fibrisolvens* strain MC2013, *Butyrivibrio* sp. 2 strains WCD3002 and LC3010, *Butyrivibrio* sp. 3 strains FCS014 and AE2032. From these findings it is proposed that these five *Butyrivibrio* strains have truly novel metabolic pathways (or enzymes), the mechanisms of which await further investigation.

4.8.3 Flagellar operons and motility

Genetic characterisation of the flagellum operons using the motility assay (Chapter 3, Figures 3.9-3.12) coupled with the whole-genome sequences, identified discrepancies in the motility phenotype of *Butyrivibrio*. It is known that while *B. proteoclasticus* B316 has a flagellum, it is non-motile (Attwood *et al.*, 1996). An analysis of the flagellum synthesis/assembly operons in *B. proteoclasticus* B316 (Kelly et al, 2010) *B. hungatei* MB2003 and *Pseudobutyrivibrio xylanivorans* MA3014 (Palevich, 2011) using draft genome sequences showed variations in the region encoding key flagellum genes and the order of these genes for all strains. MA3014 lacked gene insertions, in the operon prompting a hypothesis that the gene insertion was correlated with a lack of motility (Palevich, 2011). Analysis of the flagellum operon in the *Butyrivibrio-Pseudobutyrivibrio* assemblage, however, showed variability in the gene organisation that did not correlate with motility. All *B. fibrisolvens* strains lacked gene insertions in the flagellar operon, yet only one strain, MC2013, was motile (Figures 4.24-4.26). Furthermore, *Butyrivibrio* sp. 2 strain FC2001 presented a motile phenotype despite an insertion between *flgE* and *motA*

genes. This suggests that the insertions present in the flagellum biosynthetic loci of both MB2003 and B316 non-motile controls, as well as other non-motile *Butyrivibrio* strains, does not affect the motility of the microorganisms, and that mutations elsewhere in structural or regulatory genes for the flagellum structure and function may be involved.

The presence of a non-functional flagellum begs the question as to its purpose. It is possible that *Butyrivibrio* use their flagellum for adherence rather than motility, as shown for archaeal adherence to protozoa (Shimoyama *et al.*, 2009). The ability to adhere to plant material is an essential property of the plant-fibre-degrading bacteria within the rumen, and it is possible that *Butyrivibrio* utilise their flagellum as a means of attachment to the insoluble substrates. It is possible that *Butyrivibrio* such as *B. hungatei* MB2003 that possess a flagellum but are not motile and are unable to breakdown complex polysaccharides such as xylan or pectin alone (see Chapter 5 for details), use their flagellum for attachment to other *Butyrivibrio* such as *B. proteoclasticus* B316 that initiate the breakdown process of the complex substrates. However, the mechanisms involved in bacterial adherence to plant material in the rumen are not fully understood and further investigation, including morphological confirmation of flagella via microscopy will be required to resolve the true function(s) of the flagellum in *Butyrivibrio*.

In conclusion, the comparative pan-genomic analyses of 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains has provided a wealth of knowledge and insight into the genomic diversity, metabolism and polysaccharide-degrading potential of these bacteria. This analysis has shown that *Butyrivibrio* and *Pseudobutyrivibrio* are distinct from one another, and that the proposed species groupings of *Butyrivibrio* are supported by their CAZyme complements and polysaccharide-degrading capabilities. The following chapter investigates the polysaccharide-degrading abilities and interactions of two different species of *Butyrivibrio* within the *Butyrivibrio* sp. 3 cluster, *B. proteoclasticus* B316 and *B. hungatei* MB2003, in mono- and co-culture growth on complex substrates, xylan and pectin.

Chapter 5

Xylan and pectin utilisation by *B. hungatei* MB2003 and *B. proteoclasticus* B316

5.1 Introduction

The comparative pan-genome analysis of 40 *Butyrivibrio* in the previous chapter has provided insight into the genomic diversity of these bacteria and divides them into the *B. fibrisolvens*, *Butyrivibrio* sp. 2 and *Butyrivibrio* sp. 3 species clusters. Analysis of the functional domains of CAZymes involved in the breakdown or synthesis of complex carbohydrates, has revealed the polysaccharide-degrading potential of *Butyrivibrio*. In total, 4,421 GHs, 1,283 CEs, 110 PLs, 3,605 GTs and 1,706 CBMs were identified. In general, *Butyrivibrio* species possess similar complements of CAZymes, but with variations in the absolute number of genes within each CAZy category, suggesting that *Butyrivibrio* occupy similar niches within the rumen. Although both 16*S* rRNA marker gene and genome sequence based species-group demarcations are supported by the global CAZyme profiling, these *Butyrivibrio* strains form a rather loose continuum at the species-level. This apparent functional redundancy encoded by closely related strains was investigated in more detail in this chapter, where two closely related *Butyrivibrio* sp. 3 strains, *B. hungatei* MB2003 and *B. proteoclasticus* B316, which have similar enzymatic profiles were selected for examination in relation to their use of the insoluble substrates, xylan and pectin. This was carried out both in mono-cultures grown on each of the substrates, and in co-culture situations, the aim being to see how the polysaccharide degrading potential of these strains are expressed, and how this is affected by competition for substrate with another closely related strain.

5.2 *Butyrivibrio hungatei* MB2003

Butyrivibrio hungatei MB2003 was isolated from the plant-adherent fraction of rumen contents from a pasture-grazed dairy cow as part of a rumen bacterial cultivation study (Noel, 2013). MB2003 was originally selected for genome sequencing as a representative of the *B. hungatei* species (Figure 5.1). MB2003 grows readily in broth culture, making it amenable to experimentation in the laboratory. A summary of the genome project information is shown in Appendix III Tables A3.1 and A3.2. A phylogenetic analysis of the MB2003 full-length 16*S* rRNA gene sequence places it within the *B. hungatei* species, and it is approximately 99% similar to the *Butyrivibrio hungatei* NCDO 2398 and the *Butyrivibrio hungatei* type strain, JK615 (Figure 3.1). In these phylogenetic analyses, MB2003 was assigned to the *Butyrivibrio* sp. 3 species group.

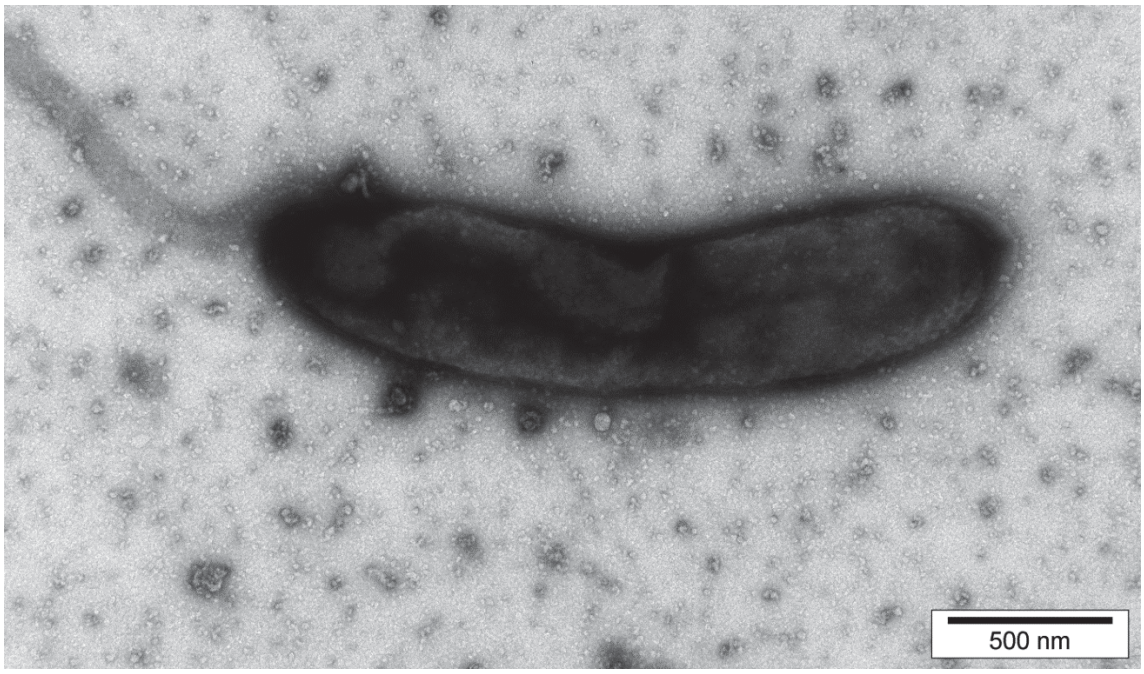


Figure 5.1. Transmission electron micrograph of a negatively stained *B. hungatei* MB2003 cell.

5.2.1 *B. hungatei* MB2003 genome properties and comparison to *B. proteoclasticus* B316.

The genome of *B. hungatei* MB2003 consists of four replicons (Palevich, 2011); a single chromosome (3,143,784 bp, %G+C 39.91), a chromid (BhuII, 91,776 bp, %G+C 37.71), a megaplasmid (pNP144, 144,470 bp, %G+C 36.86) and a plasmid (pNP6, 6,284 bp, %G+C 35.71) (Figure 5.2). The total size of the closed genome is 3,386,314 bp with an overall %G+C content of 39.71%. A total of 3,072 genes were predicted, of which 3,007 (97.89%) were protein-coding genes. A putative function was assigned to 2,258 of the protein-coding genes, while 749 protein coding genes were annotated as hypothetical proteins. The MB2003 chromosome encodes 2,767 genes, and BhuII, pNP144 and pNP6 encode 88, 147 and 5 genes respectively. The properties and statistics of the genome are summarised in Appendix III Tables A3.3 and A3.4.

Functional comparison of *B. hungatei* MB2003 and *B. proteoclasticus* B316 genomes based on COG category (Appendix III Table A3.5) and CAZy profile (Figures 4.9, 4.11, 4.13, 4.14, 4.19 and 4.22) analysis, support the notion that both these *Butyrivibrio* sp. 3 strains are functionally similar but are at the opposite ends of the continuum in terms of their total CAZyme number complements.

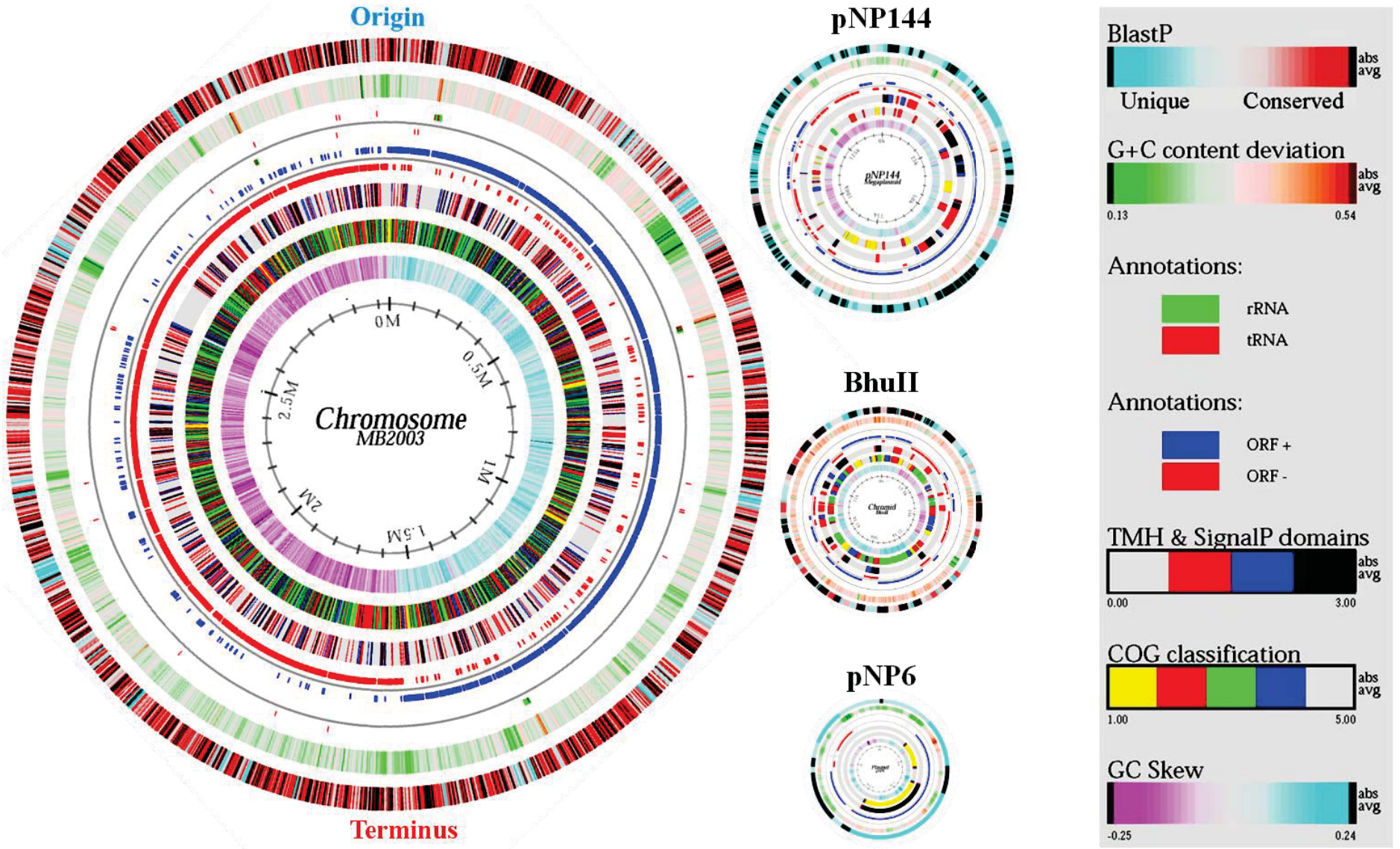

Origin
Terminus
Chromosome
MB2003
0M
0.5M
1M
1.5M
2M
2.5M
pNP144
BhuII
pNP6
BlastP
Unique
Conserved
abs
avg
G+C content deviation
0.13
0.54
Annotations:
rRNA
tRNA
Annotations:
ORF +
ORF -
TMH & SignalP domains
0.00
3.00
COG classification
1.00
5.00
GC Skew
-0.25
0.24

Figure 5.2. Genome atlas of *B. hungatei* MB2003. The figure represents a circular view of the four replicons that make up the *B. hungatei* MB2003 genome. The key at the right describes the concentric circles within each replicon in the outermost to innermost direction. The diagram was created using GENEWIZ (Jensen *et al.*, 1999) and custom-developed software. The innermost circle 1 shows GC-skew; Circle 2 shows COG classification: predicted ORFs were analysed using the COG database and grouped into the five major categories: yellow, information storage and processing; red, cellular processes and signalling; green, metabolism; blue, poorly characterised; and uncoloured, ORFs with uncharacterised COGs or no COG assignment. Circle 3 shows transmembrane helices (TMH) and SignalP domains: the four categories represent, uncoloured, absent; red, TMH; blue, SignalP; and black, both TMH and SignalP present. Circle 4 shows ORF orientation: ORFs in sense orientation (ORF+) are shown in blue; ORFs oriented in antisense direction (ORF-) are shown in red. Circle 5 shows ribosomal machinery: tRNAs and rRNAs are shown as green or red lines, respectively. Clusters are represented as coloured boxes to maintain readability. Circle 6 shows G+C content deviation from the average: GC-content are shown in either green (low GC spike) or orange (high GC spike). A box filter was applied to visualise contiguous regions of low or high GC deviations. Circle 7 shows BLAST similarities: deduced amino acid sequences were compared against the nonredundant (nr) database using gapped BLASTP (Altschul *et al.*, 1997). Regions in blue represent unique proteins in MB2003, whereas highly conserved features relative to sequences in the nr database are shown in red. The degree of colour saturation corresponds to the level of similarity. The predicted origin and terminus of DNA replication are indicated in large print.

Although the types of CAZymes were very similar (GHs, CEs, PLs, CBMs, GTs, etc.), B316 contained many more genes encoding the CAZymes, in total 342, compared to MB2003 which contained only 225 genes (Figure 4.9). From this observation it was hypothesised that the more developed polysaccharide degrading repertoire of B316 may offer B316 a competitive edge over MB2003 when grown in co-culture on insoluble substrates such as xylan and pectin. To test this hypothesis, the two strains were grown on xylan and pectin, separately in mono-cultures and together in co-cultures, to compare their individual substrate utilisation abilities and any competitive interactions that occurred during co-culture growth. Bacterial growth on xylan and pectin cannot be measured by changes in OD_{600nm} due to the cloudy nature of the medium containing these substrates. Therefore, quantitative PCR assays were developed so that each strain could be enumerated separately, or collectively when co-cultured, during their growth on these insoluble substrates. Monosaccharide release from these polymers and their subsequent utilisation was measured, along with fermentation end products and gene transcript abundance (using RNA-seq), to assess each strain's metabolism of xylan and pectin and to compare gene expression changes when grown as monocultures or when cultured together, on the same substrate.

5.3 Mono- and co-culture growth experiments

5.3.1 Quantitative PCR assays

To identify target genes for the qPCR assays, a comparison of the complete *B. hungatei* MB2003 and *B. proteoclasticus* B316 genomes was conducted using DBA which identified unique and conserved regions between genomes. A significant percentage of ORFs were very similar between *B. hungatei* MB2003 and *B. proteoclasticus* B316 (Appendix III Tables A3.6-A3.8). The gene selected as a target for the universal *Butyrivibrio* primer/probe assay was the butyrate kinase (*buk*) gene, while a cell wall binding domain-containing protein (CBD), and the α-L-arabinofuranosidase (*arf51C*) gene were selected as specific gene targets for B316 and MB2003, respectively (Table 2.8). The specificity of each primer/probe assay was demonstrated by amplification of the gDNAs of the respective target *Butyrivibrio* species and non-amplification of non-target DNAs by standard PCRs (Appendix III Figure A3.1).

A pilot study was carried out, where DNA was extracted from co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin for 24 h. qPCRs were performed on dilutions of the extracted DNAs (undiluted, 10-, 100- and 1,000-fold dilutions) using the primer/probe assays, and a linear response was detected for each strain (Appendix III Figure A3.2). The qPCR results also showed that after 24 h, growth on pectin was lower than growth on xylan (Figure 5.3). There was little sample variation amongst all triplicate samples, except for the *buk*-assayed xylan and pectin samples. For both xylan and pectin samples, the total copy numbers (per µl of DNA) detected using the universal *buk* assay were almost equal to the sum of the B316 (CBD) and MB2003 (*arf51C*) assays.

5.3.2 *Butyrivibrio* co-culture growth on xylan and pectin

The B316 and MB2003 strains were grown on xylan and pectin, either in mono- or in co-culture to examine competition between the strains for these insoluble substrates (see Section 2.2.22). The qPCR results from the mono-cultures showed that B316 was able to utilise both xylan and pectin for growth (Figure 5.4), while MB2003 showed only minimal growth on pectin, and no significant growth on xylan. In co-cultures, MB2003 displayed similar growth to B316 up until the 16 h time period when either xylan or pectin were supplied as substrates, while B316 showed reduced growth compared to the mono-cultures.

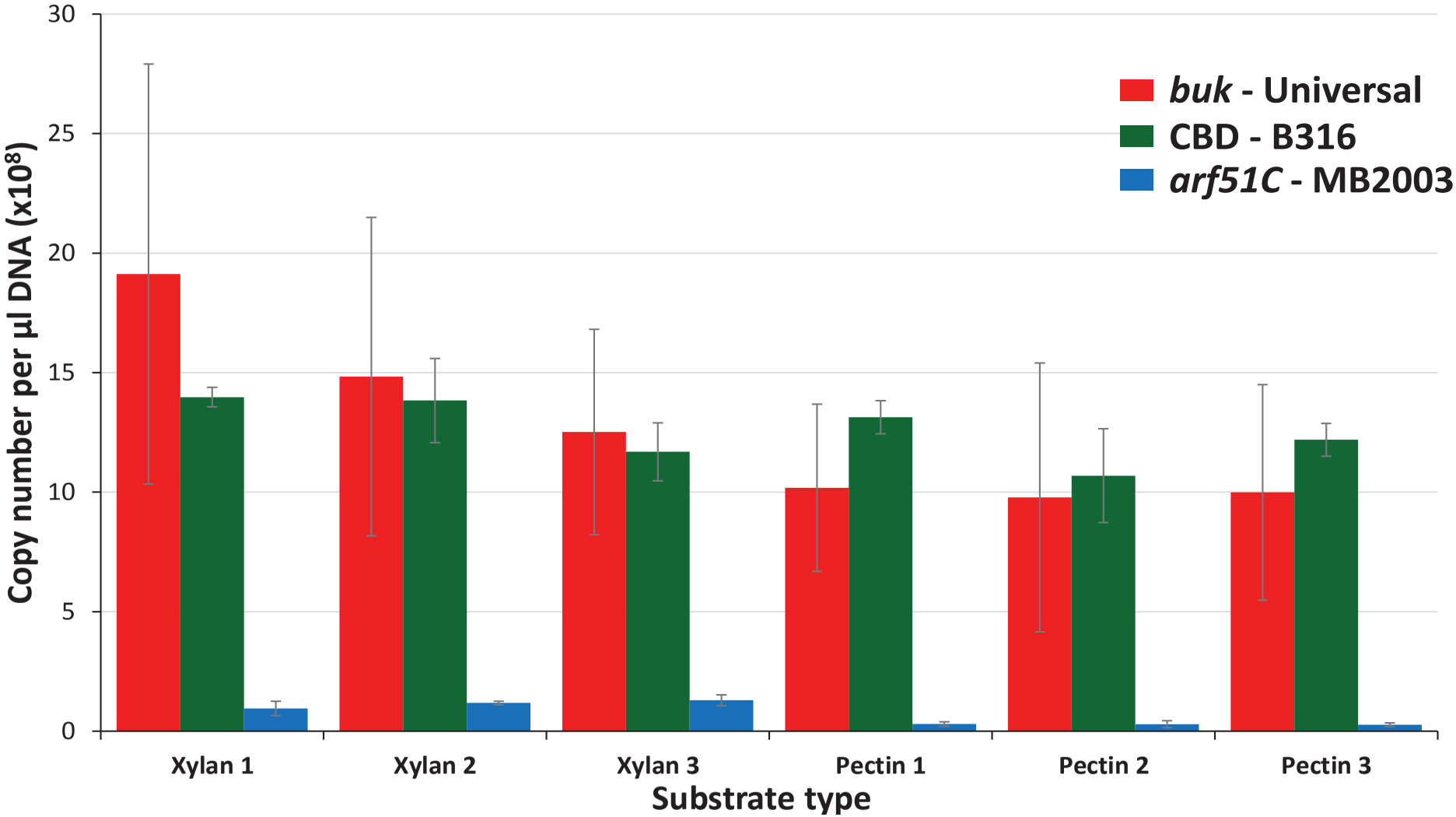


Figure 5.3. qPCR of *B. proteoclasticus* B316 and *B. hungatei* MB2003 co-cultures grown on xylan and pectin. Error bars represent ±1 standard deviation of the mean.

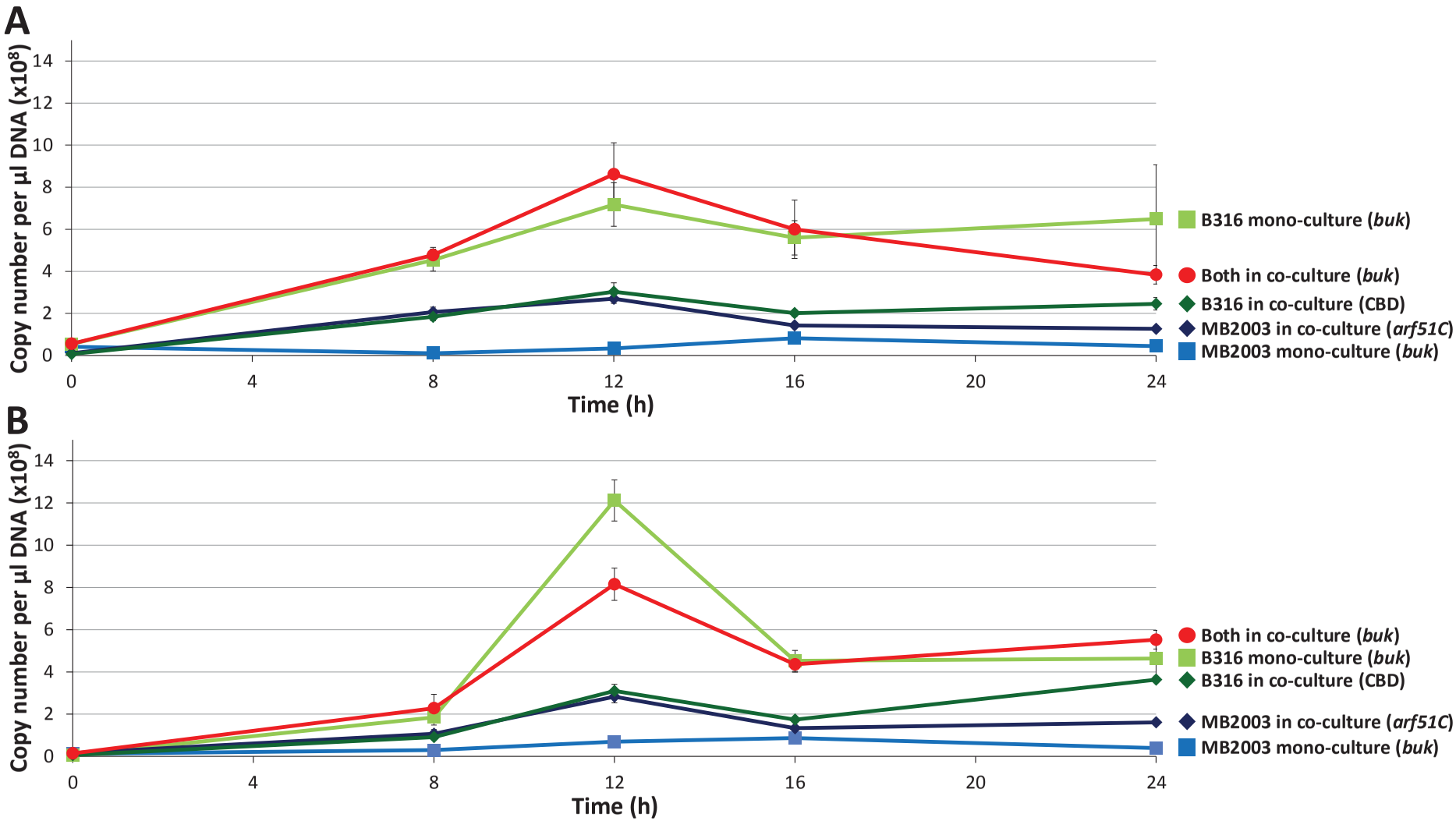


Figure 5.4. Mono- and co-culture growth of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin as determined by qPCR. A, xylan-grown cultures and B, pectin-grown cultures. Error bars represent ±1 standard deviation of the mean.

Butyrivibrio growth is typically accompanied with a change in pH of cultures, therefore pH change was used as an additional indicator of growth (Figure 5.5). The B316 mono-culture and the co-culture demonstrated a reduction of pH by approximately 0.15 and 0.2 units, respectively, for both xylan and pectin substrates over a 24 h time period. Most of the change in pH for these samples occurred within the first 8 h, and was maintained until the 24 h sampling. The pHs of the negative controls (media only and substrate supplemented media), did not change significantly during the experiments (data not shown). For MB2003 cultures, the pH did not change over the 24 h growth period for the xylan substrate. However, there was a reduction of approximately 0.05 pH units at the 8 h sampling time-point for the pectin grown samples, and the pH was further reduced by 0.12 at the 16 h time period. The changes in pH for the mono- and co-culture samples grown on the xylan and pectin substrates were statistically significant, follow the same pattern of growth as determined by the qPCR assays.

The VFAs produced after 12 h growth (mid-log phase) by MB2003 and B316 in mono- and co-cultures grown on either xylan or pectin were determined (Table 5.1). There was little sample variation amongst all triplicate samples or growth conditions. The VFAs detected were acetate, formate and butyrate, whereas lactate and succinate were not detected. Overall, large amounts of formate was produced in all B316 mono-cultures (6 mM) and co-cultures. Butyrate was the second most abundant VFA product in xylan grown cultures (1.9 mM), and the most abundant VFA product when cultures were grown on pectin was acetate (6.7 mM). For MB2003 grown on xylan or pectin in mono-culture, only trace amounts of butyrate were produced (less than 0.3 mM for both substrates) while in pectin grown cells, the total amount of VFAs produced was double that with xylan. Formate production in particular increased (0.2 to 1.0 mM). The total amount of VFAs produced in pectin co-culture samples was less than the combined amounts of VFAs produced by MB2003 and B316 mono-cultures. In contrast, co-cultures grown on xylan produced more total VFAs than the mono-cultures combined.

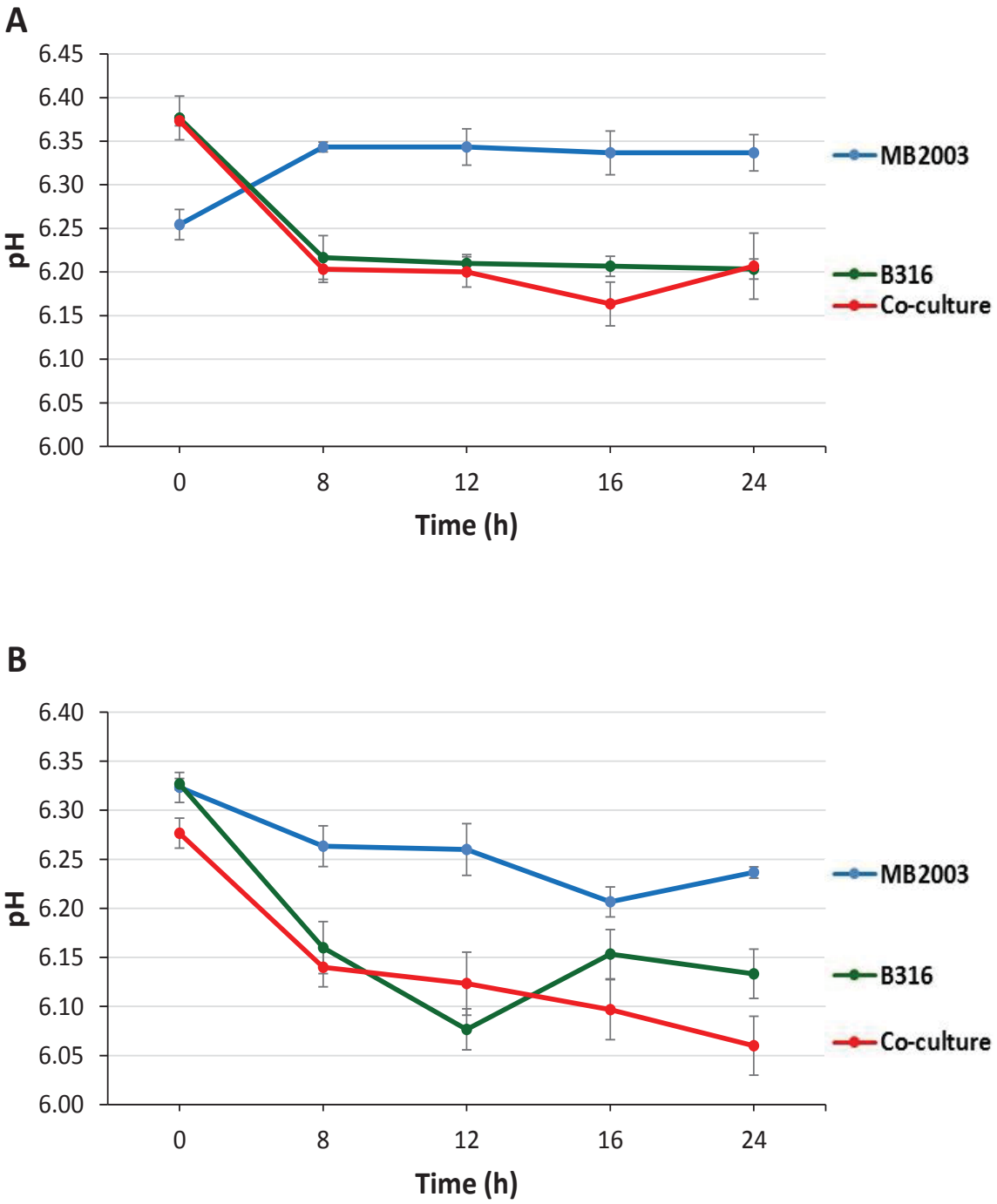


Figure 5.5. pH measurements of mono- and co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin. A, xylan-grown cultures and B, pectin-grown cultures. Error bars represent ±1 standard deviation of the mean.

Table 5.1. VFA analysis of mono- and co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan and pectin.

Substrate	Culture	VFA (mM) *					
		Acetate	SD	Butyrate	SD	Formate	SD
Xylan	*B. hungatei* MB2003	0.72	0.22	0.12	0.00	0.23	0.26
	B. proteoclasticus B316	1.38	0.36	1.94	0.04	6.14	0.04
	Co-culture	2.03	0.00	2.07	0.00	6.63	0.21
Pectin	*B. hungatei* MB2003	0.65	0.68	0.36	0.07	0.98	0.04
	B. proteoclasticus B316	6.73	0.00	0.44	0.00	6.25	0.04
	Co-culture	4.85	0.29	0.60	0.07	4.85	0.63

*, The numbers represent the change in acetate, butyrate and formate production between inoculation and 12 h post inoculation, and are the means of three replicate cultures. SD, standard deviation of the mean.

Monosaccharides released from xylan and pectin during growth were measured using ion chromatography (IC; Figures 5.6 and 5.7). For B316 cells grown on xylan, the most abundant sugar detected was xylose. The mono-culture samples showed peak concentrations of xylose (0.17 mM) at 12 h while the co-cultures showed peak concentrations (0.12 mM) at 10 h (Figure 5.6A and C). Small amounts (<0.02 mM) of arabinose and galactose were detected from 8 to 16 h of growth. The xylan-grown MB2003 mono-culture samples showed very little release of monosaccharides (Figure 5.6B). The pectin grown cultures showed similar dynamics in the mono- and co-culture samples (Figure 5.7), but the monosaccharide detected at the highest concentration was galactose, with smaller amounts of arabinose and rhamnose. Due to technical difficulties galacturonic acid could not be detected with the IC method used. The B316 mono-culture produced a sharp peak of galactose (0.08 mM) release from 6 h to 8 h (Figure 5.7A), while MB2003 released smaller amount of galactose (0.03 mM) during the same phase of growth (Figure 5.7B). A similar amount of galactose (0.04 mM) was released in the co-cultures (Figure 5.7C). Furthermore, there could have been other mono-, di- and oligosaccharides released and solubilised in the media from thermal hydrolysis of xylan and pectin upon autoclaving, and thus the correlation between bacterial growth and polysaccharide utilisation was limited to the monosaccharides that were analysed.

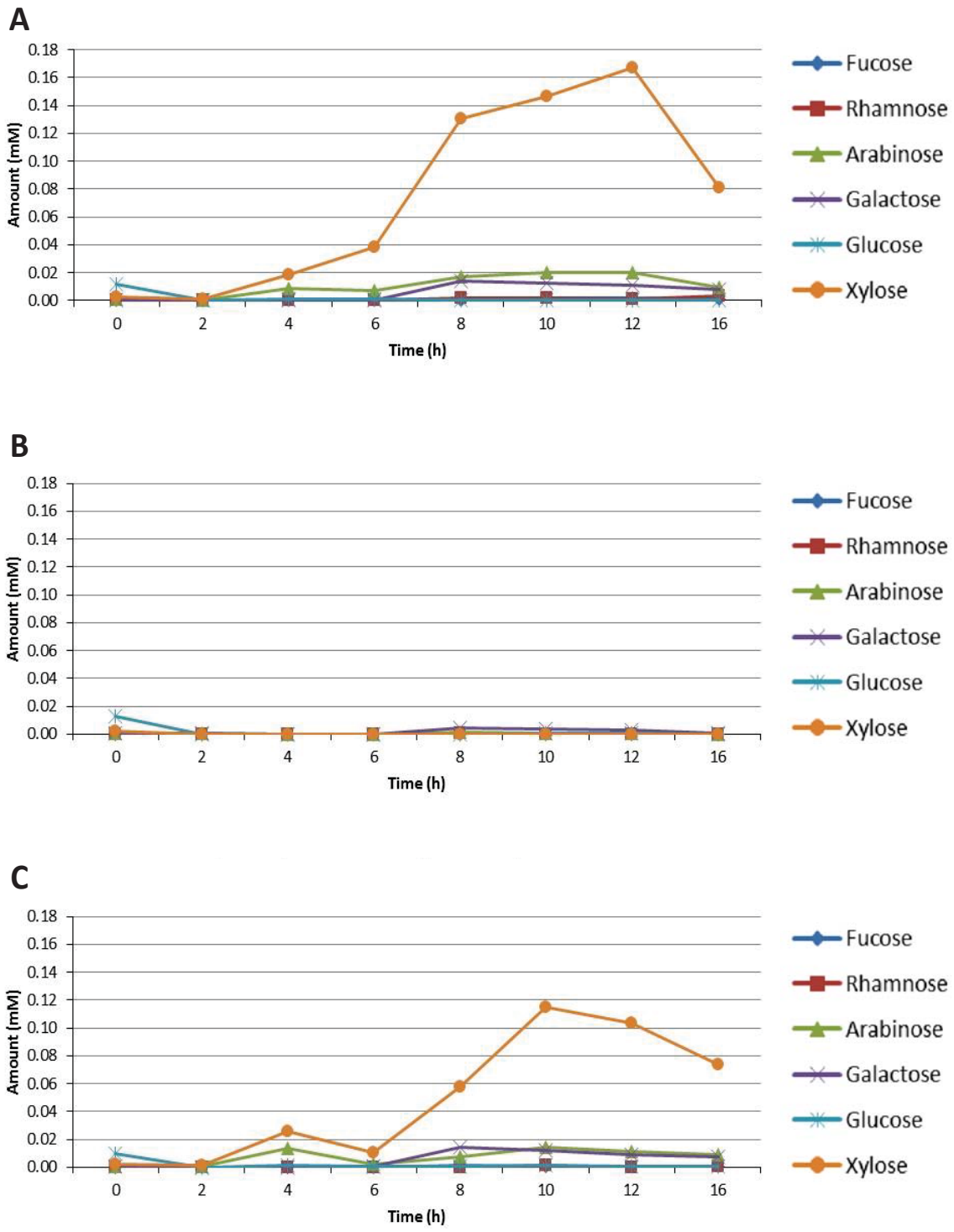


Figure 5.6. Monosaccharides released by *B. hungatei* MB2003 and *B. proteoclasticus* B316 in mono- and co-cultures grown on xylan. A, *B. proteoclasticus* B316 grown in mono-culture. B, *B. hungatei* MB2003 grown in mono-culture, C, *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown in co-culture.

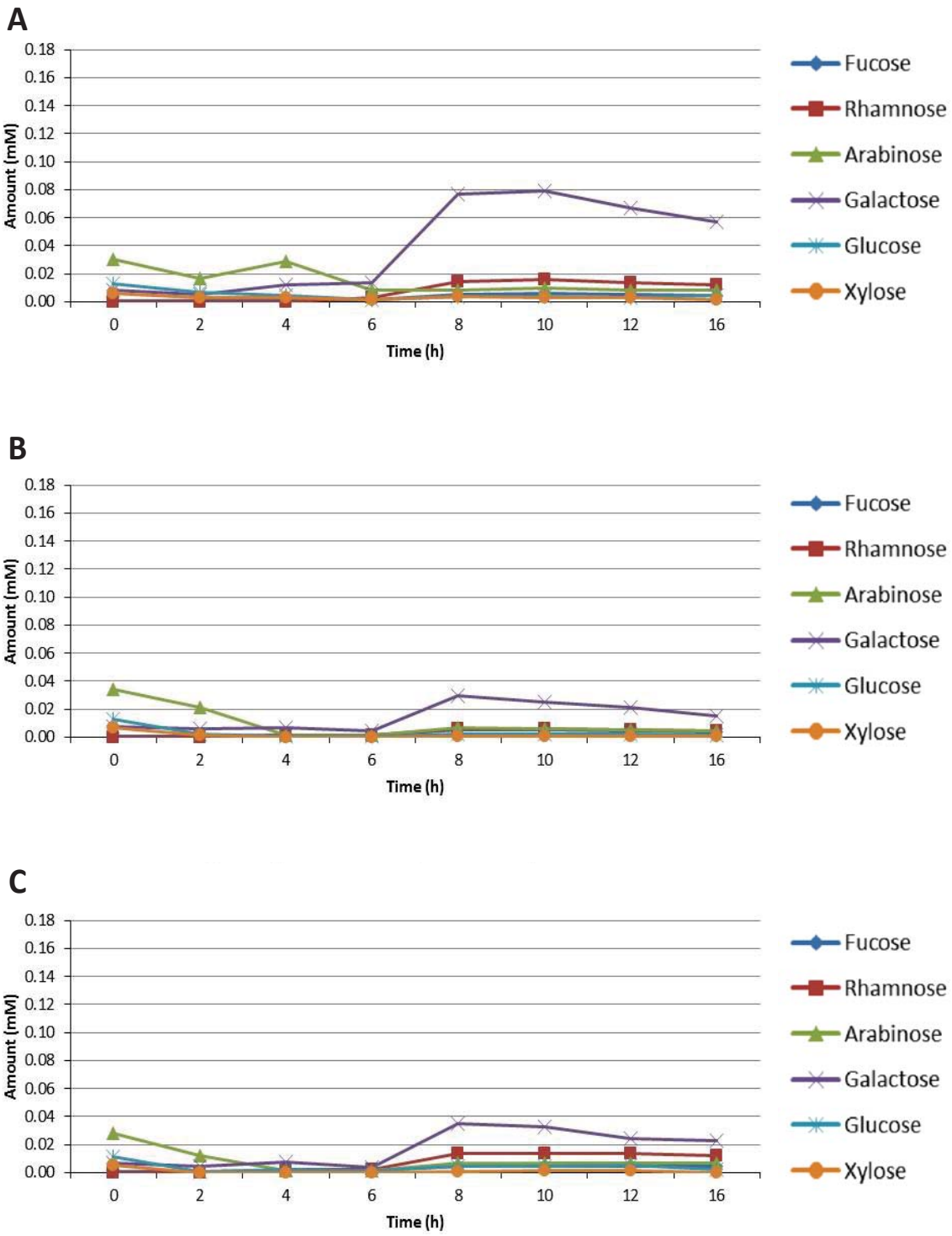


Figure 5.7. Monosaccharides released by *B. hungatei* MB2003 and *B. proteoclasticus* B316 in mono- and co-cultures grown on pectin. A, *B. proteoclasticus* B316 grown in mono-culture. B, *B. hungatei* MB2003 grown in mono-culture, C, *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown in co-culture.

5.4 Transcriptome profiling of mono- and co-cultures of *B. hungatei* MB2003 and *B. proteoclasticus* B316 grown on xylan or pectin

Mono- and co-cultures of B316 and MB2003 grown on xylan or pectin were sampled at mid-log growth (12 h) and subjected to transcriptome analyses using an RNA-seq approach. The quality and alignment statistics of the trimmed read sequences for the B316 and MB2003 genomes are detailed in Appendix III Tables A3.9-A3.11, respectively. Approximately 12,500,000 reads were generated for all the RNA-seq samples (Table A3.9). For the B316 and MB2003 mono-culture RNA-seq samples, over 80% of reads were aligned to their respective genome sequences for both the xylan and pectin growth conditions (Tables A3.10 and A3.11). For the xylan co-cultures, B316 RNA-seq samples had an alignment rate of approximately 70%, while approximately 18% of the reads aligned to MB2003. For the pectin co-cultures, B316 RNA-seq samples had a reduced alignment rate of 38%, compared to 54% for MB2003. For B316 xylan mono- and co-cultures, 58% reads mapped to protein coding genes and 39% to unannotated regions of the chromosome. These alignment proportions were similar for MB2003 mono- and co-culture xylan samples. In general, approximately 82% of B316 RNA-seq reads for the pectin mono- and co-culture conditions mapped to unannotated regions of the B316 genome and only 18% to protein coding genes. In MB2003, 67% of the reads aligned to protein coding genes in mono-culture, and to approximately 27% in the co-culture samples.

To examine possible differences in the expression of functionally relevant polysaccharide degrading and metabolism genes, the total transcript abundances of protein coding genes expressed by MB2003 and B316 in mono- and co-culture treatments were determined (Appendix III Tables A3.12 and A3.13). A total of 1,900 genes in B316 and 3,000 genes in MB2003 were investigated, and among these 1,300 and 2,206 genes had functional annotations in B316 and MB2003, respectively. In the 200 most abundant and up-regulated B316 genes were those associated with post-translational modifications, carbohydrate metabolism (various CAZymes), sugar ABC transport systems, flagella biosynthesis and some transposases (Table A3.12). The 200 most abundant and up-regulated genes in MB2003 across all treatment conditions were associated with acetyl-CoA metabolism, sugar ABC transport system, DNA replication, carbohydrate metabolism (various CAZymes) and flagella biosynthesis (Table A3.13).

5.4.1 Global transcriptome dynamics

Two forms of MDS, together with NMDS (Figures 5.8 and 5.9) were used to explore the RNA-seq data (methods described in Section 2.2.23.4). Both analyses used the full transcript datasets from the xylan and pectin conditions after stringent *Q-value* cut-off of < 0.05 (Appendix III Figures A3.3 and A3.4). The effects of the use of different distance measures on the tests for association were considered, and the Bray-Curtis distances of the normalised counts were used to measure pairwise similarity between the sample replicates. The Bray Curtis distances were calculated between pairs of expression vectors of sample replicates to obtain a distance matrix of dimensions 12 × 12 for both B316 and MB2003. The two vectors generated were used as coordinates *x* and *y* for each of the 3,811 and 3,007 data points (B316 and MB2003 plots respectively), each representing one transcript per sample, and were plotted according to the sample treatment, to make a scatter plot in a 2-dimensional plane (Figures 5.8 and 5.9).

NMDS analyses of the full datasets revealed that for the B316 transcriptome, replicate 3 for xylan mono-culture (XNC_C1_R3) and replicate 3 for pectin co-culture (PNC_C2_R3) were potential outliers among the triplicate samples (Figure 5.8A). This was displayed more clearly when the xylan and pectin datasets were analysed separately (Figure 5.8C and D). For the MB2003 transcriptome only the second replicate of the pectin co-culture (PNC_C2_R2) did not group with the rest of the pectin co-culture samples (Figure 5.9A), otherwise all sample replicates were visually inseparable from their respective condition/treatment.

Overall, the pectin MB2003 mono-culture samples grouped separately from the other samples (Figure 5.9A and B). For both B316 and MB2003, the xylan ordinates grouped with each other and separately from the pectin ordinates, with the mono- and co-culture samples also grouping separately. The differences in transcripts after the *Q-value* cut-off of < 0.05 criteria (Figure 5.9B), showed that few transcripts were differentially abundant in MB2003 for both xylan and pectin conditions inferred from non-grouping of samples.

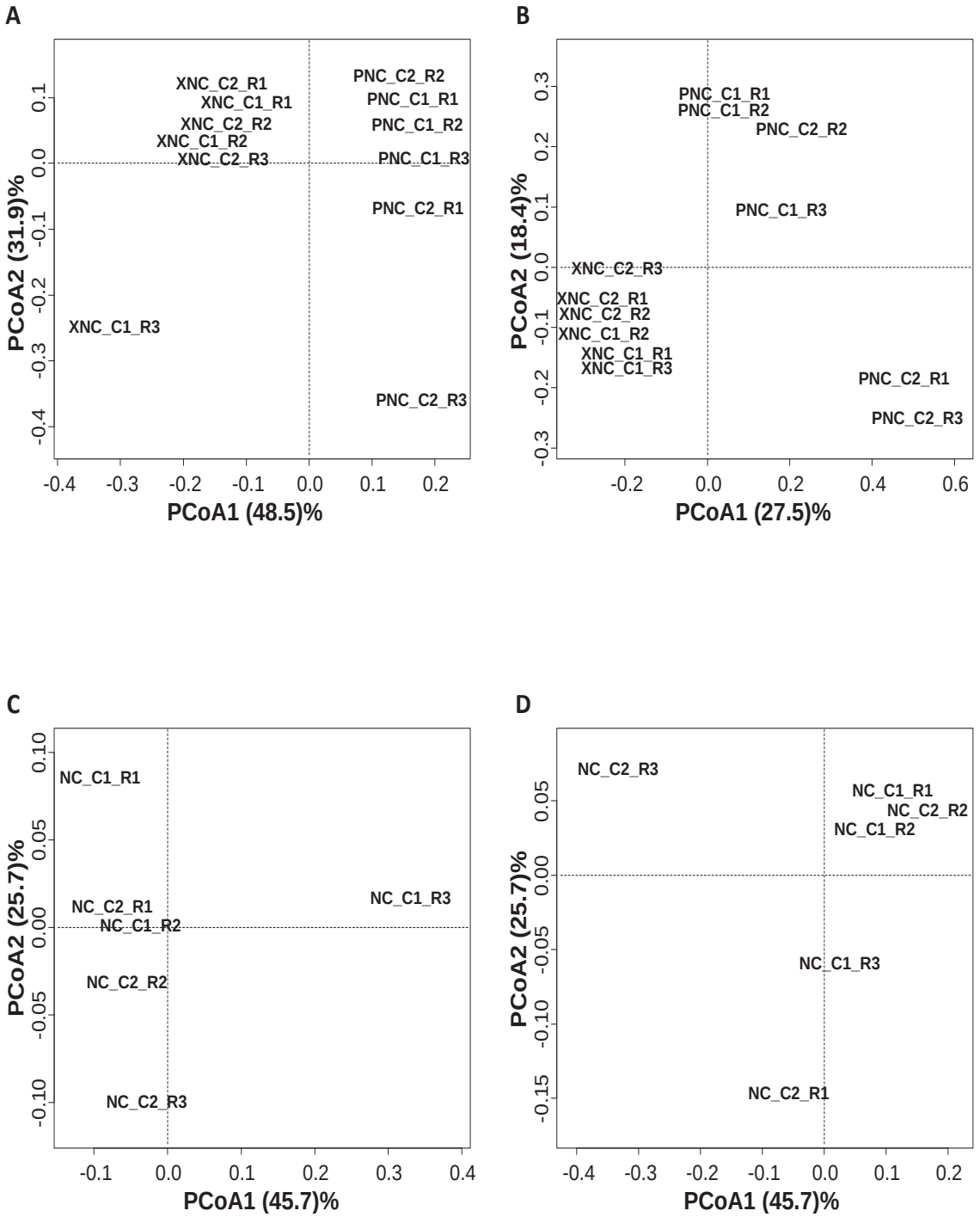


Figure 5.8. NMDS ordination showing transcriptional similarity between *B. proteoclasticus* B316 samples. A, B316 entire dataset. B, B316 subset dataset (*Q-value* < 0.05). C, B316 grown on xylan dataset. D, B316 grown on pectin dataset. Abbreviations: X, xylan; P, pectin; NC, normalised counts for transcripts; C1, mono-culture conditions; C2, co-culture condition; R, refers to sample replicates 1, 2, 3. Dashed lines represent the intersection between the first dimension on the x axis and the second dimension on the y axis.

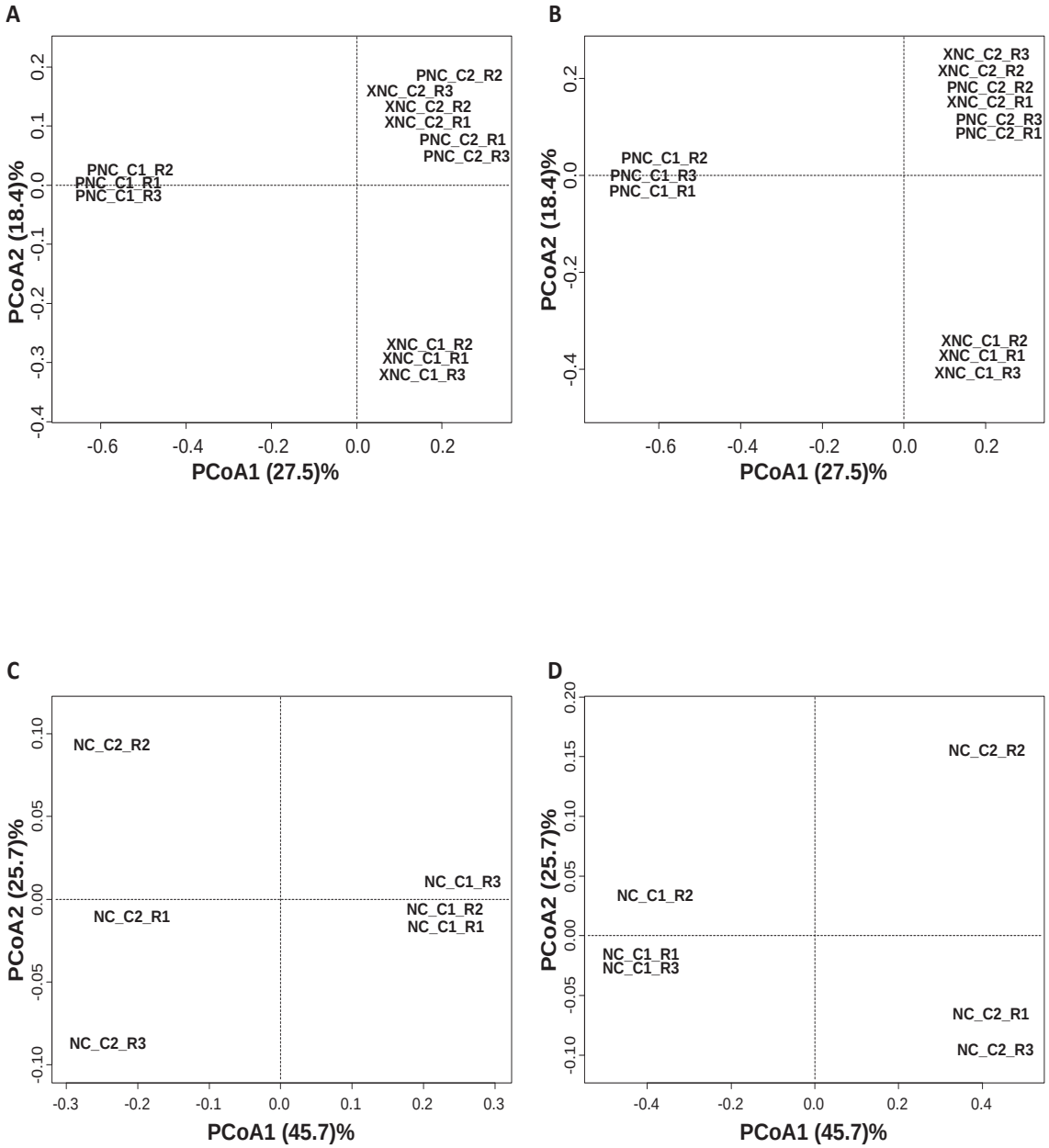


Figure 5.9. NMDS ordination showing transcriptional similarity between *B. hungatei* MB2003 samples. A, MB2003 entire dataset. B, MB2003 subset dataset (*Q-value* < 0.05). C, MB2003 grown on xylan dataset. D, MB2003 grown on pectin dataset. Abbreviations: X, xylan; P, pectin; NC, normalised counts for transcripts; C1, mono-culture conditions; C2, co-culture condition; R, refers to sample replicates 1, 2, 3. Dashed lines represent the intersection between the first dimension on the x axis and the second dimension on the y axis.

To check the homogeneity of variation within the groupings of replicate samples, a permutation based test, that examines the variance within treatment groups versus the variability among different groups of samples, was considered. The permutation-based test was considered for homogeneity of group dispersions for the entire and sub-group datasets (*Q-value* < 0.05 cut-off) and metric scaling in the form of Principal-Coordinate Analysis (PCoA) was used (Figure 5.10). The analysis of the B316 data indicates that even though the xylan and pectin treatments scattered differently, the scatter is uniform for the mono- and co-culture groups (Figure 5.10A and B). The large distances from individual points to group centroids in the xylan mono-culture (XC1) and pectin co-culture (PC2) samples, are indicative of sample variation. For both MB2003 plots, the co-culture groups for both xylan and pectin treatments were located close to one another, indicative of low sample variation among these groups (Figure 5.10C and D). Interestingly, a considerable distance was observed between both mono-culture groups and the co-culture groups, indicating significant differences between the mono- and co-culture samples that use the xylan and pectin as the major carbon source.

A simple CA on the two-way frequency *Q-value* cut-off data (*Q* < 0.05) of B316 and MB2003 was carried to examine the change in proportion and composition of various differentially expressed genes against the mono- and co-culture, xylan and pectin growth conditions. The CA results only reveal the associations between gene expression and condition/treatment groupings in two dimensions (Appendix III Figures A3.5-A3.10). Out of the total number of genes in B316 (3,813), 47 were found with 0 variance, thus the pool of genes for analysis was reduced to 3,766 in total. For MB2003, no genes had 0 variance, hence the pool of genes for analysis was 3,007. The gene subsets that complied with the FDR *Q-value* cut-off of < 0.05 numbered 313 and 617 for B316 and MB2003, respectively.

The B316 and MB2003 scatterplots show there are no particular groups of individual genes or treatment samples and both datasets are quite homogeneous (Figures A3.5 and A3.6). For the B316 total dataset (Figure A3.7), there were large numbers of up- and down-regulated genes for all conditions; however, there were differences in up-regulated genes between the xylan and pectin conditions. From the B316 subset dataset (Figure A3.8), the gene expression pattern in the pectin mono- (PC1) and co-culture (PC2) conditions appear to diverge from that in the xylan-containing mono- (XC1) and co-culture (XC2).

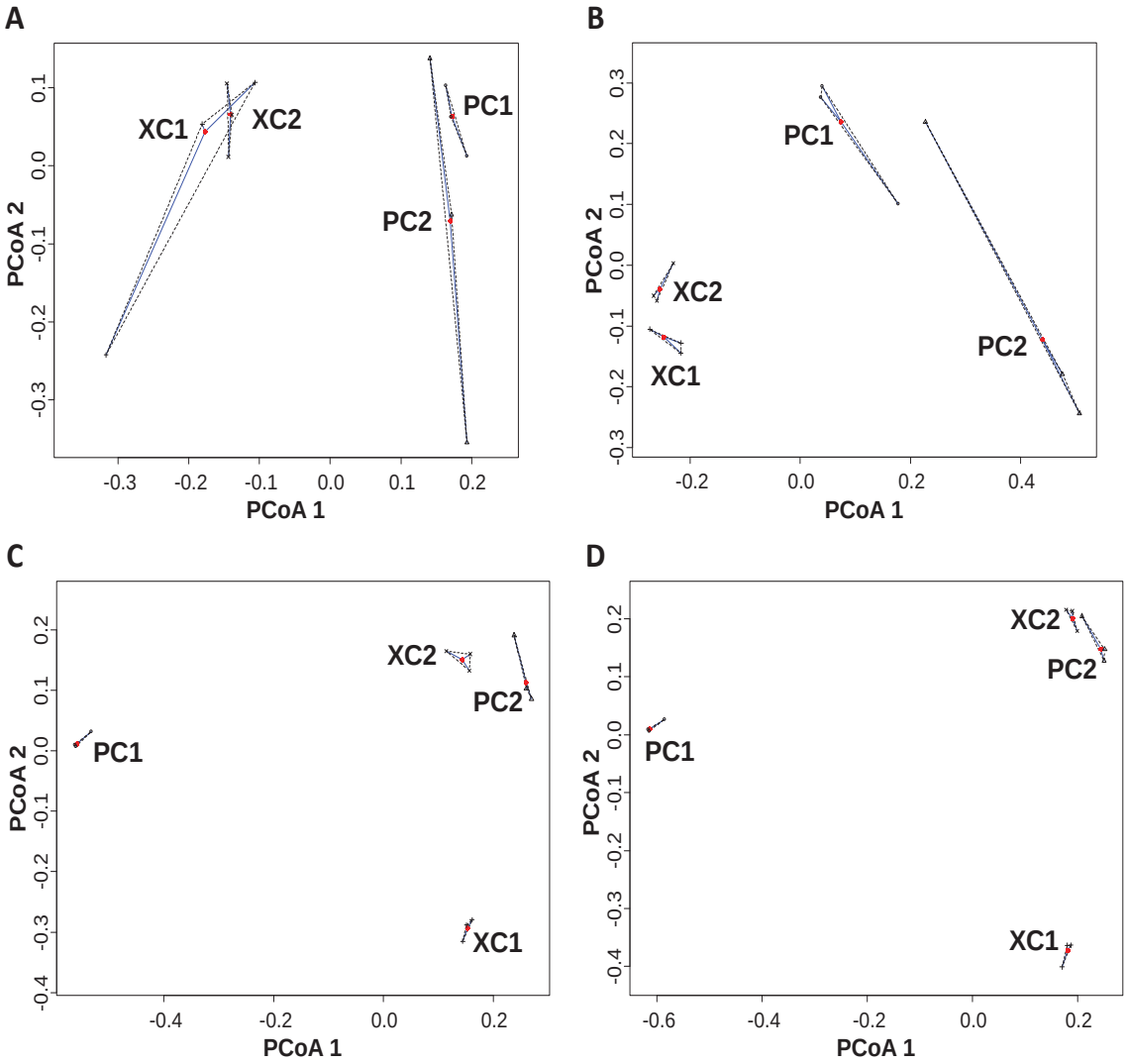


Figure 5.10. Permutation-based test of multivariate homogeneity of group dispersions and variances. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. A, *B. proteoclasticus* B316 transcriptome and B, *B. proteoclasticus* B316 $Q < 0.05$ subset transcriptome; C *B. hungatei* MB2003 transcriptome, D, *B. hungatei* MB2003 $Q < 0.05$ subset transcriptome.

For the xylan condition (Figure A3.8C and D), a similar number of genes were either up- or down-regulated in mono-culture (XC1), whereas in the co-culture (XC2), there were substantially more genes significantly up-regulated, as shown by the large distances of the genes from the centroid line. For the B316 pectin subset data (Figure A3.8A and B), there was also a similar pattern between the mono- (PC1) and co-culture (PC2).

To investigate correlations between gene expression and the various treatment groups, a network analysis of the B316 dataset genes with a FDR *Q-value* cut-off of < 0.05 was done (Figure 5.11). A total of 247 genes within this FDR cut-off were identified whose transcripts increased by CA values of 0.15 or more. All the up-regulated genes were associated with the pectin treatment (PC1 and PC2). A group of 50 genes were also down-regulated in xylan mono- and/or co-culture treatments (Appendix III Table A3.14). Among these, 27 genes were significantly (CA values of 0.16 to 0.17 ↑↑) up-regulated (5 in PC1 and remainder in PC2), and 11 strongly up-regulated genes (CA values of 0.17 or more ↑↑↑), predominantly represented by hypothetical proteins (Table 5.2). MB2003 had a large number of genes that were significantly up-regulated in both the full dataset and the $Q < 0.05$ subset (Figures A3.9 and A3.10).

To display all the relevant associations in one graph, CIMs were produced using the regularised correlation analysis. CIMs generated using the entire B316 and MB2003 datasets are shown in Appendix III Figures A3.11 and A3.12, while Figures A3.13 and A3.14 show CIMs constructed from the B316 and MB2003 $Q < 0.05$ subsets, respectively. A network plot for MB2003 was not included due to the large number of transcripts.

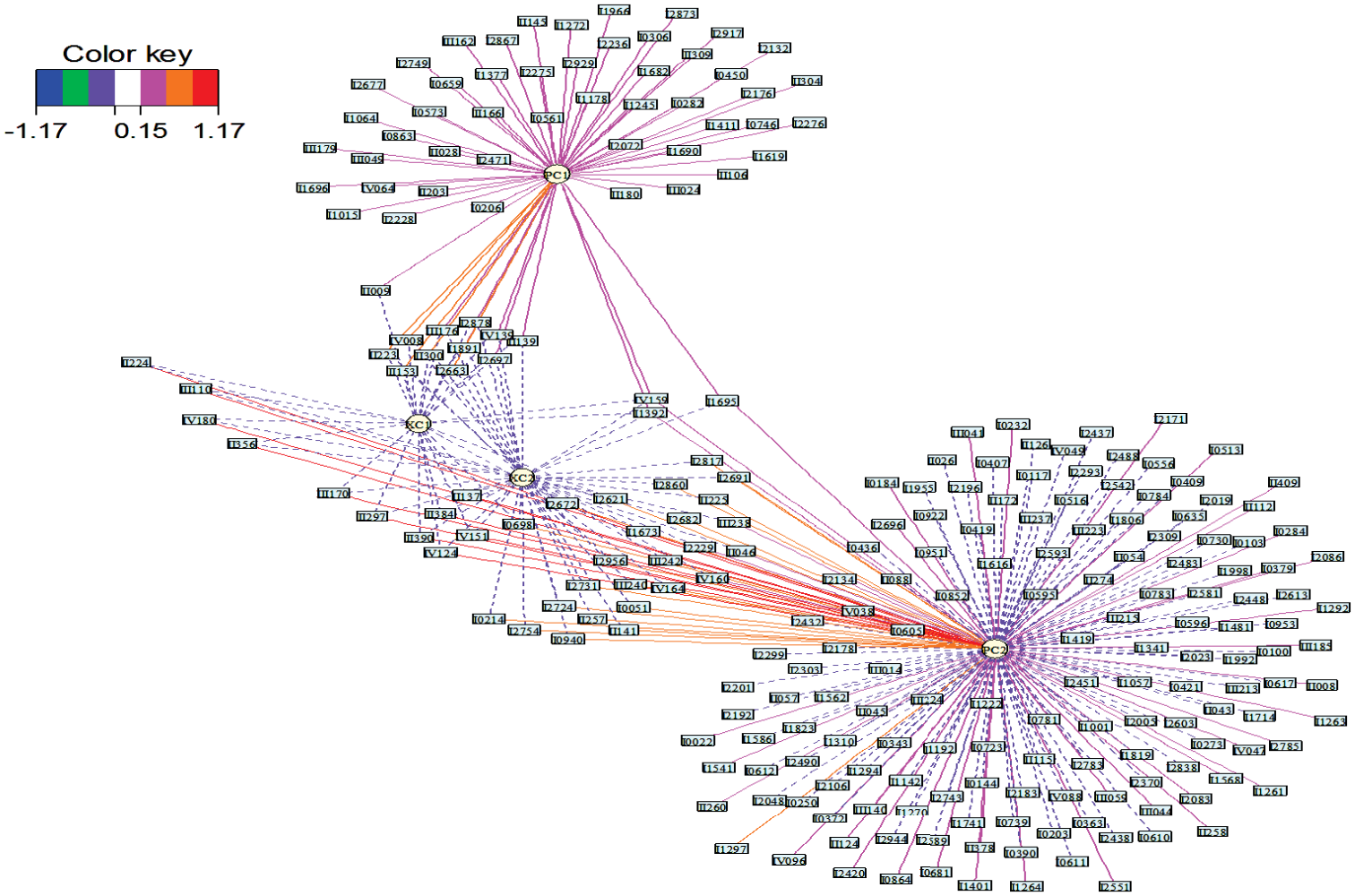


Figure 5.11. Network analysis of the subset *B. proteoclasticus* B316 dataset. The network was constructed based on CA of the *B. proteoclasticus* B316 dataset genes with FDR *Q-value* cut-off of < 0.05. The colour key indicates expression levels; the threshold for expression was set at 0.15 CA units. Each sample is represented with the following abbreviations: X, xylan treatment; P, pectin treatment; C1, mono-culture; C2, co-culture.

Table 5.2. List of *B. proteoclasticus* B316 subset genes from Network analysis.

Locus tag	Gene annotation	Name	Treatment correlation *		
bpr_II153	hypothetical protein	-	↑↑PC1	↓XC1	
bpr_I2663	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2
bpr_II300	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2
bpr_II223	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2
bpr_IV008	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2
bpr_II224	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_II356	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_II297	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_II390	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_II384	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_IV180	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_IV124	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_IV151	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_II137	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_III170	HTH domain-containing protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_III110	response regulator domain-containing protein	-	↑↑↑PC2	↓XC1	↓XC2
bpr_I2621	cysteine desulfurase	-	↑↑PC2	↓XC1	↓XC2
bpr_I2672	GNAT family acetyltransferase	-	↑↑PC2	↓XC1	↓XC2
bpr_I0698	sugar ABC transporter permease	-	↑↑PC2	↓XC1	↓XC2
bpr_I2817	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I2691	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I2860	glycoside hydrolase 31	-	↑↑PC2	↓XC2	
bpr_I2682	sugar ABC transporter permease	-	↑↑PC2	↓XC2	
bpr_I1673	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I2956	tRNA modification GTPase	*trmE*	↑↑PC2	↓XC2	
bpr_I2731	GNAT family acetyltransferase	-	↑↑PC2	↓XC2	
bpr_I2724	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I2754	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I0214	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I0940	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_I0051	FeS assembly ATPase	*sufC*	↑↑PC2	↓XC2	
bpr_II225	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_II141	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_II257	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_IV164	hypothetical protein	-	↑↑PC2	↓XC2	
bpr_III240	sugar ABC transporter substrate-binding protein	-	↑↑PC2	↓XC2	
bpr_III242	sugar ABC transporter permease	-	↑↑PC2	↓XC2	
bpr_I1297	SsrA-binding protein	*smpB*	↑↑PC2		

*, Abbreviations: X, xylan treatment; P, pectin treatment; C1, mono-culture condition; C2, co-culture condition. The network was based on CA of the *B. proteoclasticus* B316 dataset with FDR *Q-value* cut-off of < 0.05. The increase and/or decrease in gene expression is represented by ↑ and ↓, respectively (see colour key in Figure 5.11 for expression levels). The threshold for expression was set at a CA value of 0.15.

5.4.2 Transcriptome assembly and differential gene expression

After sequence quality control assessments and quantile normalization, 3,813 and 3,007 transcripts corresponding to protein-coding genes from B316 and MB2003, respectively, were used in the analyses (Tables 5.3 and 5.4). Once the non-coding RNA (ncRNA) transcripts were removed, the total transcript counts for both B316 and MB2003 matched the total numbers of ORFs within each genome. This indicates that the transcriptome assembly performed by the Rockhopper system was accurate. MB2003 did not have any reads mapping to tRNAs, compared to B316 that had 47 tRNA reads for the chromosome, two for chromid and one for the plasmid pCY186, equating a total of 50 tRNAs for both xylan and pectin conditions.

Differential transcript abundance between the mono-culture and co-culture growth of B316 and MB2003 on the insoluble substrates xylan and pectin were determined. The FDR was calculated using the Benjamini-Hochberg procedure and reported as *Q*-values. A mean detection-above-background *Q-value* cut-off of < 0.05 was used, which produced 184 and 556 DEGs for growth on xylan, and 238 and 1,195 for growth on pectin, for B316 and MB2003 respectively (Tables 5.3 and 5.4).

To test the strength of the *p*- and *Q-values* generated from the Rockhopper software the KW analysis of variance was used. The total numbers of DEGs that also satisfied the KW adjusted $p < 0.05$ cut-off were 75 and 461 for xylan condition, 57 and 807 for pectin condition in B316 and MB2003 respectively (Tables 5.3 and 5.4). Also the Benjamini Hochberg analysis was done based on the KW *p-values* to calculate the adjusted *p-values*.

The full data set was comprised of the merged mono- and co-culture transcripts of B316 and MB2003 on both xylan and pectin growth conditions. Resulting *p-values* from the ANOVA and *t*-tests were corrected for multiple comparisons using the Benjamini Hochberg method (Benjamini and Hochberg, 1995) into *Q-values*. These *p-values* from all statistical analyses were merged together and compared with *p*- and *Q-values* from Rockhopper software for the xylan and pectin subsets separately to determine significant differentially expressed genes that comply with all analyses (BHaov2t $p < 0.05$ & BHkrus $p < 0.05$). Conservative thresholds of FDR *Q-value* < 0.05, KW adjusted *p-values* of < 0.05 and $\geq$ 2-fold $\log_2$-transformed signal intensity difference were used to define DEGs. The numbers of significant DEGs ($\geq$ 2-fold $\log_2$ difference) for the mono- and co-culture conditions following the FDR ($Q < 0.05$), KW ($p < 0.05$) cut-offs, are also shown in Tables 5.3 and 5.4. The full results for these statistical analyses are displayed in Appendix III Figure A3.15.

Table 5.3. Transcriptome assembly and differential gene expression analysis for growth on xylan.

				Protein coding transcript counts				
Sample	**Replicon**	**ORFs in genome**	**Total Mapped**	**FDR [A] $Q < 0.05$**	**KW [B] adjusted $p < 0.05$**	**Q/p [C] < 0.05 & $\log_2 \geq 2$**	**Number of DEGs [*] in mono-culture**	**Number of DEGs [*] in co-culture**
B316 (Xylan)	chromosome	2939	2939	152	66	14	8	6
	pCY290	251	251	12	3	1	1	0
	pCY360	425	425	10	4	1	0	1
	pCY186	198	198	10	2	1	1	0
	Total	**3813**	**3813**	**184**	**75**	**17**	**10**	**7**
MB2003 (Xylan)	chromosome	2767	2767	501	413	279	189	90
	BhuII	88	88	25	24	13	11	2
	pNP144	147	147	30	24	15	15	0
	pNP6	5	5	0	0	0	0	0
	Total	**3007**	**3007**	**556**	**461**	**307**	**215**	**92**

[A], FDR reported as *Q-values*. [B], KW analysis of variance reported as *p-values*. [C], DEGs. *, Numbers of DEGs were calculated by considering the mono- and co-culture conditions separately.

Table 5.4. Transcriptome assembly and differential gene expression analysis for growth on pectin.

				Protein coding transcript counts				
Sample	**Replicon**	**ORFs in genome**	**Total Mapped**	**FDR [A] $Q < 0.05$**	**KW [B] adjusted $p < 0.05$**	**Q/p [C] < 0.05 & $\log_2 \geq 2$**	**Number of DEGs [*] in mono-culture**	**Number of DEGs [*] in co-culture**
B316 (Pectin)	chromosome	2939	2939	167	42	42	38	4
	pCY290	251	251	13	5	4	2	2
	pCY360	425	425	45	9	9	8	1
	pCY186	198	198	13	1	1	1	0
	Total	**3813**	**3813**	**238**	**57**	**56**	**49**	**7**
MB2003 (Pectin)	chromosome	2767	2767	986	671	659	604	55
	BhuII	88	88	80	57	51	47	4
	pNP144	147	147	129	79	77	74	3
	pNP6	5	5	0	0	0	0	0
	Total	**3007**	**3007**	**1195**	**807**	**787**	**725**	**62**

[A], FDR reported as *Q-values*. [B], KW analysis of variance reported as *p-values*. [C], DEGs. *, Numbers of DEGs were calculated by considering the mono- and co-culture conditions separately.

For B316, only 0.5% and 1.5% of the total transcripts complied with these cut-offs for the xylan and pectin conditions, respectively, compared to MB2003 where 10.2% and 26.2% complied for the xylan and pectin conditions, respectively.

CIMs were constructed based on CA analysis of the significantly up-regulated genes for B316 and MB2003 and highlight the overall expression differences between the two bacteria across all growth conditions and on both substrates. B316 significantly up-regulated expression of only a select few genes across all growth conditions, most notably when grown on pectin in both mono- and co-culture (Figure 5.12). In contrast, MB2003 significantly up-regulated many genes during mono-culture growth on pectin (Figure 5.13). The gene names are not included in the CIM for MB2003 due to the large number of transcripts and instead are listed in Appendix III Tables A3.16 and A3.18.

To investigate the correlations between gene expression and various treatment groups, a network analysis of the CA significantly up-regulated genes was carried out. For B316, a total of 43 genes were identified at the 0.15 expression threshold (CA value scale), where all the up-regulated genes were associated with the pectin treatment and 5 genes were also down-regulated in the xylan mono- (bpr_II009) and co-culture (bpr_II257, bpr_II046, bpr_III242 and bpr_I2860) treatments. Among these, bpr_II009 (PC1) and bpr_I2420 (PC2) were significantly up-regulated, where bpr_II257, bpr_III242 and bpr_I2860 associated with PC2 were strongly up-regulated (Figure 5.14). A network plot for MB2003 is not included due to the large number of transcripts. This analysis has shown that gene expression of DEGs differs among the four treatment groups for B316, whereas for MB2003 the vast majority of DEGs were expressed in only the PC1 treatment group.

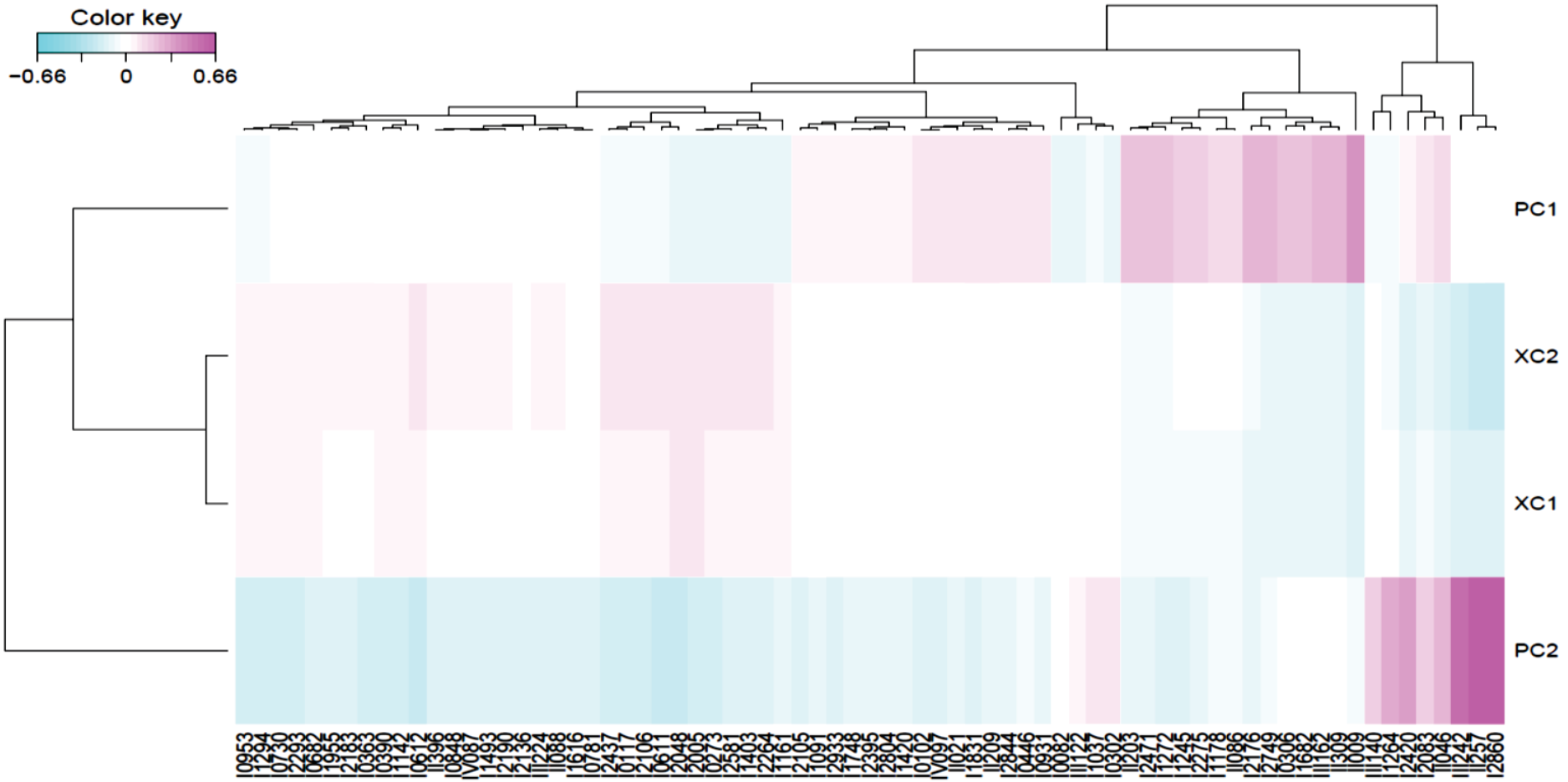


Figure 5.12. CIM analysis of genes significantly up-regulated in *B. proteoclasticus* B316. Genes were deemed "significantly expressed" if all three thresholds were met: FDR calculated using the Benjamini-Hochberg procedure *Q*-values < 0.05, KW analysis of variance with $p < 0.05$, and ≥ 2-fold $\log_2$ difference in expression. CIM was constructed using the similarity matrix from hierarchical clustering based on CA of the *B. proteoclasticus* B316 dataset. The colour key at the top left indicates expression levels. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture.

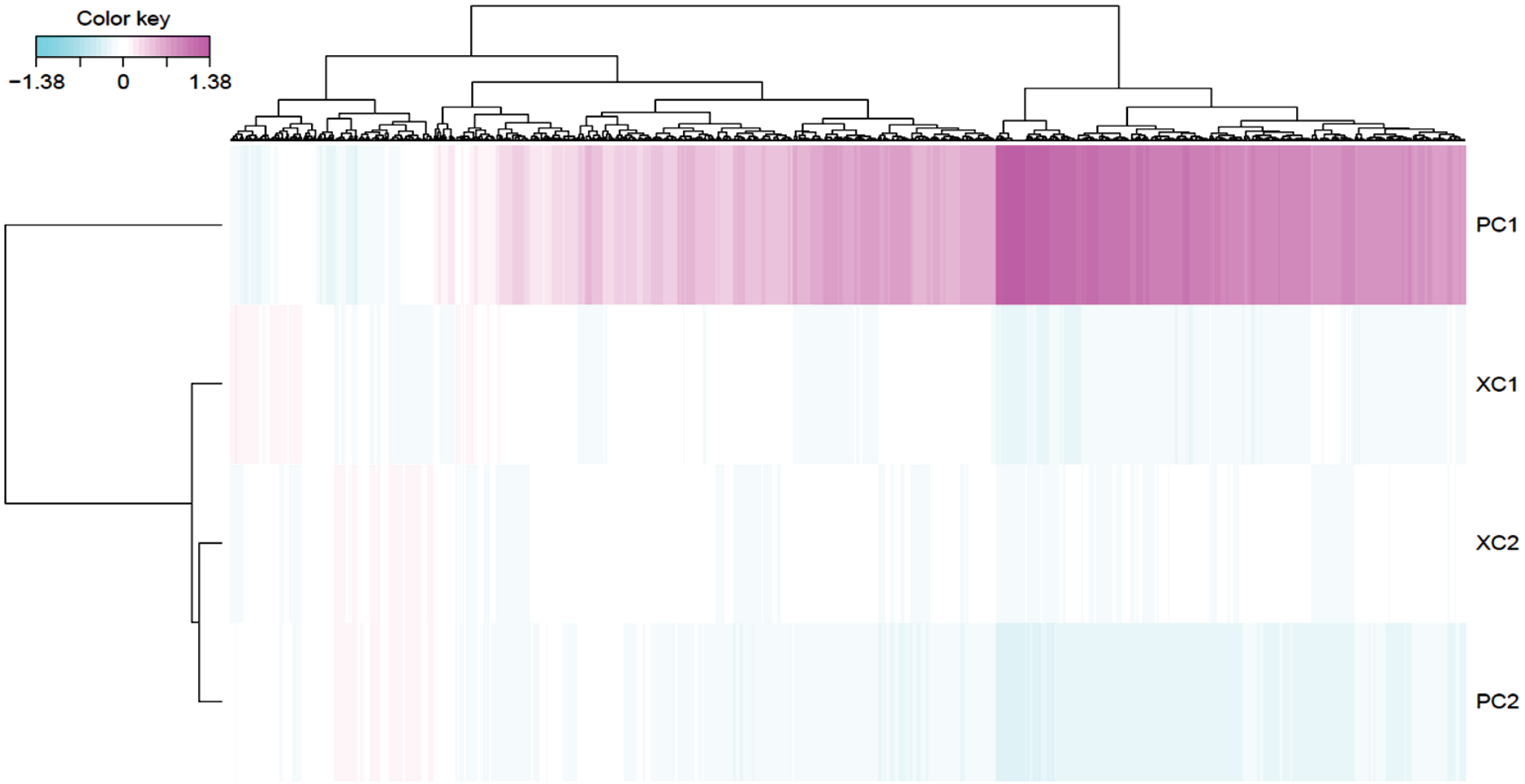


Figure 5.13. CIM analysis of genes significantly up-regulated in *B. hungatei* MB2003. Genes were deemed "significantly expressed" if all three thresholds were met: FDR calculated using the Benjamini-Hochberg procedure *Q*-values < 0.05, KW analysis of variance with $p < 0.05$, and ≥ 2-fold $\log_2$ difference in expression. CIM was constructed using the similarity matrix from hierarchical clustering based on CA of the *B. hungatei* MB2003 dataset. The colour key at the top left indicates expression levels. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture.

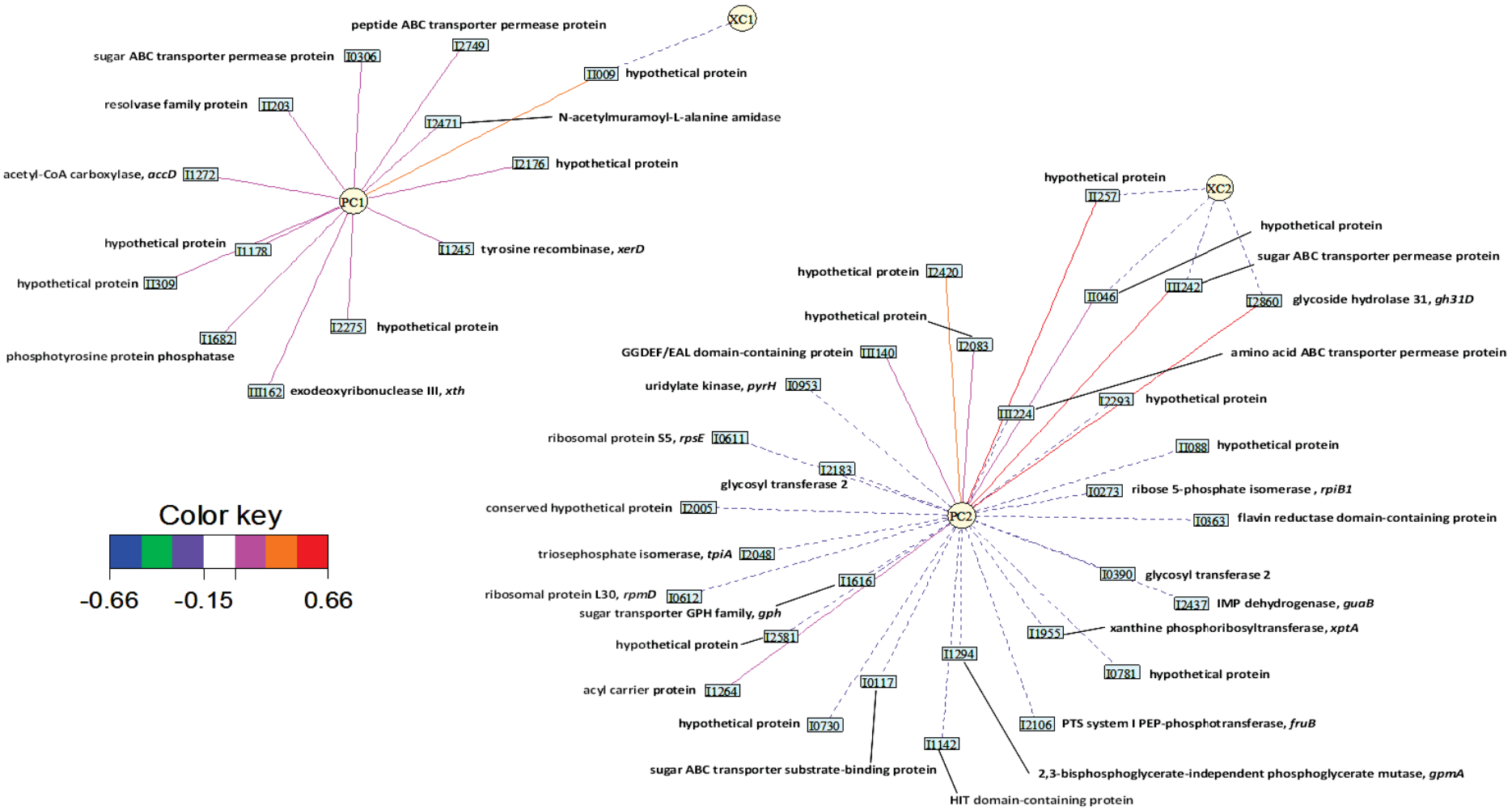


Figure 5.14. Network analysis of genes significantly up-regulated in *B. proteoclasticus* B316. The network was constructed based on CA of the *B. proteoclasticus* B316 dataset with *Q-value* cut-off of < 0.05. The colour key at the top left indicates expression levels and the threshold for expression was set at 0.15. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture.

5.4.3 Biological processes associated with differentially expressed genes

The statistically significant DEGs for B316 and MB2003 under all growth conditions, along with their gene annotations, CAZy families, and COG functional assignments are shown in Appendix III Tables A3.15 to A3.18. During growth on xylan, B316 had 10 genes up-regulated in mono-culture and 7 up-regulated in co-culture conditions (Table A3.15). In contrast, MB2003 cells grown on xylan had 215 genes up-regulated in mono-culture and 92 up-regulated in co-culture conditions (Table A3.16). MB2003 also had 14 genes significantly up-regulated in the xylan mono-culture that were not expressed at all in the co-culture, and 4 genes in the co-culture conditions that were not expressed at all in the mono-culture. During growth on pectin, B316 up-regulated 49 genes in mono-culture while only 7 in co-culture conditions (Table A3.17). There were 26 genes that were only induced in the mono-culture and there were no genes induced only in the co-culture. MB2003 grown on pectin was very different in terms of DEGs; 725 genes were up-regulated in mono-culture and 62 were up-regulated under co-culture (Table A3.18). MB2003 had 107 genes up-regulated in the mono-culture that were not expressed at all in the co-culture, and no genes were expressed exclusively in the co-culture. These gene expression profiles indicate that B316 grown on either substrate, shows relatively little up-regulation of genes when co-cultured with MB2003. In contrast, MB2003 showed poor utilisation of either xylan or pectin for growth.

COG classifications were assigned to 56% and 66% of B316 DEGs from cells grown on xylan and pectin, respectively, while 75% and 63% of MB2003 DEGs from cells grown on xylan and pectin, respectively were assigned to COGs. For MB2003 grown on xylan, the number of co-culture DEGs assigned to the carbohydrate metabolism category (G) was similar to the mono-culture, while the number of co-culture DEGs assigned to the cell motility COG category (N) was three times higher than the mono-culture (Figure 5.15). This suggests that MB2003 up-regulates its motility functions when co-cultured with B316 on xylan. For the pectin grown cultures of MB2003, DEGs belonging to every COG category analysed were up-regulated in the mono-culture to a significant extent compared to the co-culture (Figure 5.16).

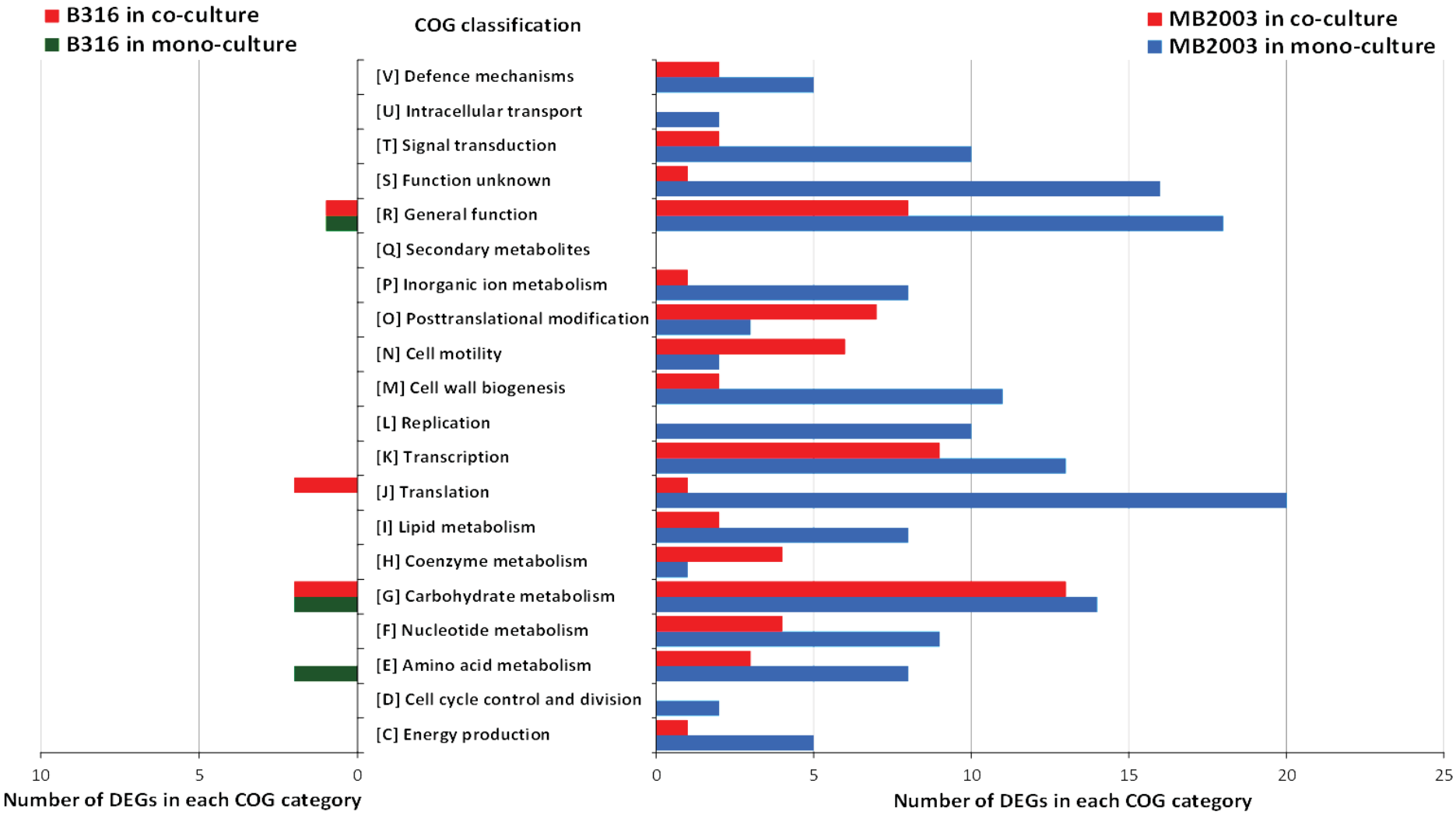


Figure 5.15. COG classifications of DEGs from *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown in mono- and co-culture on xylan. Analysis includes all DEGs with COG classifications.

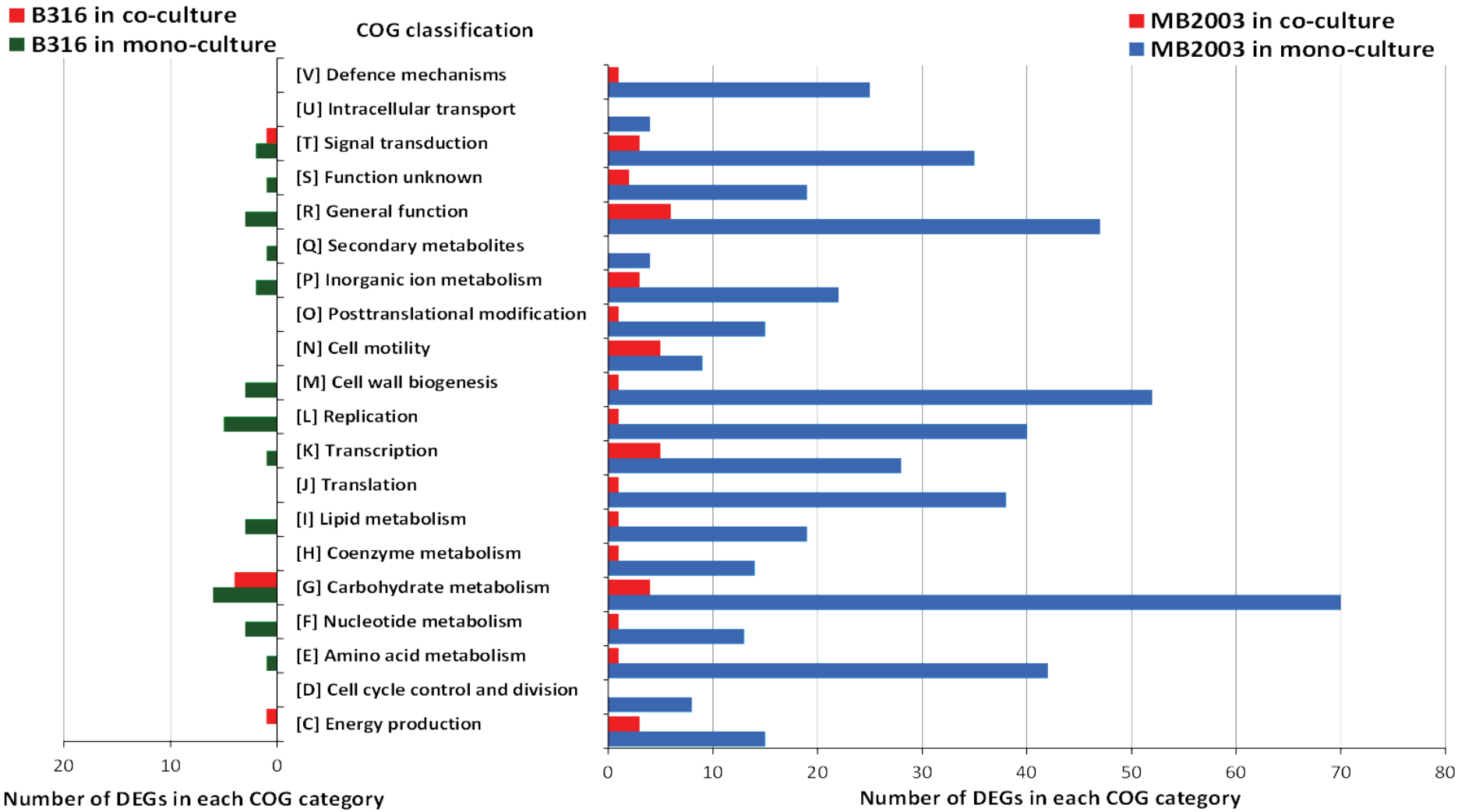


Figure 5.16. COG classifications of DEGs from *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown in mono- and co-culture on pectin. Analysis includes all DEGs with COG classifications.

5.5 Carbohydrate degradation and metabolism

The CAZyme assignments of the B316 and MB2003 DEGs (Appendix III Tables A3.15-A3.18) indicate differences in carbohydrate metabolism. Overall, B316 did not significantly up-regulate many genes that encode CAZymes. For CAZymes belonging to the GH and CE families, B316 had only one gene (*xsa43A*; xylosidase/ arabinofuranosidase) from xylan-containing mono-cultures and one gene (*gh31D*; glycoside hydrolase 31) from pectin-containing co-cultures that were differentially abundant (Figures 5.17A and 5.18A). *Xsa43A* was highly up-regulated at 3.36 $\log_2$-fold change, while the *gh31D* gene (a possible α-glucosidase (EC 3.2.1.20) or α-xylosidase (EC 3.2.1.-)) had a $\log_2$-fold change of 3.64.

In comparison, MB2003 expressed 14 CAZyme genes differentially in the xylan-containing cultures (8 in mono-culture and 6 in co-culture), and 38 in the pectin-containing cultures. Only two of these genes were up-regulated in the co-culture condition (Figures 5.17 and 5.18). The three most up-regulated CAZyme genes in MB2003 xylan-grown mono-culture were *cel9A* (endo-1, 4-β-glucanase, GH9/Pfam00759), *xyn8A* (reducing end xylose-releasing exo-oligoxylanase, GH8/Pfam01270), and *est4C* (polysaccharide deacetylase; CE4/Pfam01522), each with > 3.0 $\log_2$-fold change (Figure 5.17B). The 36 CAZyme genes up-regulated in MB2003 pectin-grown mono-culture varied greatly in their predicted enzymatic capability (Figure 5.18B) and included genes with multiple domains (acetyl-xylan esterase containing a GH2 sugar-binding domain and domain of unknown function DUF303 (Pfam03629)), genes containing two identical catalytic domains within the same gene, (β-glucosidase *bgl3D* with two GH3 domains (Pfam00933)) and the two β-galactosidases *bga2A* and *bga2C* which each contain two GH2 (Pfam00703) domains. Examples of genes with additional non-catalytic domains include *lyc25B* lysozyme (GH25 and two SH3 domains) and *xsa43A,* xylosidase/ arabinofuranosidase (GH43 and CBM6 domains). The polysaccharide lyase, *pl11A* (PL11/COG14111 domains), also known as rhamnogalacturonan lyase (EC 4.2.2.-) or exo-unsaturated rhamnogalacturonan lyase (EC 4.2.2.-), was highly up-regulated in the MB2003 pectin-grown mono-culture with a $\log_2$-fold change of 5.51. In total, MB2003 has 8 genes encoding secreted enzymes involved in pectin breakdown.

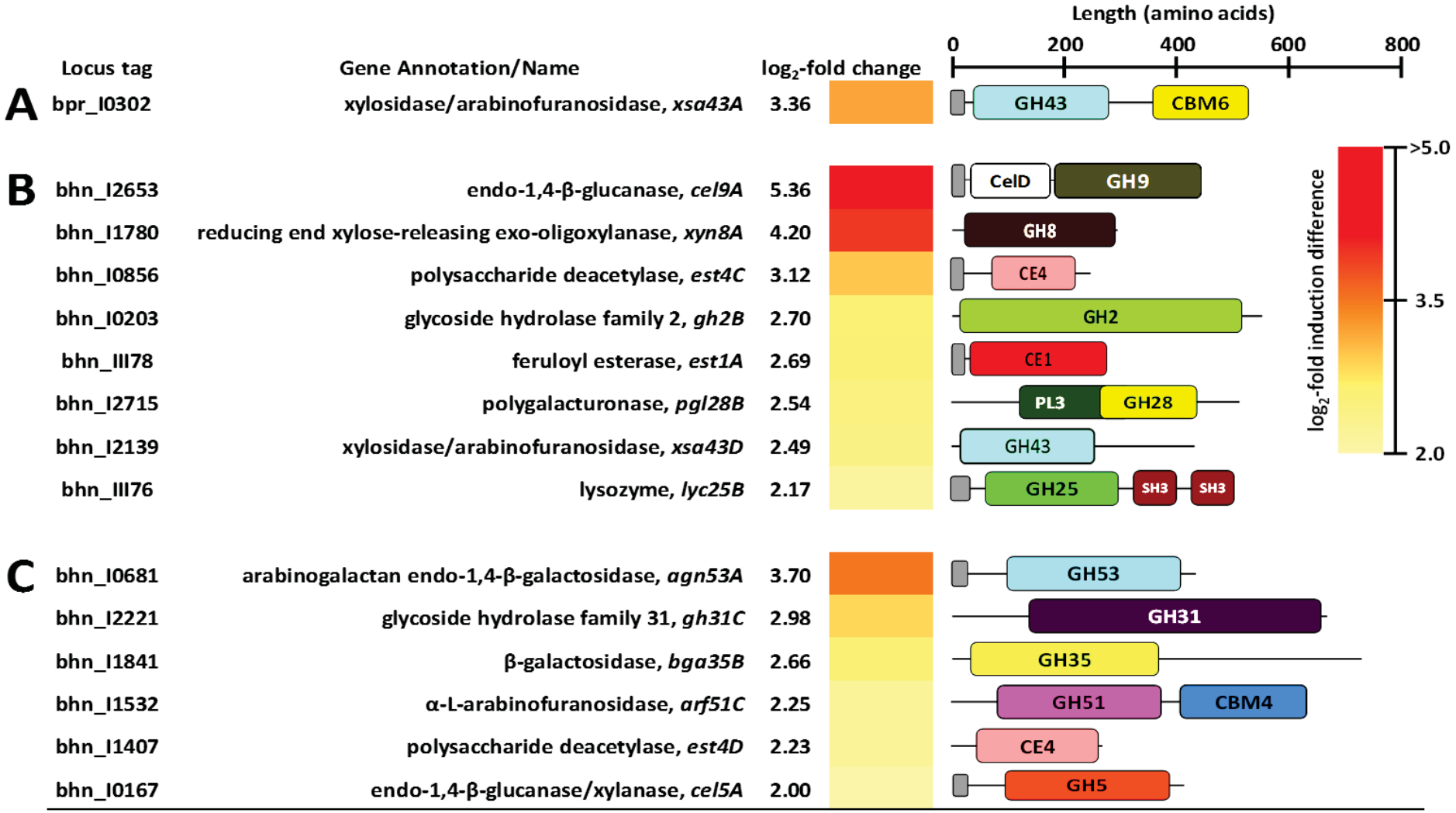
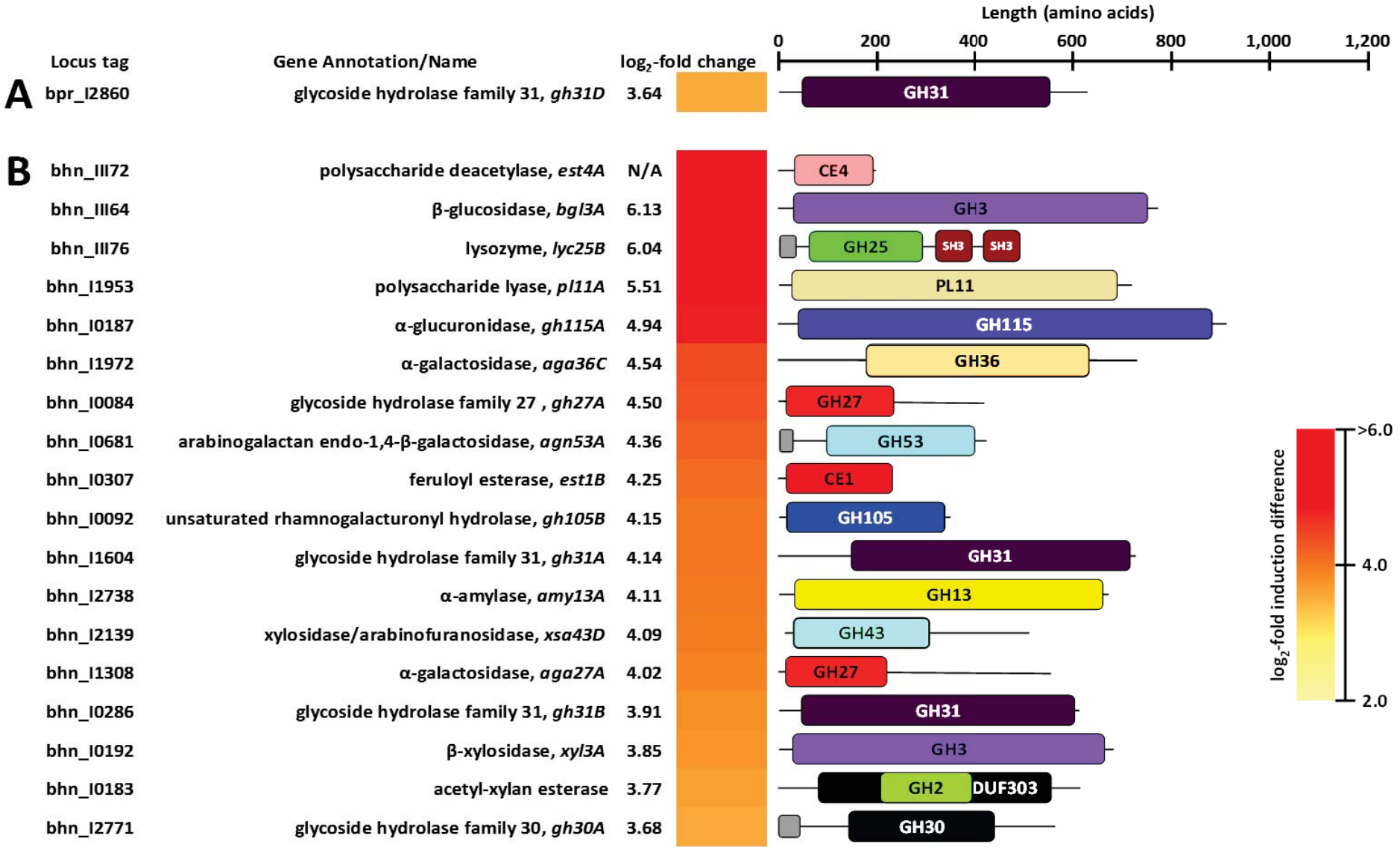


Figure 5.17. CAZyme encoding DEGs up-regulated in xylan-grown cultures. A, B316 grown in mono-culture; B, MB2003 in mono-culture; C, MB2003 grown in co-culture. ▯, signal peptide sequences.


Length (amino acids)
0
200
400
600
800
1,000
1,200
Locus tag
Gene Annotation/Name
log2-fold change
A
bpr_I2860
glycoside hydrolase family 31, gh31D
3.64
GH31
B
bhn_III72
polysaccharide deacetylase, est4A
N/A
CE4
bhn_III64
β-glucosidase, bgl3A
6.13
GH3
bhn_III76
lysozyme, lyc25B
6.04
GH25
SH3
SH3
bhn_I1953
polysaccharide lyase, pl11A
5.51
PL11
bhn_I0187
α-glucuronidase, gh115A
4.94
GH115
bhn_I1972
α-galactosidase, aga36C
4.54
GH36
bhn_I0084
glycoside hydrolase family 27 , gh27A
4.50
GH27
bhn_I0681
arabinogalactan endo-1,4-β-galactosidase, agn53A
4.36
GH53
bhn_I0307
feruloyl esterase, est1B
4.25
CE1
bhn_I0092
unsaturated rhamnogalacturonyl hydrolase, gh105B
4.15
GH105
bhn_I1604
glycoside hydrolase family 31, gh31A
4.14
GH31
bhn_I2738
α-amylase, amy13A
4.11
GH13
bhn_I2139
xylosidase/arabinofuranosidase, xsa43D
4.09
GH43
bhn_I1308
α-galactosidase, aga27A
4.02
GH27
bhn_I0286
glycoside hydrolase family 31, gh31B
3.91
GH31
bhn_I0192
β-xylosidase, xyl3A
3.85
GH3
bhn_I0183
acetyl-xylan esterase
3.77
GH2
DUF303
bhn_I2771
glycoside hydrolase family 30, gh30A
3.68
GH30
log2-fold induction difference
>6.0
4.0
2.0

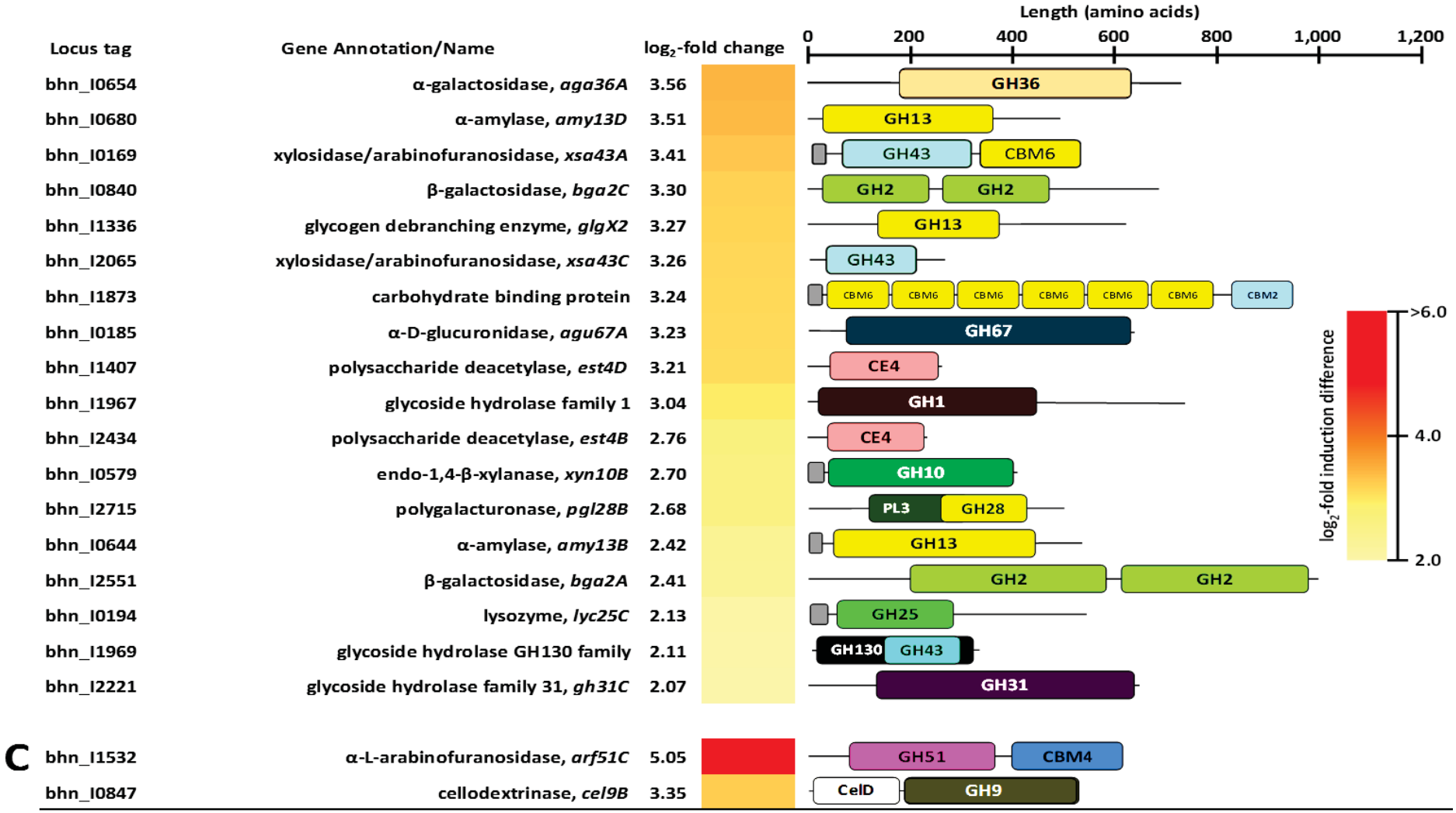
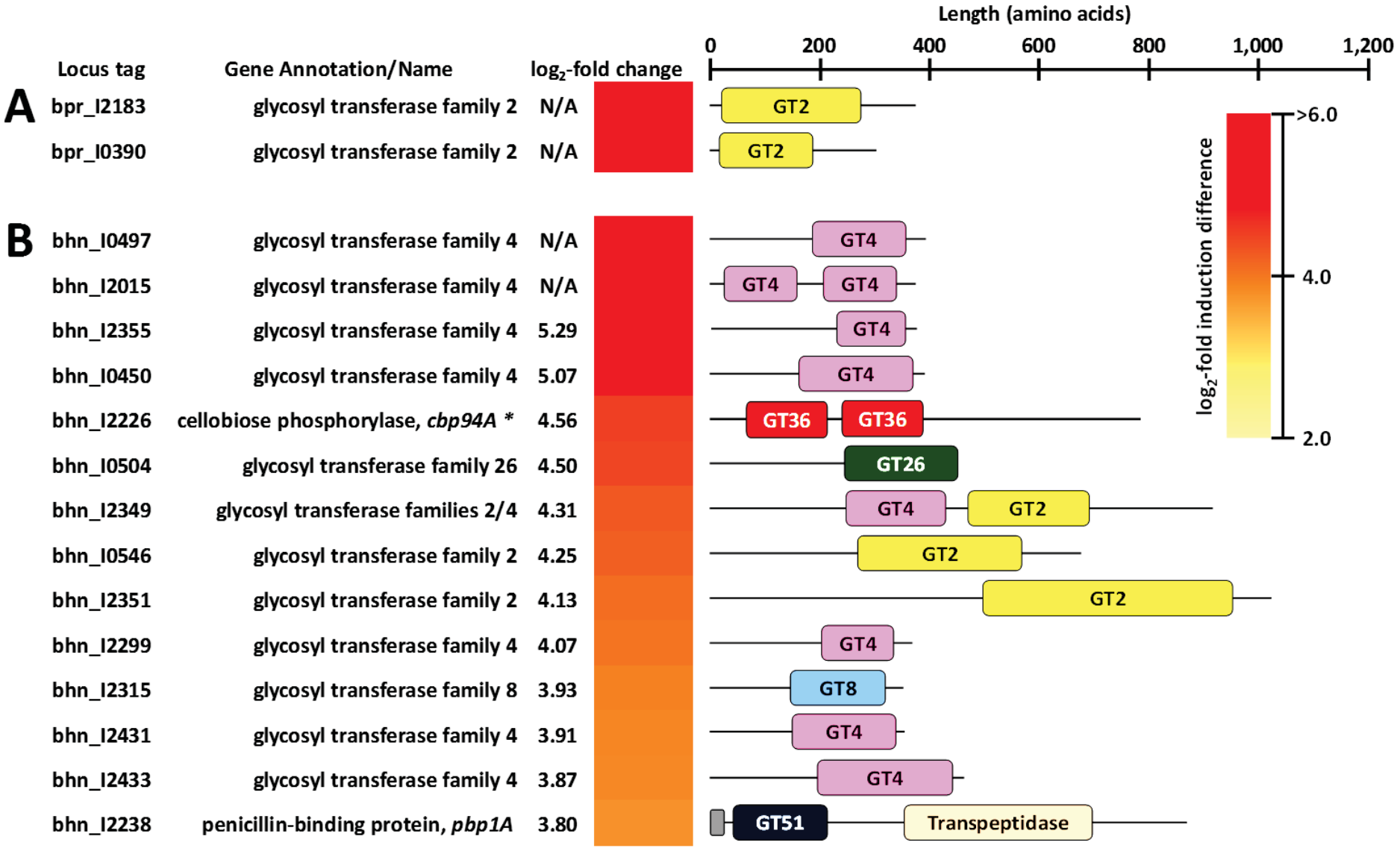


Figure 5.18. CAZyme encoding DEGs up-regulated in pectin-grown cultures. A, B316 grown in co-culture; B, MB2003 in mono-culture; C, MB2003 grown in co-culture. ▯, signal peptide sequences.


Length (amino acids)
0
200
400
600
800
1,000
1,200
Locus tag
Gene Annotation/Name
log2-fold change
A
bpr_I2183
glycosyl transferase family 2
N/A
GT2
bpr_I0390
glycosyl transferase family 2
N/A
GT2
B
bhn_I0497
glycosyl transferase family 4
N/A
GT4
bhn_I2015
glycosyl transferase family 4
N/A
GT4
GT4
bhn_I2355
glycosyl transferase family 4
5.29
GT4
bhn_I0450
glycosyl transferase family 4
5.07
GT4
bhn_I2226
cellobiose phosphorylase, cbp94A *
4.56
GT36
GT36
bhn_I0504
glycosyl transferase family 26
4.50
GT26
bhn_I2349
glycosyl transferase families 2/4
4.31
GT4
GT2
bhn_I0546
glycosyl transferase family 2
4.25
GT2
bhn_I2351
glycosyl transferase family 2
4.13
GT2
bhn_I2299
glycosyl transferase family 4
4.07
GT4
bhn_I2315
glycosyl transferase family 8
3.93
GT8
bhn_I2431
glycosyl transferase family 4
3.91
GT4
bhn_I2433
glycosyl transferase family 4
3.87
GT4
bhn_I2238
penicillin-binding protein, pbp1A
3.80
GT51
Transpeptidase
log2-fold induction difference
>6.0
4.0
2.0

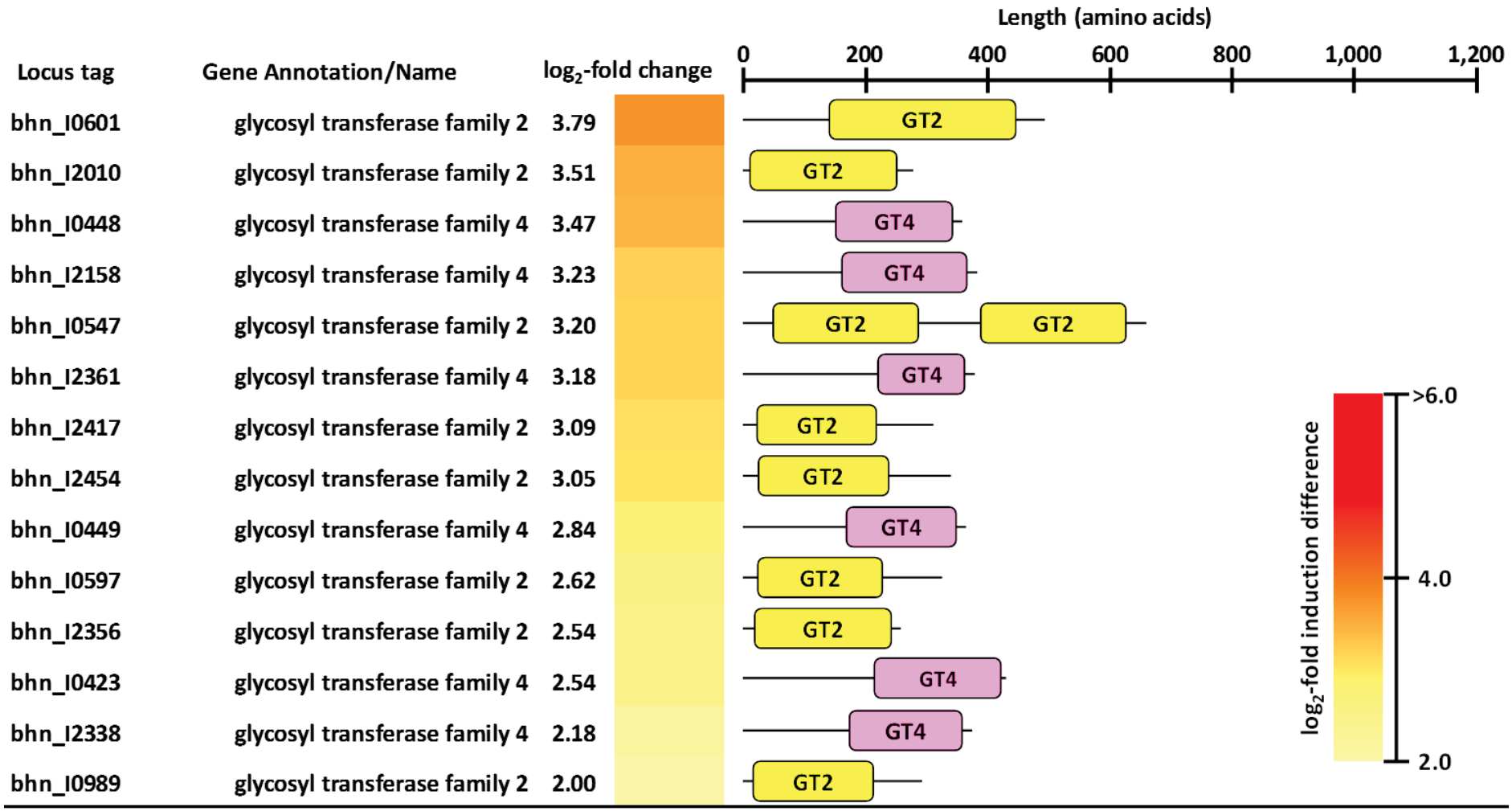


Figure 5.19. GTs up-regulated during mono-culture growth on pectin. A, B316 grown in mono-culture; B, MB2003 in mono-culture. *, Domains for cellobiose phosphorylase were initially designated as GT36, these have recently been reclassified as GH94 hence the designation in the *cbp94A* gene name. ▯, signal peptide sequences.

DEGs encoding GT CAZy functions were only found in the pectin-grown mono-cultures for both B316 and MB2003 (Figure 5.19). Two GT domain-containing genes were found for B316 (Figure 5.19A), compared to 28 in MB2003 (Figure 5.19B). The two GT genes found in B316 both possess a single GT2 (Pfam00535) domain, approximately 400 aa in length. The GTs of MB2003 consist of ten GT2, thirteen GT4, and single genes containing GTs 8, 26, 51, 36 and both GT2 and GT4 domains.

5.6 Oligosaccharide transport and assimilation

Functional annotations of DEGs assigned to the "Carbohydrate transport and metabolism" COG category [G] were investigated in order to determine which carbohydrates are taken up by B316 and MB2003 cultures grown on xylan and pectin. The analysis revealed a substantial number of ABC transporters up-regulated in MB2003 grown on xylan or pectin in both co- and mono-cultures. Analysis of the DEGs associated with ABC system carbohydrate transport and their organisation in relation to the surrounding genes encoding polysaccharide-degrading enzymes in B316 and MB2003 grown on xylan and pectin, are shown in Figure 5.20.

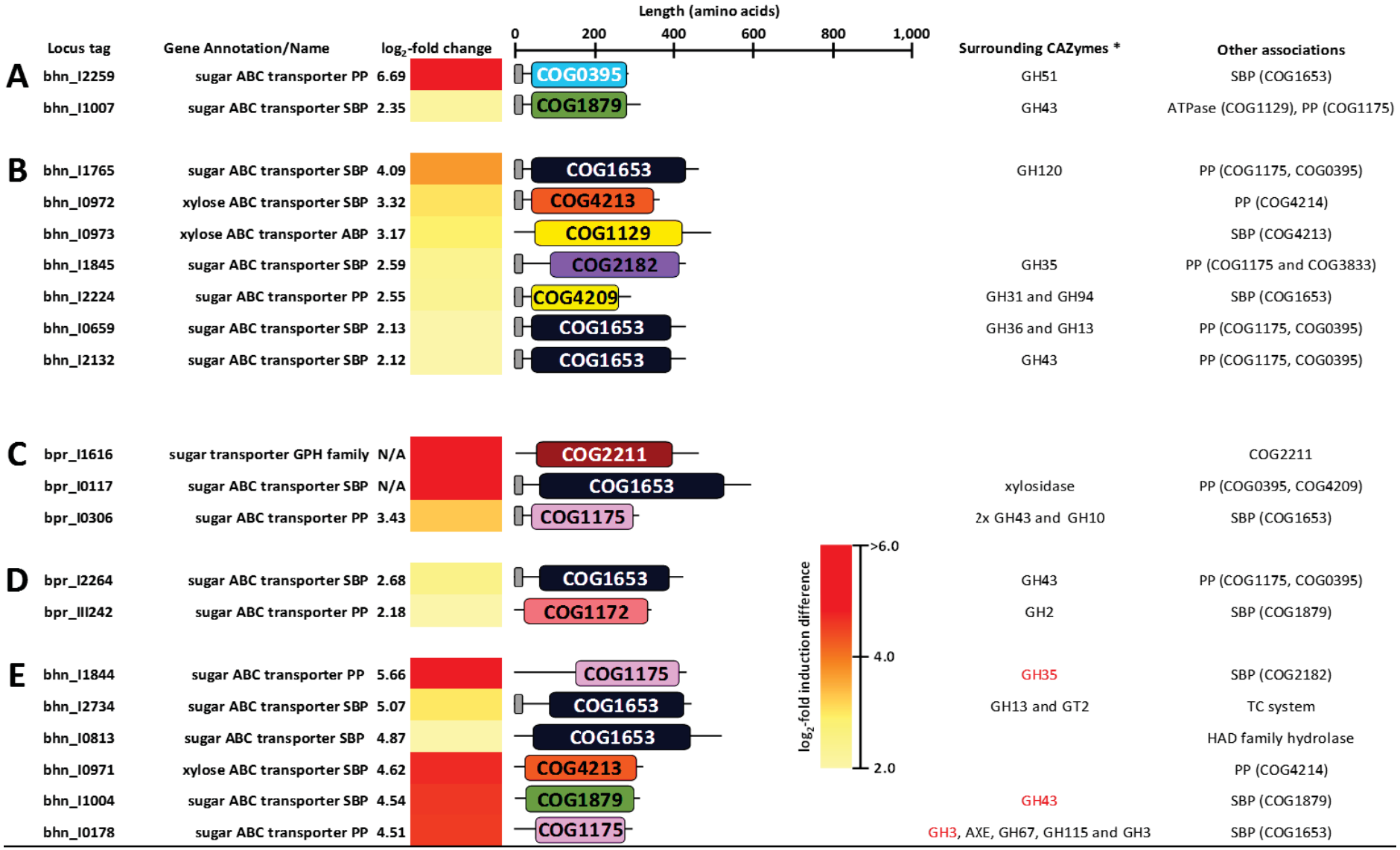
Analysis of the sugar ABC transport systems COG classifications identified predominantly SBPs and PPs that contained a carbohydrate uptake transporter type 1 (CUT1) or a CUT2 domain. The CUT1 domains were represented by COG1653 (SBPs), and COG0395, COG1175 and COG4209 (PPs). The CUT2 domains were represented by COG1879 and COG4213 (SBPs), as well as COG1175 and COG4214 (PPs).

While numerous DEGs associated with ABC system carbohydrate transport were found in MB2003 grown on xylan or pectin (Figure 5.20A, B, E and F), SBPs and PPs in B316, were identified in the pectin growth condition only. In total, five DEGs with predicted functions as sugar ABC transporters were found for B316 grown on pectin; three genes in the mono-culture, and two genes in the co-culture condition (Figure 5.20C and D). The polysaccharide-degrading enzymes found adjacent to these SBPs include a number of xylosidases/arabinofuranosidases (GH43), endo-1,4-β-xylanase (GH10), and a β-mannosidase (GH2).

MB2003 cells grown on xylan and pectin up-regulated 9 and 17 genes respectively, associated with sugar ABC transport systems and target substrates. Examples of the surrounding polysaccharide-degrading enzymes found in MB2003 xylan grown cells include: α-L-arabinofuranosidase (GH51), xylosidases/arabinofuranosidases (GH120, GH30 and GH43), α- and β-galactosidases (GH35 and GH36), sucrose phosphorylase

(GH13), and glycosidases (GH31) (Figure 5.20A and B). MB2003 grown in co-culture on xylan and mono-culture on pectin up-regulated bhn_I0972 and bhn_I0973, and bhn_I0971 genes respectively, associated with xylose transport (Figure 5.20B and E).

A variety of polysaccharide-degrading enzymes have been identified proximal to the genes up-regulated in pectin-grown MB2003 cultures and examples include: xylosidases/ arabinofuranosidases (GH43, GH51), endo-1,4-β-xylanase (GH10), acetyl-xylan esterase, endo-1,4-β-glucanase/xylanase (GH5), β-glucosidase (GH31), α-D-glucuronidase (GH67), α-glucuronidase (GH115), β-xylosidase (GH30), pectin esterase (CE12) and rhamnogalacturonan lyases (PL11) (Figure 5.20E). Also, a total of nine GT2 genes were also identified proximal to the SBPs. The CAZymes surrounding the various SBPs were highly up-regulated and many of these were only expressed in the mono-culture MB2003 cells (Table A3.18). In contrast, only one gene (bhn_I2222) was significantly up-regulated in co-culture, in proximity to a glucosidase (GH31) and cellobiose phosphorylase (GH94) (Figure 5.20F). Overall, SBPs and PPs with the COG1653 were the most abundant domains identified in MB2003 pectin grown cells.


Length (amino acids)
0
200
400
600
800
1,000
Locus tag
Gene Annotation/Name
log2-fold change
Surrounding CAZymes *
Other associations
A
bhn_I2259 sugar ABC transporter PP 6.69 COG0395 GH51 SBP (COG1653)
bhn_I1007 sugar ABC transporter SBP 2.35 COG1879 GH43 ATPase (COG1129), PP (COG1175)
B
bhn_I1765 sugar ABC transporter SBP 4.09 COG1653 GH120 PP (COG1175, COG0395)
bhn_I0972 xylose ABC transporter SBP 3.32 COG4213 PP (COG4214)
bhn_I0973 xylose ABC transporter ABP 3.17 COG1129 SBP (COG4213)
bhn_I1845 sugar ABC transporter SBP 2.59 COG2182 GH35 PP (COG1175 and COG3833)
bhn_I2224 sugar ABC transporter PP 2.55 COG4209 GH31 and GH94 SBP (COG1653)
bhn_I0659 sugar ABC transporter SBP 2.13 COG1653 GH36 and GH13 PP (COG1175, COG0395)
bhn_I2132 sugar ABC transporter SBP 2.12 COG1653 GH43 PP (COG1175, COG0395)
C
bpr_I1616 sugar transporter GPH family N/A COG2211 COG2211
bpr_I0117 sugar ABC transporter SBP N/A COG1653 xylosidase PP (COG0395, COG4209)
bpr_I0306 sugar ABC transporter PP 3.43 COG1175 2x GH43 and GH10 SBP (COG1653)
D
bpr_I2264 sugar ABC transporter SBP 2.68 COG1653 GH43 PP (COG1175, COG0395)
bpr_III242 sugar ABC transporter PP 2.18 COG1172 GH2 SBP (COG1879)
E
bhn_I1844 sugar ABC transporter PP 5.66 COG1175 GH35 SBP (COG2182)
bhn_I2734 sugar ABC transporter SBP 5.07 COG1653 GH13 and GT2 TC system
bhn_I0813 sugar ABC transporter SBP 4.87 COG1653 HAD family hydrolase
bhn_I0971 xylose ABC transporter SBP 4.62 COG4213 PP (COG4214)
bhn_I1004 sugar ABC transporter SBP 4.54 COG1879 GH43 SBP (COG1879)
bhn_I0178 sugar ABC transporter PP 4.51 COG1175 GH3, AXE, GH67, GH115 and GH3 SBP (COG1653)
log2-fold induction difference
>6.0
4.0
2.0

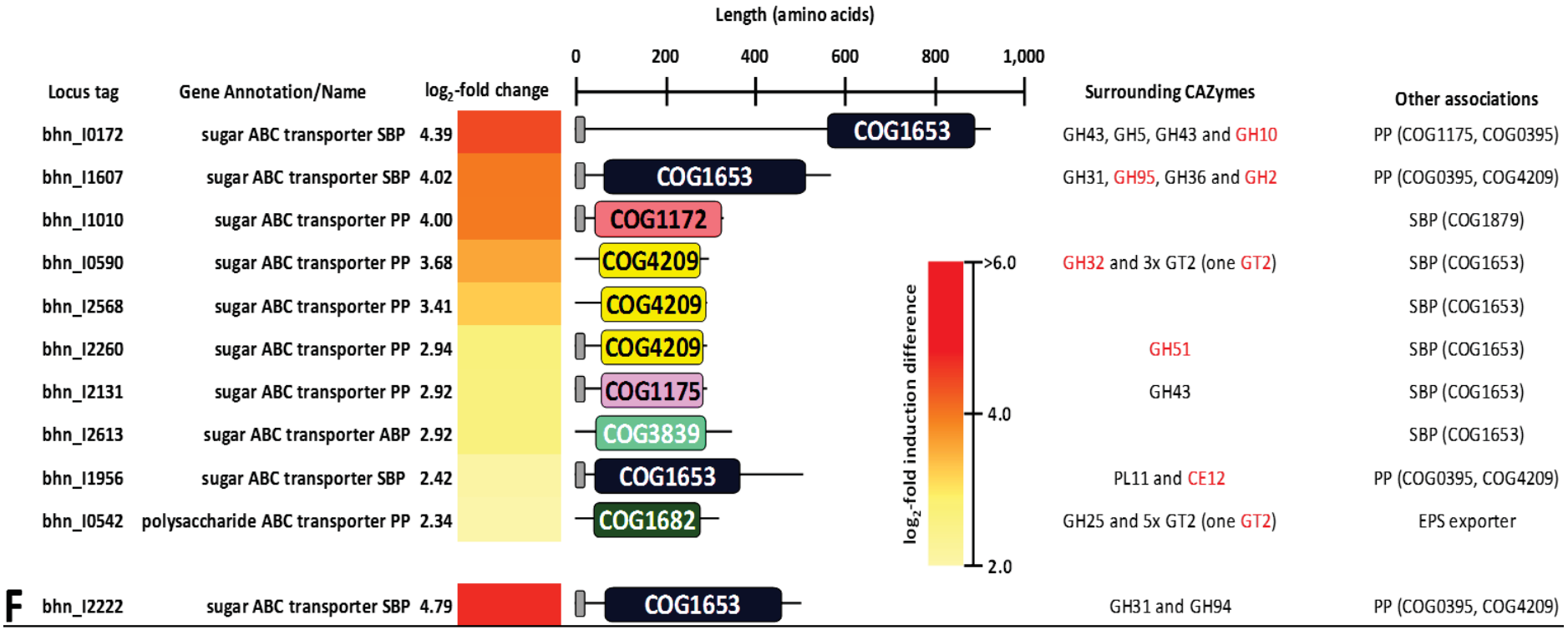


Figure 5.20. Functional domains of DEGs encoding carbohydrate transport proteins and surrounding CAZymes identified in xylan- or pectin-grown cultures. A, MB2003 mono-culture on xylan; B, MB2003 co-culture on xylan; C, B316 mono-culture on pectin; D, B316 co-culture on pectin; E, MB2003 mono-culture on pectin; F, MB2003 co-culture on pectin. *, CAZyme families of genes and DEGs (red) encoding polysaccharide-degrading enzymes co-localised with carbohydrate transport genes. Abbreviations: PP, permease protein; AXE, acetyl-xylan esterase; ABP, ATP-binding protein; EPS, exopolysaccharide; TC system, two component system histidine kinase. COG designations: COG0395, COG1175, COG1653 and COG1879, periplasmic component of ABC-type sugar transport system; COG1682, ABC-type polysaccharide/polyol phosphate export systems; COG2182, maltose-binding periplasmic proteins/domains; COG3833, permease component of ABC-type maltose transport systems; COG2211, Na^+/melibiose symporter and related transporters; COG4213, periplasmic component of ABC-type xylose transport system; COG1129 and COG3839, ATPase component of ABC-type sugar transport systems; COG4209, permease component of ABC-type polysaccharide transport system; COG1172, permease component of ribose/xylose/arabinose/galactoside ABC-type transport systems. ▯, signal peptide sequences.

5.7 Discussion

The analysis of relative growth, metabolites and transcriptomes examined the relationship between *B. hungatei* MB2003 and *B. proteoclasticus* B316 in the terms of insoluble substrate utilisation, when grown in co-culture on xylan and pectin. The analysis of cell counts for each strain using qPCR from mono-culture samples confirmed the initial carbon source utilisation characterisation of *B. proteoclasticus* B316 as a xylan and pectin degrader, while *B. hungatei* MB2003 was unable to utilise either substrate to support significant growth (Figure 5.4). However, co-cultures of B316 with MB2003 revealed a commensal relationship in which the non-xylanolytic/-pectinolytic MB2003 showed growth when either xylan or pectin were supplied as substrates. Commensalism is defined as an association between two different species which results in a benefit to one species, while the other species is not significantly affected. The relationship between MB2003 and B316 is clearly a commensalistic interaction and not classical symbiosis. This is because the number of B316 cells in co-cultures with MB2003 was lower compared to mono-cultures at the late log phase of growth. Hence the effect of MB2003 on the growth of B316 was negative.

Xylan and pectin are the most abundant plant structural polysaccharides after cellulose, and are major sources of energy for microbial fermentation within ruminants. The xylan derived from oat-spelts used in this experiment, is largely composed of GAX which is comprised of a xylose monomer backbone (Borneman *et al.*, 1990). The pectin used for this experiment was derived from apple and is largely comprised of RG-I and XGA, with backbones composed of galacturonic acid and rhamnose monomers (Cosgrove, 2005). Rumen bacteria degrade xylans and pectins to predominantly xylose, arabinose, rhamnose, galactose and a variety of higher oligosaccharides (Howard, 1957; Dehority, 1973; Chesson and Forsberg, 1997). For this reason, the release of arabinose, fucose, galactose, glucose, rhamnose and xylose monosaccharides for all conditions was analysed for 2 h time increments of growth, using ion chromatography (IC) (Figures 5.6 and 5.7).

For the xylan grown samples, the IC results showed that fermentation in the B316 mono- and co-culture samples was complete by late log phase, at 12 h post-inoculation, after which the levels of xylose decreased dramatically (Figure 5.6A). The complete absence of released monosaccharides in MB2003 samples supports the VFA and pH experimental evidence for its inability to utilise xylan in mono-culture (Figure 5.6B). However, the total pool of xylose detected and the reduced xylose concentration at 12 h

post-inoculation, was not as dramatic in the co-culture samples as in the B316 mono-culture (Figure 5.6C). It is proposed that this was due to the presence of MB2003 cross-feeding on the released xylose via a commensal interaction. Furthermore, it is possible that the other sugars analysed (or not detected) could have been released and used immediately (or at time points not analysed) such that their concentrations in the samples did not appear to increase.

In pectin-supplemented mono- and co-cultures, galactose was the predominant monosaccharide detected. B316 grown in mono-culture contained the highest concentrations (Figure 5.7) which is consistent with B316 being more efficient at breaking down pectin, in contrast to the very poor, if any, ability of MB2003 to utilise this substrate (Figure 3.12). However, the rate of galactose utilisation after the early log phase was similar between the mono-cultures and co-culture, suggesting that MB2003 could compete with B316 for the released galactose. Also, it is possible that other mono-saccharides not detected in this analysis, such as galacturonic acid, may have been released during hydrolysis of pectin, that B316 may have utilised while MB2003 could not. Furthermore, in a previous study on B316 grown on pectin (Villas-Bôas *et al.*, 2006), an increase in, rhamnose, arabinose and other mono- and disaccharides was observed, but not galacturonic acid or galactose. It is proposed that the galacturonic acid released from hydrolysis of pectin may have interfered with the galactose measurements in the inoculum and any galacturonic acid formed was rapidly consumed.

The growth and metabolite observations were complemented by transcriptome analysis. This analysis has shown that significant gene up-regulation in B316 was confined to genes required for substrate utilisation (Figures 5.15 and 5.16). In contrast, MB2003 cells grown on xylan and particularly on pectin in mono-culture, significantly up-regulated every biological process in an apparent attempt to grow. MB2003 grown in co-culture on xylan and pectin, up-regulated genes for almost every biological process but to lesser extent than in the monocultures. It appears that this is due to the availability of released mono- and oligo-saccharides, from the primary degradation of xylan and pectin by the co-culture partner, B316.

Previous genome sequence information (Kelly *et al.*, 2010) and proteome analyses (Dunne *et al.*, 2011; Bond *et al.*, 2012) led us to predict that *B. proteoclasticus* B316 primarily attacks the xylan backbone and main substituent groups of hemicellulose outside the cell via secreted enzymes, and the variable length substituted or un-substituted xylo-oligosaccharides are then transported into the cell where the final stages of

degradation are performed. It appears that B316 applies the same approach for the degradation of pectin, but the backbone is composed of galacturonic acid. *B. hungatei* MB2003 does not initiate the breakdown of the backbones of either xylan or pectin, but appears to act as a scavenger of sugars released from the insoluble substrates through the primary degradation activity of B316. MB2003 appears to produce the enzymatic machinery required to utilise the oligosaccharides released from the initial degradation of xylan and pectin by B316. The transcriptome analysis of the CAZymes encoded by *B. hungatei* MB2003 and *B. proteoclasticus* B316 supports the proposed interactions between these two rumen bacteria.

The presumption that B316 and MB2003 cannot completely degrade complex GAX, XGA and RG-I extracellularly to monosaccharides, implies that cells must transport a variety of substituted xylan and pectin oligomers across the cell wall. The degradation of these oligomers into their constituent monomers is achieved through the activity of several classes of cytosolic enzymes, including β-xylosidases, α-galactosidases and α-glucuronidases that contain GH3, GH27, GH115 and GH67 CAZy domains. MB2003 cells grown on pectin in mono-culture, significantly expressed intracellular genes encoding enzymes containing GH3, GH27, GH115 and GH67 CAZy domains (Figure 5.18B).

Overall, analysis of up-regulated SBPs and their functional domains, along with surrounding genes encoding polysaccharide-degrading enzymes, revealed that B316 did not display any preference for a particular substrate in xylan grown cells (Figure 5.20C and D), as there were no SBP genes up-regulated. Furthermore, among all expressed ABC transport system SBP genes identified, the most prevalent functional category was the COG1653 domain that is known to be associated with oligosaccharide transport (Bond *et al.*, 2012). In pectin grown co-cultures, B316 up-regulated the bpr_III242 gene, predicted to encode a sugar ABC transporter permease protein which binds ribose, xylose, arabinose and galactose, suggesting a preference for these substrates. In contrast, the substantial up-regulation by MB2003 of sugar ABC transport system genes and a large variety of co-localised genes encoding polysaccharide-degrading enzymes (Figure 5.20), supports the findings in this thesis that MB2003 is able to grow only in co-culture on xylan and pectin through cross-feeding on the released oligosaccharides such as xylose, arabinose and rhamnose. Transcriptome analysis suggests that the ABC-mediated transport of polysaccharide-derived soluble sugars, such as xylo- and pectic-

oligosaccharides, is an important component for xylan and pectin degradation and utilisation systems employed by B316 and MB2003.

In summary, the growth experiments showed that MB2003 was unable to utilise xylan and had little ability to utilise pectin when grown in mono-culture, but was capable of significant growth on both substrates only when co-cultured with B316. MB2003 appears to be able to sustain itself in co-culture until primary degradation of xylan and pectin by B316 is underway, allowing for the uptake of released soluble carbohydrates. The MB2003 transcriptome analysis revealed the deliberate and specific up-regulation of primary CAZymes and sugar ABC transporters focussed on the uptake of available substrates when co-cultured with B316. This is therefore a commensalistic interaction between MB2003 and B316, in which the former benefits from the supply of released monosaccharides made available due to the degradative activities of the latter, while B316 itself remains relatively unaffected.

Chapter 6
General discussion, conclusions and future directions

6.1 Introduction

Butyrivibrio and *Pseudobutyrivibrio* form a significant group of rumen bacteria and understanding their role in the rumen microbiome and their enzymatic activities involved in plant polysaccharide breakdown is important in maximising digestion and productivity in ruminants. The aim of this thesis was to explore four specific research questions:

i) How does the composition and organisation of the genes encoding polysaccharide-degrading enzymes vary within the genomes of *Butyrivibrio* and *Pseudobutyrivibrio* species?
ii) Is there conservation of genes that might imply particular importance in polysaccharide utilisation in these genera?
iii) Are there interspecies competitive or collaborative strategies among *Butyrivibrio* strains?
iv) How are these different strategies encoded by the respective genomes?

This thesis presents the first comprehensive genomic characterisation of the genera *Butyrivibrio* and *Pseudobutyrivibrio* in order to compare their gene complements, assess their enzymatic capabilities and understand how their members coexist, collaborate and/or compete in the rumen environment.

6.2 Differentiation of the *Butyrivibrio* and *Pseudobutyrivibrio* genera

Until 1996, all *Butyrivibrio* and *Pseudobutyrivibrio* strains were assigned to a single species, *Butyrivibrio fibrisolvens,* due to their phenotypic and metabolic similarities (Bryant and Small, 1956a). Subsequent 16*S* rRNA gene sequence analyses of 48 *Butyrivibrio* strains indicated that rumen-derived *B. fibrisolvens* strains clearly fell into two phylogenetic groups (Forster *et al.*, 1997), however division into two genera could not be supported due to lack of phenotypic data (Willems *et al.*, 1996). Additional work led to the establishment of a new genus named *Pseudobutyrivibrio* (Van Gylswyk *et al.*, 1996). Prior to the start of this thesis, rumen *Butyrivibrio* and *Pseudobutyrivibrio* were divided into six species represented by *B. fibrisolvens*, *B. hungatei*, *B. proteoclasticus* and *B. crossotus*, and the *Pseudobutyrivibrio* species, *P. xylanivorans* and *P. ruminis* (Figure 1.6), all of which belong to the genetically diverse Lachnospiraceae family, order Clostridiales (Moon *et al.*, 2008). The *B. crossotus* species is represented by the type strain *B. crossotus* NCFB 2416 obtained from human faeces (Moore *et al.*, 1976; Willems *et al.*, 1996), and thus has not been investigated in this thesis due to its phenotypic and genomic differences compared to the ruminal *B. fibrisolvens* strains.

This thesis reports the phenotypic and genotypic analysis of 30 *Butyrivibrio* strains which were grouped into three main species clusters based on analysis of full-length 16*S* rRNA gene sequences (Figure 3.1). Cluster 1 contained the sequences of the type strains of *B. proteoclasticus* (B316) and *B. hungatei* (JK615) and ten *Butyrivibrio* species 3 strains, Cluster 2 contained twelve *Butyrivibrio* species 2 strains with no type strains, while Cluster 3 consisted of eight strains containing the *B. fibrisolvens* type strain (D1) and closely related strains. Along with the 30 *Butyrivibrio* strains that were phenotypically characterised (Chapter 3), genome sequences of an additional 10 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains were included in the comparative genomics analyses (Table 4.1).

FGD genome clustering and ANI analysis were used to examine whole-genome sequence similarities for the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* strains to estimate genomic diversity among this group and to infer genomic relationships between groupings. Clustering microbes based on genome comparisons provides a more detailed depiction of their evolutionary history (Konstantinidis and Tiedje, 2007; Altermann, 2012; Kim *et al.*, 2014), and will undoubtedly contribute towards more accurate *Butyrivibrio* and *Pseudobutyrivibrio* taxonomies in the future. The clustering comparisons of the *Butyrivibrio* and *Pseudobutyrivibrio* genomes based on Pfams, COGs, TIGRfams and KO protein/functional family types (Figures 4.4 and A2.3), confirmed the initial 16*S* rRNA gene-based species grouping and genome similarity comparisons. Despite the distinct demarcation between these two genera, the boundaries between species within each genus were less distinct.

The genome-wide analyses concluded that the *B. fibrisolvens*, *Butyrivibrio* sp. 2, *Butyrivibrio* sp. 3 and *Pseudobutyrivibrio* cluster groups were phylogenetically separate from one another and functionally distinct at the genome level. Within each of the three *Butyrivibrio* cluster groups, strains displayed differences and some formed clearly separate groupings: i) *Butyrivibrio* sp. MC2013 and *Butyrivibrio* sp. NC3005 from the *B. fibrisolvens* group, ii) *Butyrivibrio* sp. AE3003, *Butyrivibrio.* sp. AE2005, *Butyrivibrio* sp. LB2008 and *B. hungatei* NK4A153 from *Butyrivibrio* sp. 3 (Figure 4.3). Based on the differences found in the genome-wide analyses, *B. hungatei* NK4A153 does not constitute a true *B. hungatei*, and likely represents a new species of *Butyrivibrio*, therefore disagreeing with the 16*S* rRNA gene analysis.

6.3 *Butyrivibrio* and *Pseudobutyrivibrio* genome features

PFGE analyses of genomic DNAs digested with *Apa*I and I-*Ceu*I (Appendix I Figures A1.2-A1.6) produced unique banding patterns for all of the *Butyrivibrio* strains analysed, providing evidence for differences at the genomic DNA level between these organisms. The genome size estimates from the RE digests range from approximately 3.5 Mb to 5.6 Mb (Appendix I Table A1.1), and these size estimations were within ± 500 Kb of the actual genome sequences for all *Butyrivibrio* strains analysed. Overall, the average genome size of the 46 *Butyrivibrio* and *Pseudobutyrivibrio* genomes was 4.14 Mb. An alignment of the draft *Butyrivibrio* genomes with the first completely sequenced genome, *B. proteoclasticus* B316 (Appendix II Figures A2.4-A2.6), revealed that only *Butyrivibrio* sp. XBB1001, *Butyrivibrio* sp. VCB2006 and *B. proteoclasticus* FD2007 from *Butyrivibrio* sp. 3 group have significant synteny to B316. These three strains displayed high similarity to B316 in the FGD, ANI and protein family analyses and therefore should be reclassified as *B. proteoclasticus*.

Early studies reported the presence of large extrachromosomal elements in a number of *Butyrivibrio* strains (Teather, 1982), and recently megaplasmids and chromids were described for *B. proteoclasticus* B316 and *B. hungatei* MB2003 (Kelly *et al.*, 2010; Palevich, 2011; Yeoman *et al.*, 2011). PFGE analyses of undigested genomic DNAs identified the average size range of extrachromosomal elements to be between 300 and 500 Kb, with the largest being 869.2 Kb (from *Butyrivibrio* sp. XPD2002) and the smallest being 99.1 Kb (from *Butyrivibrio* sp. AE3006; Appendix I Table A1.1). Comparisons of the draft genomes in regard to the sizes of extrachromosomal elements identified that two of the six *Butyrivibrio* sp. AE3004 contigs match the 433.1 and 350.1 Kb bands observed. Only the *B. proteoclasticus* FD2007 genome has a secondary chromosome similar to that from *B. proteoclasticus* B316. Overall, the patterns from the PFGE analysis indicate that extrachromosomal elements are a common genomic characteristic of *Butyrivibrio* and *Pseudobutyrivibrio*.

Megaplasmids in *Butyrivibrio* species are predicted to encode genes important for survival in the rumen, but no genes have been identified that contribute to the breakdown of lignocellulose (Yeoman *et al.*, 2011) on these episomes. Chromids are replicons that have %G+C content similar to that of their main chromosome, but have plasmid type maintenance and replication systems. They are smaller than the chromosome and usually (but not obligatory) larger than any other plasmids present. Chromids contain genes essential for growth and maintenance of the organism along with several core genus-

specific genes that can be found on the chromosome in other species of bacteria (Harrison *et al.*, 2010). The reasons for the existence of large extrachromosomal replicons such as megaplasmids and chromids in the *Butyrivibrio* genus remain unclear.

Rumen bacteria live in a diverse community of microorganisms that together fulfil the complex interactions required for rumen function, including fibre degradation. A recent metagenomic study on the bovine rumen "plasmidome" has emphasised the essential evolutionary and cooperative roles between plasmids and their host bacteria (Kav *et al.*, 2012). Due to incidence of related plasmids in phylogenetically distant bacteria coupled with ability to be horizontally transferred by conjugation, plasmids are suggested to play a role as a communication channel for the horizontal exchange of genomic information conveying advantageous functions between the rumen microbes. The types of traits that are transferred by plasmids include those implicated in amino acid, protein and carbohydrate metabolism, which are essential in the rumen ecosystem and make the metabolic burden of sustaining the plasmid worthwhile for the host (Jain *et al.*, 2003). The transfer of genes encoding degradative systems (Selinger *et al.*, 1996), exopolysaccharide production (Cheng and Costerton, 1980), bacteriocin production (Kalmokoff *et al.*, 1996; Kalmokoff and Teather, 1997) and resistance to antibiotics (Flint *et al.*, 1988; Shoemaker *et al.*, 1992), may provide a competitive advantage within the rumen microbial ecosystem.

Extrachromosomal elements such as plasmids and megaplasmids also increase the number of genes encoded in the bacterial genome through gene dosing effects and allow for faster replication of the host genome thus increasing the growth rate of the bacterial cell (Morrison, 1996). The above mentioned plasmidome study also identified phylogenetic associations of various plasmid genes across taxonomic levels up to the phylum level (Kav *et al.*, 2012). This suggests that in *Butyrivibrio*, extrachromosomal elements serve as vehicles for the exchange of genomic information between different strains and species, and potentially to other genera such as *Pseudobutyrivibrio*.

6.4 Cell motility and operons encoding flagella genes

Previous analysis of the flagella synthesis/assembly operons in the non-motile *B. proteoclasticus* B316 (Kelly et al, 2010) and *B. hungatei* MB2003, and the motile *P. xylanivorans* MA3014 (Palevich, 2011) revealed sequence variations in the region encoding key flagella genes. The non-motile B316 and MB2003 contained distinct additional genes within the main flagellar operon, between genes *flgE* and *motA*. In contrast, no gene insertions were found in the intergenic space between *flgE* and *motA* genes in the genome of the motile MA3014. Based on these findings, an hypothesis was formulated that the additional genes inserted between *flgE* and *motA* disrupted flagellar function and were correlated with a non-motile phenotype (Palevich, 2011). Only *B. fibrisolvens* MC2013 and *Butyrivibrio* sp. 2 FC2001 displayed a motile phenotype in the agar stab assay (Table 3.4). However the gene organisation hypothesised to correlate with motility, varied among the *Butyrivibrio-Pseudobutyrivibrio* assemblage. All *B. fibrisolvens* strains, and many of the *Butyrivibrio* sp. 3 strains did not have gene insertions, yet only MC2013 was motile. Conversely, *Butyrivibrio* sp. 2 FC2001 contained an insert, but nevertheless was motile. These findings refute the original hypothesis that gene insertions between *flgE* and *motA* genes are correlated with inactivation of flagella genes leading to a non-motile phenotype. The cause of the lack of motility in *Butyrivibrio* likely depends on other, unidentified mutations in the flagella operons or mutations in genes that regulate expression of this operon or potentially a lack of induction of these genes under the growth conditions used in the motility assay.

The performance of a bacterium in a motility assay does not necessarily imply the absence of a flagellum. It is known that while *B. proteoclasticus* B316 and *B. hungatei* MB2003 are non-motile, they have flagella (Attwood *et al.*, 1996; Palevich, 2011). It is therefore possible that flagella are present in most non-motile strains analysed in this work. The usefulness of a non-functional flagellum is curious. It is possible that *Butyrivibrio* use their flagella for adherence rather than motility, as shown for some archaea (Shimoyama *et al.*, 2009). The ability to adhere to plant material is an essential property of the plant-fibre-degrading bacteria within the rumen, and it is possible that *Butyrivibrio* utilise their flagella as a means of attachment to the insoluble substrates.

Another potential role of flagella in non-motile *Butyrivibrio* is attachment to other *Butyrivibrio* that colonise and degrade plant fibre. This would allow *Butyrivibrio* strains that cannot break down insoluble substrates, to co-locate themselves close to strains that did have this capability, thus positioning them to take advantage of soluble sugars

released from the breakdown of plant material. The *B. hungatei* MB2003 genome contains approximately 40 genes associated with flagella biosynthesis and motor functions, and in the xylan co-culture growth experiment almost all of these genes were up-regulated (Appendix III Table A3.13). MB2003 may thus use its flagellum for attachment to *B. proteoclasticus* B316, thereby facilitating the commensal interaction observed between these two bacteria. The mechanisms involved in bacterial adherence to plant material and other microorganisms in the rumen are not fully understood and further investigation is required to resolve the potential role of flagella in *Butyrivibrio* adherence.

6.5 Plant fibre degradation by *Butyrivibrio* and *Pseudobutyrivibrio*

The fibrolytic rumen microbiome consists of both cellulolytic and non-cellulolytic bacteria (Koike *et al.*, 2003b), and it is generally accepted that the bacterial community responsible for plant fibre degradation within the rumen is diverse and that intimate associations among morphologically and phylogentically different, but physiologically complementary bacterial species form to facilitate this process (McAllister *et al.*, 1994). *Butyrivibrio* and *Pseudobutyrivibrio* are common bacteria in the rumen (Henderson *et al.*, 2015), and are among a small number of rumen genera capable of efficiently utilising the insoluble, complex polysaccharides xylan and pectin (Bryant and Small, 1956a; Hungate, 1966; Van Gylswyk *et al.*, 1996). To understand the metabolic capabilities that enable *Butyrivibrio* to degrade and utilise these complex polysaccharides, the carbon sources used by these bacteria were determined in this study.

Previous work on substrate utilisation by *Butyrivibrio* (Kopecny *et al.*, 2003; Moon *et al.*, 2008), found that glucose, cellobiose, maltose and esculin are common substrates, while cellulose was not used. *Butyrivibrio* characteristically ferment carbohydrates to butyrate, formate, lactate and acetate. Growth tests and VFA production analyses in this thesis support the reported carbon-source utilisation findings and the distinct ability of *Butyrivibrio* and *Pseudobutyrivibrio* to utilise a wide variety of soluble sugars and insoluble polysaccharides. This is particularly true of *B. fibrisolvens* and *Butyrivibrio* sp. 3 strains that were able to grow on pectin and xylan (Table 3.5 and Figure 3.12). The VFA data also showed that *Butyrivibrio* utilisation of polysaccharides and their ability to assimilate the degradation products of these polysaccharides was variable with clear differences identified between species groups based on the phylogenetic and pan-genome analyses. For example, the *B. fibrisolvens* MD2001, AB2020, FE2007, ND3005, YRB2005 and WTE3004 strains displayed similar growth patterns, in particular their

ability to utilise pectin,xylan and xylose, whereas *B. fibrisolvens* sp. MC2013 and NC3005 expressed different growth patterns that coincide with their phylogenetic and genomic divergence within this cluster (Figure 4.3). All *Butyrivibrio* sp. 2 strains could utilise xylose and strains AE3004, XPD2002 and VCD2006 displayed similar growth patterns on all insoluble substrates analysed that is also reflected by their genome similarities. The *Butyrivibrio* sp. 3 strains B316, VCB2006, XBB1001 and AE3009 that grouped closely at the genome and phylogenetic levels displayed similar growth patterns across all insoluble substrates analysed, and curiously only AE2015 could utilise xylose.

Butyrivibrio are metabolically versatile species of ruminal bacteria and are predicted to encode extensive xylanolytic and pectinolytic activities (Kelly *et al.*, 2010). In this thesis, the polysaccharide-degrading capabilities of the *Butyrivibrio* and *Pseudobutyrivibrio* genomes were predicted based on CAZyme analysis, where a total of 163 CAZyme families were identified, consisting of 64 GHs, 14 CEs, 7 PLs, 38 CBMs and 36 GTs (Figure 4.8). Previous analysis of the *B. proteoclasticus* B316 genome described that two-thirds of its CAZymes involved in polysaccharide breakdown were predicted to be intracellular, and only a few CBMs were present (Kelly *et al.*, 2010). The analysis of the Pfam domains from the most abundant GH families (GH2, GH31 and most GH3, GH13 and GH43) showed they did not contain signal sequences and hence were predicted to be located intracellularly (Figure 4.13). Similarly, CAZymes with predicted roles in xylan and pectin degradation, the GH8, GH28, GH39, GH51, GH67, GH88, GH105, GH115, CE2 and CE10 families were also predicted to be intracellular. These findings suggest that a variety of complex oligosaccharides resulting from extracellular hydrolysis are metabolised within the cell.

Within *Butyrivibrio* species, strains generally had similar types of CAZy enzymes, but their absolute number of genes within each CAZy category varied considerably. The extensive repertoire of GH, PL and CE domain-containing CAZymes of *B. fibrisolvens* and *Butyrivibrio* sp. 3 supports their phenotypic characterisation as specialist degraders of xylan and pectin when compared to other *Butyrivibrio* and *Pseudobutyrivibrio*. In particular the above mentioned *B. fibrisolvens* strains MD2001, AB2020, YRB2005 and WTE3004 and *Butyrivibrio* sp. 3 strains VCB2006, XBB1001 and AE3009, possess the largest number of CAZymes within their respective species group (Figure 4.9) and also group together based on their relative abundance of CAZymes (Figure 4.11). Among the *Butyrivibrio* sp. 3 group in particular, xylan and pectin utilisation varied greatly, as did the total abundance of CAZymes that comprise the polysaccharide-degrading enzymes of

the 19 strains within this group (Figure 4.9). This suggests that members of *Butyrivibrio* sp. 3 play an important role in polysaccharide degradation in the rumen and are significant energy suppliers to ruminants, through production of VFAs. The growth experiments and comparative glycobiome analyses have shown that strains belonging to this species group are diverse in both the substrates they can utilise and the set of CAZymes they encode and may occupy similar niches in the rumen, which begs the question as to how these microorganisms can coexist in the competitive rumen environment.

At present, there are four recognised species of *Butyrivibrio* and two of *Pseudobutyrivibrio* species (Moon *et al.*, 2008), and interactions between these species are likely to contribute to fibre degradation in the rumen. Previous co-culture studies have identified symbiotic relationships between specific rumen bacteria (Cheng *et al.*, 1991; Chen and Weimer, 2001; Sawanon and Kobayashi, 2006), but as only a small proportion of rumen bacteria can be grown in culture, the majority of potential interactions between rumen bacteria remain unexplored. Competition and cooperation interactions in co-culture between different rumen bacteria have previously been reported, where *S. ruminantium* cross-feeding of xylo-oligosaccharides produced from the initial xylan degradation by *B. fibrisolvens* actually resulted in a slight decrease in the utilisation of xylan (Cotta and Zeltwanger, 1995). Miron and colleagues have previously demonstrated complementary activities of the xylanolytic *B. fibrisolvens* with the cellulolytic *Ruminococcus* species and *F. succinogenes* in the adhesion and digestion of plant cell walls (Miron and Ben-Ghedalia, 1992; Miron and Ben-Ghedalia, 1993a; Miron and Ben-Ghedalia, 1993b; Miron and Ben-Ghedalia, 1993c; Miron *et al.*, 1994). The hypothesis that certain *Butyrivibrio* strains are unable to initiate significant degradation of the insoluble polysaccharides alone and rely on other *Butyrivibrio* to initiate the degradation process was investigated.

The genomes of two closely related members of *Butyrivibrio* sp. 3, *B. proteoclasticus* B316 and *B. hungatei* MB2003 were previously compared in detail to examine this question (Palevich, 2011). The B316 genome encodes a much more numerous and diverse range of polysaccharide-degrading enzymes compared to MB2003. It was therefore hypothesised that B316 would be better equipped to grow on insoluble, hemicellulosic substrates in comparison to MB2003. However, on the basis of its CAZyme complement, MB2003 appeared to encode the enzymatic machinery required for xylan and pectin degradation. In Chapter 5, the interactions between two *Butyrivibrio* sp. 3 strains, *B. proteoclasticus* B316 and *B. hungatei* MB2003, and mechanisms by

which these bacteria work in concert to degrade these different forms of polysaccharides were explored. To investigate these interactions, the two strains were grown on xylan and pectin, either alone in mono-cultures, or in direct competition in co-cultures. However, despite its genetic makeup, the results demonstrated that MB2003 actually was unable to utilise xylan and displayed very poor ability to utilise pectin alone. MB2003 was, however, capable of significant growth on both substrates when co-cultured with B316. This indicates a commensalistic interaction between these species in which B316 initiates the primary attack on the insoluble substrates, while MB2003 has a secondary role, competing for the released soluble sugars.

From the co-culture experiments, the increase in the total cell count (via qPCR) of the culture was correlated with VFA production (acetate, butyrate and formate) and a reduction of the culture pH (Table 5.1 and Figure 5.5). The lack of VFA production or change in pH lead to the conclusion that MB2003 was not able to utilise xylan at all and showed minimal growth when supplemented with pectin in mono-culture. These observations were confirmed by species-specific quantitative PCR analysis (Figure 5.4). It is interesting that MB2003 displayed similar growth to B316 up until the 16 h time period when co-cultured with either xylan or pectin supplied as substrates. In contrast, a decrease in B316 growth was identified in co-culture compared to the mono-culture samples, particularly in pectin-supplemented medium that was supported by a decrease in formate and acetate production as well as the pH measurements of the 12 h samples (Figure 5.4B and 5.5B).

The net difference between production and utilisation of xylan and pectin during mono- and co-culture growth of MB2003 and B316 was detected using ion chromatography (IC) to quantify the released monosaccharides (Figures 5.6 and 5.7). For B316 and co-culture cells grown on xylan, the most abundant sugar detected was xylose. In contrast, pectin grown cultures galactose was the most abundant sugar in all mono- and co-culture samples, with small amounts of arabinose and rhamnose also detected. All the mono- and co-culture samples showed similar dynamics where galactose release occurred from 6 h to 8 h (Figure 5.7), except that B316 mono-culture samples produced approximately three times the amount of galactose compared to MB2003 mono- and co-culture samples. Furthermore, rhamnose, a large proportion of the RG-I backbone of pectin, was only detected in small amounts in B316 mono- and co-culture samples. It is proposed that hydrolysis of pectin, as opposed to xylan, releases a greater variety of soluble sugars suitable for MB2003 growth, as it is forced to scavenge on the released

monosaccharides that are made available through secondary hydrolysis of the side chains of these complex molecules. This is supported by the release of predominantly galactose in pectin-supplemented MB2003 cultures, where a significant increase in growth was observed when co-cultured with B316 (Figure 5.7). These results support the proposed hypothesis that MB2003 is unable to initiate significant degradation of the insoluble substrates xylan or pectin alone. MB2003 therefore relies on B316 to initiate the breakdown of xylan and pectin in co-culture and then cross-feeds on the released soluble substrates in a commensal interaction.

In the co-cultures grown with xylan or pectin, free arabinose was not detected, and in pectin-grown co-cultures rhamnose and xylose were not detected during the 16 h of incubation (Figures 5.6 and 5.7). Previous work has shown that rumen bacteria degrade xylans and pectins to predominantly xylose, arabinose, rhamnose, galactose and a variety of higher oligosaccharides (Howard, 1957; Dehority, 1973; Chesson and Forsberg, 1997). Cotta and Zeltwanger (1995) demonstrated that the rapid hydrolysis of oat-spelt xylan by enzymes derived from *B. fibrisolvens* H17c (recently described as belonging to the genus *Pseudobutyrivibrio*) caused the aggregation of extracellular soluble, low molecular weight xylo-oligosaccharides, composed of xylose, xylobiose and xylotriose, and low amounts of glucuronic acids. The ability of a *B. fibrisolvens* to degrade and utilise different forms of xylan also showed similar findings (Hespell and Cotta, 1995). Following 12 h growth on xylan they identified that arabinose was still present in the cultures in the form of residual undegraded xylan and larger xylan degradation products or xylo-oligosaccharides (Hespell and Cotta, 1995). It is likely that a similar explanation applies to pectin-grown cultures in the present study, where arabinose, rhamnose and xylose were not detected as they were present in the form of galacturonans and pectic-oligosaccharides following initial hydrolysis of pectin. Although the IC method used was able to detect some of the soluble monosaccharide intermediates, the accumulation of the majority of monosaccharides analysed was not observed possibly due to a large turnover of these sugars. Also, the chromatography method used was not sufficient for analysis of the more complex xylo- and pectic-oligosaccharides and is a point for future work. The absence of free monosaccharides observed in both mono- and co-culture implies that, following the initial extracellular degradation of xylan and pectin, B316 and MB2003 must transport the resulting large xylo- and pectic-oligosaccharides into the cytoplasm via specific carbohydrate transport systems, where the side groups are debranched and the constituent monosaccharides can be metabolised.

6.6 Proposed models for degradation of xylan and pectin by *Butyrivibrio*

The predominant form of xylan found in ruminant forage feed is GAX (Ishii, 1997; Kulkarni *et al.*, 1999; Cosgrove, 2005; Ebringerová *et al.*, 2005; Scheller and Ulvskov, 2010), which is a large component of oat-spelt derived xylan used in this study (Figure 6.1A). Pectic polysaccharides form a significant portion of forage and polygalacturonans such as xylogalacturonan and rhamnogalacturonan are the main pectic components of alfalfa and lucerne (Hatfield, 1989; Hatfield and Weimer, 1995; Hatfield *et al.*, 1999a; Saulnier and Thibault, 1999; Nakamura *et al.*, 2002; Jung and Lamb, 2003). The two main components of apple derived pectin are XGA and RG-I (Figure 6.1B), that are largely composed of galacturonic acid linked at the *O*-1 and the *O*-4 positions with neutral sugar side chains such as arabinans, galactans or arabinogalactans attached to most rhamnose residues at the *O*-4 position (Caffall and Mohnen, 2009; Yapo, 2011).

A model for the fibrolytic potential of *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown on xylan and pectin is proposed, based on total transcript abundances and significantly up-regulated genes encoding CAZymes (Appendix III Tables A3.12-A3.13 and Figures 5.17-5.19), and from previous analysis of the intra- and extra-cellular proteome of B316 (Dunne, 2010; Dunne *et al.*, 2011; Bond *et al.*, 2012; Dunne *et al.*, 2015). The model predicts GAX, XGA and RG-I degradation occurs outside the cell, followed by carbohydrate uptake and breakdown of internalised xylo- and pectic-oligosaccharides (Figures 6.2 and 6.3).

In GAX, xylose makes up the backbone of the polymer and is also found in the various side chains of both GAX and XGA, whereas galactose is a large component of RG-I side chains, thus explaining the appearance of xylose and galactose as important monosaccharides released during xylan and pectin degradation. These findings support the use of GAX, XGA and RG-I structures as the basis for the proposed utilisation model, as these monosaccharides are important components of the backbones and are also found in various side chains of these polymers. Efficient hydrolysis of xylan and pectin within the rumen requires a variety of poly-saccharolytic enzymes, including endo-xylanases, β-xylosidases, α-D-glucuronidases α-L-arabinofuranosidases, acetyl-xylan esterases, feruloyl and cinnamoyl esterases, polysaccharide deacetylases, and polygalacturonases among others, which work in concert to breakdown these complex polysaccharides. Liberation of GAX, XGA and RG-I from the surrounding polymers also requires numerous non-polysaccharidases, identified in the co-culture experiment.

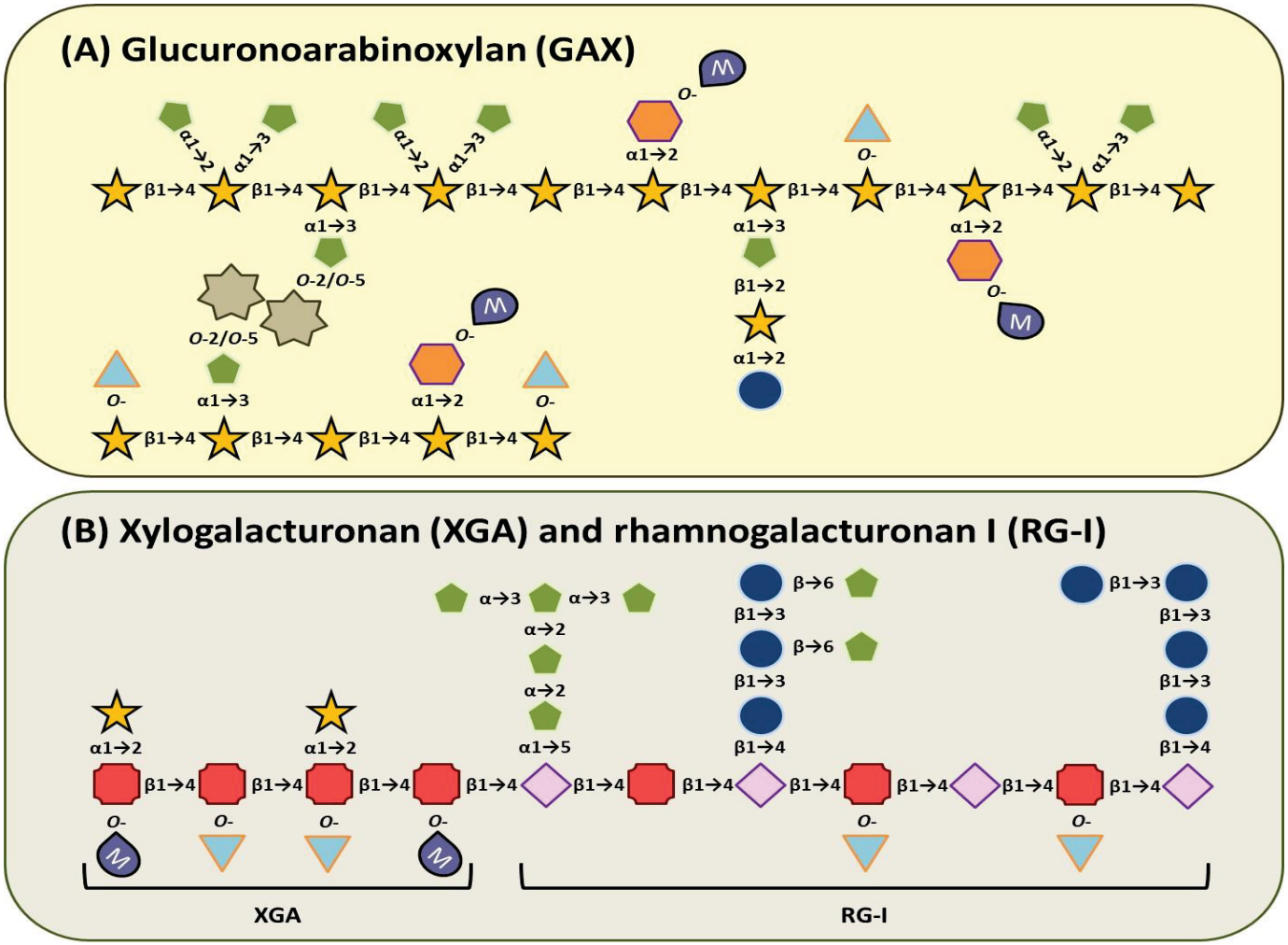


Figure 6.1. Schematic diagrams of the structures of the main classes of xylan and pectin. The xylan (A) and pectin (B) structures shown are not quantitatively accurate. Symbols: ★, *D*-xylose; △, acetyl group; ◒, methyl group; ⬟, *L*-arabinose; ●, *D*-galactose; ⬢, *D*-galacturonic acid; ⬢, *D*-glucuronic acid; ✸, ferulic acid; ◇, *L*-rhamnose.

An important initial observation is that the functional diversity within the complements of xylan- and pectin-degrading enzymes expressed by *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown on xylan and pectin, reflects the structural complexity of these polysaccharides.

6.6.1 Proposed model for GAX degradation

The proposed model for GAX degradation based on the fibrolytic potential of *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown on xylan is shown in Figure 6.2, and describes the initial extracellular degradation of GAX, followed by transport of the xylo-oligosaccharides into the cytoplasm where the internalised xylo-oligosaccharides are broken-down and metabolised.

6.6.1.1 Extracellular GAX degradation

Previous microarray analysis of the B316 genes expressed during growth on xylan or xylose showed up-regulation of several secreted endoxylanases, an endocellulase and a feruloyl esterase in xylan grown cells (Kong, 2007). The proposed model is thus feasible as the xylanases secreted by B316 have been shown to produce oligosaccharides extracellularly from xylan (Dunne *et al.*, 2011). The secreted enzymes necessary for degradation of GAX, along with the removal of other substituent groups such as acetyl, methyl and ferulic acid groups, contain GH10, GH13, GH30, GH43, GH53, CE1, CE4 and CE8 CAZy domains that target (1,4)-β-D-linkages within the xylan backbone.

B316 grown in co-culture on xylan did not up-regulate any genes encoding CAZymes, yet it grew better than MB2003 in mono-culture and outcompeted MB2003 in co-culture. The conditions of the co-cultures were such that gene expression in B316 was unaffected by the presence of MB2003, but the transcription level of many CAZymes, including *xyn10D* (bpr_I1083), *xyn10E* (bpr_I1740), and *cel5D* (bpr_I0728) genes that encode endo-1,4-β-xylanase and endo-1,4-β-glucanase activities, respectively, as well as sugar ABC transport system genes were elevated in B316 (Table A3.12). The observed up-regulation of *xsa43A* (bpr_I0302) in B316 mono-cultures grown on xylan (Figure 5.17A), signifies that this is an important enzyme for xylan degradation. MB2003 significantly expressed secreted enzymes containing CE1 and CE4, as well as GH53, GH10 and GH5 CAZy domain containing genes in mono- and co-culture xylan-grown cells, respectively (Figure 5.17B and C).

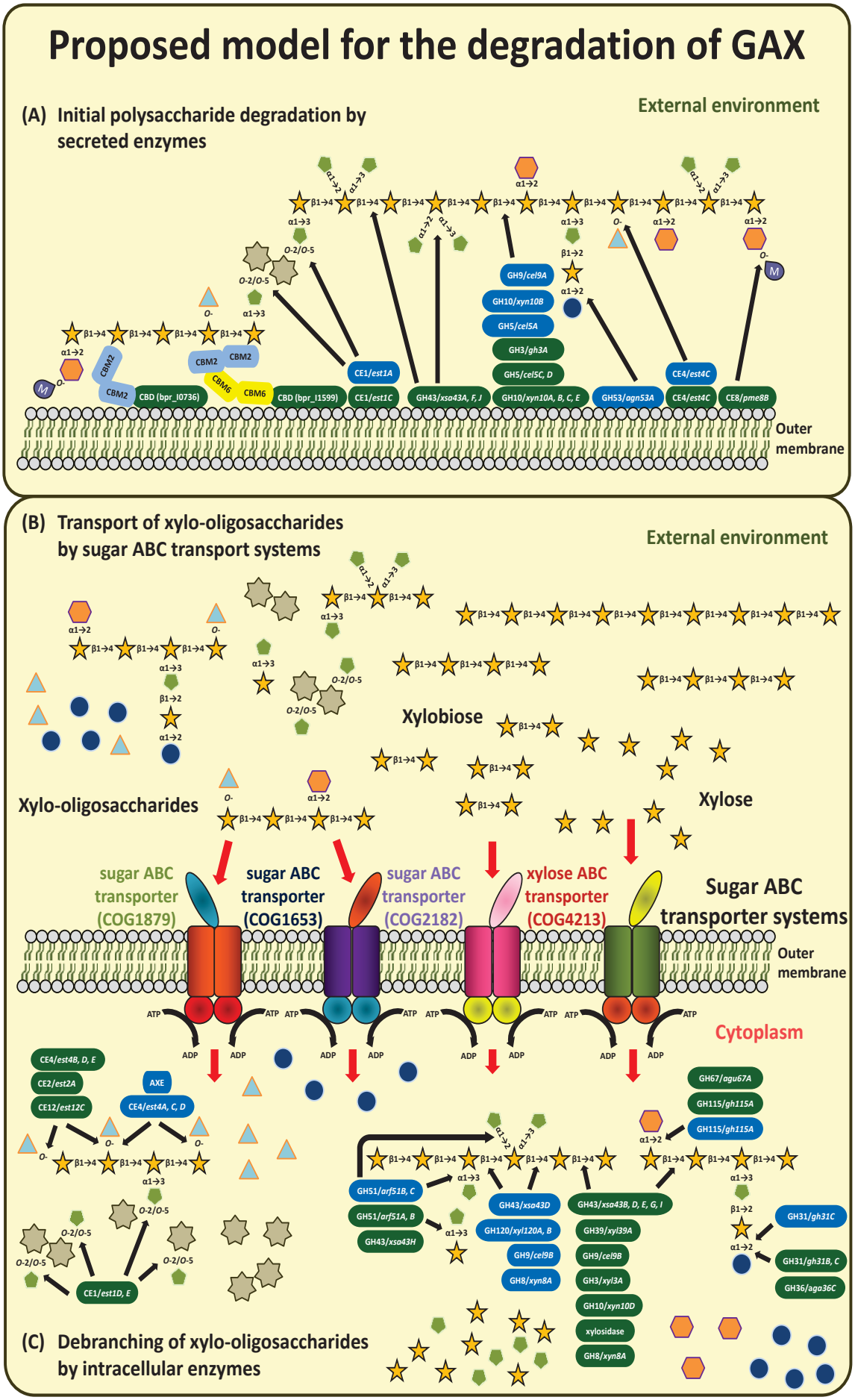

Proposed model for the degradation of GAX
(A) Initial polysaccharide degradation by secreted enzymes
External environment
CBM2
CBM6
CBD (bpr_I0736)
CBD (bpr_I1599)
CE1/est1A
CE1/est1C
GH43/xsa43A, F, J
GH9/cel9A
GH10/xyn10B
GH5/cel5A
GH3/gh3A
GH5/cel5C, D
GH10/xyn10A, B, C, E
GH53/agn53A
CE4/est4C
CE8/pme8B
Outer membrane
(B) Transport of xylo-oligosaccharides by sugar ABC transport systems
External environment
Xylo-oligosaccharides
Xylobiose
Xylose
sugar ABC transporter (COG1879)
sugar ABC transporter (COG1653)
sugar ABC transporter (COG2182)
xylose ABC transporter (COG4213)
Sugar ABC transporter systems
Outer membrane
ATP
ADP
Cytoplasm
CE4/est4B, D, E
CE2/est2A
CE12/est12C
AXE
CE4/est4A, C, D
CE1/est1D, E
GH51/arf51B, C
GH51/arf51A, B
GH43/xsa43H
GH43/xsa43D
GH120/xyl120A, B
GH9/cel9B
GH8/xyn8A
GH43/xsa43B, D, E, G, I
GH39/xyl39A
GH9/cel9B
GH3/xyl3A
GH10/xyn10D
xylosidase
GH8/xyn8A
GH67/agu67A
GH115/gh115A
GH115/gh115A
GH31/gh31C
GH31/gh31B, C
GH36/aga36C
(C) Debranching of xylo-oligosaccharides by intracellular enzymes

Figure 6.2. Model for the degradation of GAX by *B. proteoclasticus* B316 and *B. hungatei* MB2003. The model for degradation of complex GAX includes the primary attack (A), assimilation (B) and intracellular processing (C). The black arrows indicate examples of the linkages cleaved by the enzymes. The significantly up-regulated polysaccharidases are represented by ovals shown in green for *B. proteoclasticus* B316 and blue for *B. hungatei* MB2003. Enzymes are identified based on their CAZy classifications, consisting of GHs, CEs, PLs and CBMs. The red arrow heads indicate xylo-oligosaccharide transport between cellular locations. SBPs signify the surface sugar and xylose binding proteins of the ABC transporter systems (Davidson *et al.*, 2008). Schematic representation of the ATP-driven sugar ABC transport systems likely to mediate the uptake of xylan-derived soluble sugars are also included. All polysaccharidases and transport systems shown are based on transcriptional analysis of genes deemed significantly expressed in either mono- or co-culture growth conditions. Symbols: ★, *D*-xylose; ▲, acetyl group; ◆, methyl group; ⬟, *L*-arabinose; ●, *D*-galactose; ⬢, *D*-glucuronic acid; ✹, ferulic acid. Abbreviations: AXE, acetyl-xylan esterase; CBD, carbohydrate binding protein. COG designations: COG1653, ABC-type sugar transport system; COG1879, ABC-type sugar transport system; COG2182, maltose-binding periplasmic proteins/domains; COG4213, ABC-type xylose transport system. The xylan structures shown are not quantitatively accurate.

The GH43 family enzymes included in the model possess a wide range of potential functions including β-xylosidases, α-L-arabinofuranosidases, and endoxylanases and are important for the efficient degradation of hemicellulose. The β-xylosidases are exo-type glycosidases (EC 3.2.1.37), and remove β-xylosyl residues from non-reducing termini of xylobiose and xylo-oligosaccharides. This activity helps prevent the end-product inhibition of endo-xylanase during complete degradation of xylan (Coughlan and Hazlewood, 1993; Beg *et al.*, 2001). Previous functional characterization of the *xsa43E* from *B. proteoclasticus* B316 found this enzyme has dual β-xylosidase and α-L-arabinofuranosidase activities (Till *et al.*, 2014). α-L-arabinofuranosidases cleave arabinose side chains from substituted xylo-oligosacccharides derived from xylan, and are predicted to have a role in disrupting the complex inter- and intra-polymer networks within the plant cell wall (Grabber *et al.*, 2000). It is proposed that *xsa43A* is secreted into the extracellular environment where it acts on the xylan to release arabinose, and may also contribute to the hydrolysis of α-linked L-arabinofuranosyl groups from the xylan backbone prior to xylo-oligosaccharide assimilation.

The *xsa43A*, *F* and *J* genes were constitutively expressed in xylan-grown B316 cells and *xsa43A* was the only B316 gene significantly differentially expressed during growth on xylan that was predicted to be secreted. The *xsa43A* gene encodes an N-terminal GH43 (Pfam04616) catalytic domain (Figure 5.17A), and a C-terminal CBM6 (Pfam03422)

non-catalytic module that has been shown to bind xylan in other organisms (Gilbert *et al.*, 2013). Previous studies have also shown that CBM6 domains are able to recognise xylose either as a monosaccharide, at the non-reducing end of xylo-oligosaccharides, or within the side chain components of xyloglucan (Abbott *et al.*, 2009). Several CBM6 modules also recognise (1,3)-β-D-linkages at the non-reducing end of β-glucans (van Bueren *et al.*, 2005; Correia *et al.*, 2009), and appear to have co-evolved with their associated catalytic domains to acquire the same substrate specificity (Michel *et al.*, 2009). Furthermore, high arabinofuranosidase activity has previously been observed with *B. fibrisolvens* grown on xylan suggesting arabinofuranosidase activity may be regulated by the availability of arabinose (Hespell *et al.*, 1987; Sewell *et al.*, 1988; Hespell and Whitehead, 1990; Hespell, 1992a). The observation that little arabinose was detected in B316 mono- and co-culture xylan-grown cells (Figure 5.6), suggests a high turnover of arabinose due to the significant up-regulation of *xsa43A*.

Glycoside hydrolases of family GH9 are the second largest cellulase family and contain mainly endoglucanases (Henrissat *et al.*, 1989; Gilkes *et al.*, 1991). GH9 enzymes have been shown to express discernible activity on numerous plant polysaccharides including xylan, 1,3-1,4-ß-glucan, arabinoxylan, xyloglucan, and glucomannan (Master *et al.*, 2004; Yoshida and Komae, 2006; Li *et al.*, 2007). *Cel9A* was the most up-regulated gene of MB2003 mono-culture xylan-grown cells, and its encoded enzyme contains a CBM30 (Pfam02927) non-catalytic domain attached to the C-terminus of the GH9 (Pfam00759) catalytic domain. The strong up-regulation of *cel9A* suggests that it is an important xylanolytic enzyme of MB2003, however the activity/specificity of this enzyme has not been sufficiently elucidated and is a point for future investigation.

Xylan polymers are cross-linked via ester linkages with the phenolic acids, ferulic and *p*-coumaric acids, which are esterified to the arabinose side chains (Hatfield, 1993), as shown in Figure 6.1A. These linkages have been proposed to account for much of the steric hindrance to plant fibre degradation in the rumen (Bacic *et al.*, 1988), and there is significant interest in rumen microorganisms that possess feruloyl and *p*-coumaroyl esterases that belong to the CE1 (Pfam00756) CAZy family. These esterases function to remove the ferulic acid, methyl and acetyl groups from the xylan polymers and/or oligomers, thus making arabinose and xylose available for growth. The cell-associated β-xylosidases and α-L-arabinofuranosidases are then likely to target the arabinoxylan oligomers from GAX that are subsequently assimilated into the cell. A previous study has shown that high levels of extracellular cinnamoyl esterases are characteristic of a

selection of fibre-degrading ruminal bacteria and in particular are common amongst the xylanolytic *B. fibrisolvens* (McSweeney *et al.*, 1998).

The extracellular feruloyl esterases *est1A* (bhn_III78) of MB2003 and *est1C* (bpr_I0187) of B316 were up-regulated in mono-cultures on xylan, *est1A* significantly so (Figure 5.17B and Table A3.13). The homologous polysaccharide deacetylases *est4C* of MB2003 (bhn_I0856) and B316 (bpr_I0951) were up-regulated in mono-cultures grown on xylan, and are predicted to target the acetyl side groups of xylan. It has been argued that the well-developed xylanolytic and cinnamoyl esterase systems of ruminal fungi give them a distinct advantage over fibrolytic bacteria in the rumen (Christov and Prior, 1993). However, the high levels of CE1 esterases (total of 385) identified in the pan-genome CAZy comparisons (Figure 4.18A), clearly support *Butyrivibrio* and *Pseudobutyrivibrio* as important xylan- and pectin-degrading ruminal microorganisms. The feruloyl and *p*-coumaroyl esterase activities of *Butyrivibrio* and *Pseudobutyrivibrio* thus warrants further investigation in future work.

In addition to the secreted polysaccharidases identified in the *B. proteoclasticus* B316 xylan-grown cells, a curious feature was the presence of two CBPs. The two CBPs encoded by B316 (bpr_I0736 and bpr_I1599) both contain two CBM2a domains, and bpr_I1599 also contains two CBM6 domains (Kelly *et al.*, 2010). CBM6 domains characteristically bind xylose and are associated with xylanase activity with ligand specificity for xylan (McCartney *et al.*, 2004). CBM2 domains, are divided into two sub-families: 2a, that bind to crystalline cellulose even when associated with xylanases (Gilkes *et al.*, 1992), and 2b, that bind to xylan (Simpson *et al.*, 1999). The ligand specificity of the CBM2 domains has been shown to be determined by a single amino acid difference (Simpson *et al.*, 2000).

6.6.1.2 Transport of xylo-oligosaccharides and monosaccharides

In order to make use of oligo- and monosaccharides derived from extracellular digestion of lignocellulosic material, it is necessary for the released soluble sugars to be transported into the cell. Both B316 and MB2003 have Gram-positive cell wall structures even though cultures stain Gram-negative (Attwood *et al.*, 1996; Palevich, 2011). In Gram-positive bacteria, oligo- and monosaccharide transport across the bacterial cell wall is mediated by a variety of extracellular SBPs linked to dedicated sugar ABC transporter systems. SBPs are lipoproteins that tether themselves to the external surface of the cell wall or directly fuse to the membrane transporter and are strictly associated with import

systems (van der Heide and Poolman, 2002; Davidson *et al.*, 2008). The SBP component of bacterial ABC transport systems, thus act as a scavenger for substrate molecules from the environment. A common arrangement in other polysaccharide-degrading Gram-positive bacteria that has also been observed in both B316 and MB2003 genomes, is that the ABP and PP are not universally associated with SBPs (Dam *et al.*, 2011). Recent data on the B316 secreted proteome (Dunne *et al.*, 2011; Dunne *et al.*, 2015) and carbohydrate transport-associated membrane proteins (Bond *et al.*, 2012), as well as genome sequence analysis of MB2003 (Palevich, 2011), identified a large number of sugar-specific ABC transporter SBPs, which further supports the rationale for the proposed model.

The transcriptomic analysis of B316 and MB2003 grown with xylan as the sole substrate identified up-regulation of 45 and 38 genes encoding membrane proteins in B316 and MB2003 respectively, and predicted to function as carbohydrate transporters (Appendix III Tables A3.12 and A3.13). Among the 38 ABC transporter proteins expressed in MB2003, up-regulation of genes encoding three proteins with predicted functions as xylose ABC transporters was detected in xylan co-cultures. For B316 growth on xylan, no genes encoding xylose ABC transporter systems were differentially expressed. MB2003 grown in mono-culture on xylan, displayed significant expression of 13 genes associated with many different transport systems and target substrates.

Detection of a large number of significantly, differentially expressed ABC transporter SBP-encoding genes detected in the B316 and MB2003 signifies the importance of SBP-dependent ATP-driven active transport of oligo- and mono-saccharides for growth. The most prevalent up-regulated genes encoding SBPs in both xylan and pectin growth conditions belong to the Carbohydrate Uptake Transporter 1 (CUT1) family, that is comprised of a large and diverse collection of SBP-dependent ABC transporter systems dominated by those that mediate di- and oligo-saccharide uptake (Fukami-Kobayashi *et al.*, 1999). The analysis revealed that the predominant CUT1 and CUT2 family SBPs expressed in xylan-grown cells were represented by the COG1653 and COG4213 domains, where both are associated with sugar ABC transporters (Berntsson *et al.*, 2010).

All of the SBPs represented by the COG1653 domain in both B316 and MB2003 were found co-localised with other genes predicted to be involved in xylan degradation and metabolism, including various genes encoding GH43, GH10, GH51, GH53 and CE12 CAZy domains (Figure 5.20A and B). Furthermore, two SBP-encoding genes identified in xylan-grown B316 mono- and co-cultures, bpr_I0182 and bpr_I0313, have previously

been reported in B316 xylan-grown cells to be found in clusters of genes predicted to be involved in hemicellulose degradation, deemed polysaccharide utilisation loci or PULs (Kelly *et al.*, 2010). The PULs representative of these SBPs contain genes encoding β-xylosidases, α-glucuronidases, acetyl-xylan esterases, ferulic acid esterases and secreted endo-1, 4-β-xylanases (Dunne, 2010), and have been proposed as being important for efficient hemicellulose metabolism by B316.

6.6.1.3 Cytosolic degradation of xylo-oligosaccharides

Fermentation end product analysis of mono- and co-cultures of B316 and MB2003 revealed that only B316 was capable of efficiently degrading and utilising xylan (Table 5.1). However, previous investigations into their carbohydrate utilisation profiles have identified that both are capable of utilising the monosaccharides arabinose, galactose, and xylose (Palevich, 2011). The B316 and MB2003 genomes contain 90 and 71 genes respectively, predicted to encode cytosolic oligosaccharide-degrading enzymes represented by various GH, CE and PL families (Kelly *et al.*, 2010; Palevich, 2011). In addition, 59 of the MB2003 genes encoding cytosolic oligosaccharidases have corresponding homologues in B316. Recent proteomic analysis of B316 grown on xylan or xylose, has identified almost 400 intracellular enzymes predicted to metabolise assimilated products of the initial extracellular xylan degradation (Dunne *et al.*, 2015). The intracellular debranching enzymes identified in the above mentioned proteomic analysis include endo-1,4-β-xylanases (GHs 5 and 10), β-xylosidases (GHs 3, 8, 30, 43 and 120), arabinofuranosidases (GHs 43 and 51), α-glucuronidases (GHs 67 and 115), feruloyl esterases (CE1) and acetyl xylan esterases (CEs 2, 4 and 10).

From the transcriptomic analysis of B316 and MB2003 grown with xylan or pectin, the highly expressed genes predicted to encode oligosaccharide debranching enzymes and their catalytic domains are: *xyl120A, B* (GH120), *est1B, D, E* (CE1), *xsa43B-I* (GH43), *xyn8A* (GH8), *xyl3A* (GH3), *agu67A* (GH67), *arf51A-C* (GH51), *gh115A* (GH115), *gh31B, C* (GH31), *est2A* (CE2), *est4A-E* (CE4), and an unclassified acetyl-xylan esterase (AXE) (Figure 6.2C). The strong up-regulation of debranching enzymes and sugar SBP-dependent ABC transporter systems, indicates that extracellular degradation of xylan does not result in complete removal of side chain groups and that the xylo-oligosaccharides are assimilated into the cytoplasm for further degradation by intracellular enzymes, further supporting the proposed model for GAX degradation.

6.6.2 Proposed model for XGA and RG-I degradation

The proposed model for XGA and RG-I degradation based on the fibrolytic potential of *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown on pectin is shown in Figure 6.3, and describes the initial extracellular degradation of XGA and RG-I, followed by transport of the pectic-oligosaccharides into the cytoplasm where the internalised pectic-oligosaccharides are broken-down and metabolised.

6.6.2.1 Extracellular XGA and RG-I degradation

In addition to the secreted enzymes containing the various GH and CE CAZy domains for degradation of GAX and its constituent side groups, the GH13 and PL1 domains that target (1,4)-β-D-linkages within the galacturonic acid backbones are also necessary for the extracellular degradation of XGA and RG-I. MB2003 significantly expressed secreted enzymes containing GHs 53, 30, 13 as well as 43 and 10 CAZy domain containing genes (Figure 5.18B). B316 did not up-regulate any genes encoding secreted CAZymes when grown in mono- and co-culture on pectin, however it outcompeted MB2003 in co-culture and displayed better growth than MB2003 in mono-culture.

The GH43 family enzymes have predicted roles in the debranching of pectin and similar to the *xsa43E* from B316, the *xsa43A* expressed in MB2003 pectin-grown cells is also proposed to possess dual β-xylosidase and α-L-arabinofuranosidase activities. It is proposed that *xsa43A* is secreted where extracellular debranching of pectin results in release of arabinose and xylose side groups prior to pectic-oligosaccharide assimilation. The *xsa43A* gene up-regulated in MB2003 mono-culture, was only GH43 family enzyme that was expressed in pectin-grown cells and predicted to be secreted. The *xsa43A* gene of MB2003 is homologous to the *xsa43A* of B316 and encodes an N-terminal GH43 catalytic domain and a C-terminal CBM6 non-catalytic module (Figure 5.18B).

In contrast to the GAX polymers, ferulic acids are not components of XGA or RG-I and the CE family esterases instead are predicted to remove the 4-*O*-methylglucuronic acid and acetyl groups from the pectin polymers and/or oligomers to facilitate degradation of the pectin backbones. The secreted feruloyl esterases *est1A* of MB2003 and *est1C* of B316 were also up-regulated in pectin-grown cells (Table A3.13). However, the polysaccharide deacetylase *est4C* predicted to target the acetyl side groups of pectin was only up-regulated in MB2003 mono-cultures grown on pectin.

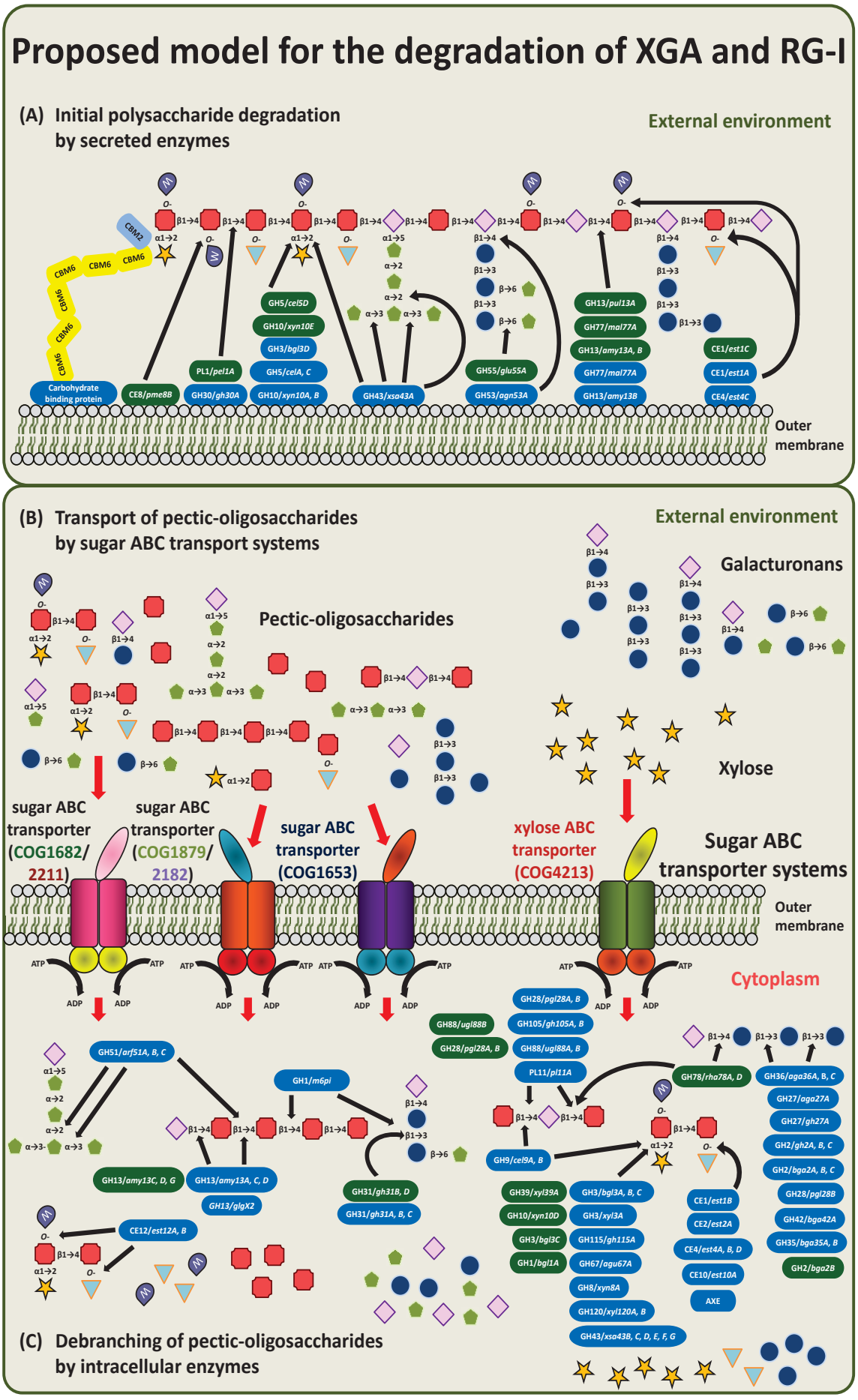

Proposed model for the degradation of XGA and RG-I
(A) Initial polysaccharide degradation by secreted enzymes
External environment
CBM2
CBM6
CBM6
CBM6
CBM6
CBM6
Carbohydrate binding protein
CE8/pme8B
PL1/pel1A
GH30/gh30A
GH5/cel5D
GH10/xyn10E
GH3/bgl3D
GH5/celA, C
GH10/xyn10A, B
GH43/xsa43A
GH55/glu55A
GH53/agn53A
GH13/pul13A
GH77/mal77A
GH13/amy13A, B
GH77/mal77A
GH13/amy13B
CE1/est1C
CE1/est1A
CE4/est4C
Outer membrane
(B) Transport of pectic-oligosaccharides by sugar ABC transport systems
External environment
Galacturonans
Pectic-oligosaccharides
Xylose
sugar ABC transporter (COG1682/2211)
sugar ABC transporter (COG1879/2182)
sugar ABC transporter (COG1653)
xylose ABC transporter (COG4213)
Sugar ABC transporter systems
Outer membrane
Cytoplasm
ATP
ADP
GH51/arf51A, B, C
GH1/m6pi
GH13/amy13C, D, G
GH13/amy13A, C, D
GH13/glgX2
GH31/gh31B, D
GH31/gh31A, B, C
CE12/est12A, B
GH88/ugl88B
GH28/pgl28A, B
GH28/pgl28A, B
GH88/ugl88A, B
PL11/pl11A
GH9/cel9A, B
GH39/xyl39A
GH10/xyn10D
GH3/bgl3C
GH1/bgl1A
GH3/bgl3A, B, C
GH3/xyl3A
GH115/gh115A
GH67/agu67A
GH8/xyn8A
GH120/xyl120A, B
GH43/xsa43B, C, D, E, F, G
GH78/rha78A, D
CE1/est1B
CE2/est2A
CE4/est4A, B, D
CE10/est10A
AXE
GH36/aga36A, B, C
GH27/aga27A
GH27/gh27A
GH2/gh2A, B, C
GH2/bga2A, B, C
GH28/pgl28B
GH42/bga42A
GH35/bga35A, B
GH2/bga2B
(C) Debranching of pectic-oligosaccharides by intracellular enzymes

Figure 6.3. Model for the degradation of XGA and RG-I by *B. proteoclasticus* B316 and *B. hungatei* MB2003. The model for degradation of complex XGA and RG-I includes the primary attack (A), assimilation (B) and intracellular processing (C). The black arrows indicate examples of the linkages cleaved by the enzymes. The significantly up-regulated polysaccharidases are represented by ovals shown in green for *B. proteoclasticus* B316 and blue for *B. hungatei* MB2003. Enzymes are identified based on their CAZy classifications, consisting of GHs, CEs, PLs and CBMs. The red arrow heads indicate xylo- or pectic-oligosaccharide transport between cellular locations. SBPs signify the surface sugar binding proteins of the ABC transporter systems (Davidson *et al.*, 2008). Schematic representation of the ATP-driven sugar ABC transport systems likely to mediate the uptake of pectin-derived soluble sugars are also included. All polysaccharidases and transport systems shown are based on transcription analysis of genes deemed significantly expressed in either mono- or co-culture growth conditions. Symbols: ★, *D*-xylose; ▲, acetyl group; ◆, methyl group; ⬟, *L*-arabinose; ●, *D*-galactose; ⯃, *D*-galacturonic acid; ◇, *L*-rhamnose. Abbreviations: AXE, acetyl-xylan esterase; CBD, carbohydrate binding protein. COG designations: COG1653, ABC-type sugar transport system; COG1682, ABC-type polysaccharide/polyol phosphate export systems; COG1879, ABC-type sugar transport system; COG2182, maltose-binding periplasmic proteins/domains; COG2211, Na^+/melibiose symporter and related transporters; COG4213, ABC-type xylose transport system. The pectin structures shown are not quantitatively accurate.

The *gh30A* (bhn_I2771) was significantly up-regulated only in MB2003 mono-cultures grown on pectin (Figure 5.18B). This enzyme has a GH30 (Pfam02055) catalytic domain that is predicted to have activities in debranching of pectin to release *D*-xylose and to hydrolyse (1,4)-β-D-linkages in xylans. However, due to the lack of homology of the GH30 domain to enzymes in the NCBI database with characterised functions, it is difficult to accurately predict the catalytic potential of this enzyme. The relevant prescribed activities for GH30 enzymes include β-xylosidase, endo-β-1,4-xylanase, β-glucosidase, β-glucuronidase, β-1,6-glucanase, glucurono-arabinoxylan endo-β-1,4-xylanase and endo-β-1,6-galactanase. Relatively high expression of a B316 GH30 homologue was also observed in previous experimental work under growth on xylan (Dunne, 2010). However, the functional role of GH30 enzymes, and their contribution to xylan and pectin degradation, will require detailed biochemical analysis.

The β-1,4-galactanases containing GH53 catalytic domains are predicted to degrade galactan and arabinogalactan side chains in the hairy regions of pectin, by attacking the 1,4-β-D-galactosidic linkages in Type I arabinogalactans (Le Nours *et al.*, 2009). In B316, *agn53A* was not expressed in either culture condition, however, *agn53A* (bhn_I0681) was

significantly expressed in mono-culture pectin-grown cells and was the most up-regulated gene in co-culture xylan-grown MB2003 cells (Figures 5.18B and 5.17C). This suggests that *agn53A* encodes an important xylan- and pectin-degrading enzyme in MB2003. Although the *agn53A* from MB2003 does not possess a CBM module, CBMs have frequently been found in other proteins with β-1,4-galactanase activity and recently the CBM61 was characterised with galactose-containing ligands (Cid *et al.*, 2010). Furthermore, given the predicted functions of *agn53A*, the absence of a CBM module may account for its up-regulation under both xylan and pectin conditions. Future biochemical characterisation of this enzyme is warranted to determine its substrate specificity.

CBM domains function to direct their associated catalytic domains to specific regions of polysaccharides where they can carry out their activity. It is generally accepted that the addition of a CBM domain to the enzymatic catalytic domain of an enzyme increases enzyme concentration at the site of activity, and improves enzyme substrate interaction and catalysis (Black *et al.*, 1996; Bolam *et al.*, 1998; Yu *et al.*, 2003). CBMs are presently classified into 71 families, where binding specificities for the major plant cell wall polysaccharides, cellulose and xylan, are well represented and reflect the importance of CBMs in the enzymatic degradation of these polysaccharides (Boraston *et al.*, 2004). A recent study has identified that a CBM35 (Pfam03422) in *Amycolatopsis orientalis* targets cellulosomal pectin acetylesterase to regions of pectin that have been cleaved by a pectate lyase (PL10) (Montanier *et al.*, 2009). However, there appears to be an underrepresentation of CBMs with binding specificity for pectin, despite the fact that pectin is a major component of the plant cell wall and this aspect of fibre degradation calls for further investigation.

MB2003 encodes one secreted CBP (bhn_I1873, 984 aa), and this protein was significantly up-regulated only in pectin-grown mono-culture cells (Figure 5.18B). The domain structure of MB2003 bhn_I1873 is unique, containing six CBM6 domains towards the N-terminus and a single C-terminal CBM2a domain. It is interesting that even though the transcription levels of the two CBPs of B316 were increased in xylan-grown cultures, the only CBP that displayed significant differential expression was the bhn_I1873 in MB2003 mono-culture pectin-grown cells compared to growth on xylan. Recent studies have shown that in discrete regions of plant cell walls, initial enzymatic attack of pectin increases the access of CBMs to cellulose (Blake *et al.*, 2006), effectively loosening the polysaccharide interactions to reveal the xylan and xyloglucan substrates

(Marcus *et al.*, 2008; Hervé *et al.*, 2009). This initial stage in enzymatic saccharification of plant cell walls termed amorphogenesis (Coughlan, 1985), is a possible role of such CBPs containing multiple non-catalytic domains. In the rumen, B316 and MB2003 may thus secrete these non-catalytic CBPs synergistically with polysaccharide-active enzymes as a mechanism to enhance the rate and extent of plant cell wall degradation by disrupting the interface between other polysaccharides. These non-catalytic CBPs are therefore of interest and future *in vitro* experimentation aimed at investigating their function, and the role of their CBMs is justified.

6.6.2.2 Transport of pectic-oligosaccharides and monosaccharides

The transcriptomic analysis of B316 and MB2003 grown with pectin as the sole substrate identified 10 and 78 membrane proteins in B316 and MB2003 respectively, predicted to function as carbohydrate transporters (Appendix III Tables A3.12 and A3.13). Among the 78 ABC transporter proteins expressed in MB2003, four proteins with predicted functions as xylose ABC transporters were identified in pectin mono-cultures. For B316 mono- and co-culture growth on pectin, the only transport-associated genes found, had predicted functions as sugar ABC transporters and in contrast to MB2003, no genes encoding xylose ABC transporter systems were differentially expressed in pectin-grown cultures. MB2003 grown in mono-culture on pectin, displayed significant expression of 46 genes associated with many different transport systems and target substrates.

The predominant SBPs expressed in B316 and MB2003 pectin-grown cells were represented by the COG1653 and COG4213 domains. The SBP represented by the COG1879 domain functions as a ribose, xylose, arabinose and galactose ABC transporter, was among the most significantly expressed SBPs in MB2003 grown in mono-culture on pectin. In addition to the SBPs represented by the COG1653 domain in both B316 and MB2003 found co-localised with other genes involved in xylan degradation and metabolism, genes containing GH13 and PL11 (only in MB2003) CAZy domains (Figure 5.20C to F), were expressed in pectin-grown cells. This implies that the availability of a particular carbon source causes activation of a wider network of genes that enable these rumen bacteria to breakdown, transport and metabolise such substrates for growth.

6.6.2.3 Cytosolic degradation of pectic-oligosaccharides

Fermentation end product analysis revealed that only B316 was capable of efficiently degrading and utilising pectin in both mono- and co-culture, while MB2003 showed minimal growth in mono-culture (Table 5.1). The arabinose, galactose, and xylose are abundant components of the side groups of both XGA and RG-I, where both B316 and MB2003 have been shown to be capable of utilising each of these monosaccharides (Palevich, 2011).

As previously mentioned, many of the MB2003 genes encoding cytosolic oligosaccharide-degrading enzymes have corresponding homologues in B316, however this is the first reported investigation into the transcriptomes of *Butyrivibrio* strains grown with pectin as the sole substrate. From these transcriptomic analyses, all of the genes up-regulated in the xylan-grown B316 and MB2003 were also up-regulated in pectin-grown cultures (Figure 6.3C), and with regard to MB2003, all genes have significant matches to B316 except *xsa43C*, *xsa43E*, *arf51C* and *est4A*. In addition, genes that were expressed only in pectin-grown cells include: *bgl3A-C* (GH3), *est10A* (CE10), *pl11A* (PL11), *ugl88A, B* (GH88), *rha78A, D* (GH78) and *amy13A, C, D, G* (GH13) (Figure 6.3C). The *pl11A*, *ugl88A, B*, *rha78A, D* and *amy13A, C, D, G*, are associated with predicted functions as pectate lyases, unsaturated glucuronyl hydrolases, rhamnosidases and α-amylases, respectively, and are proposed to hydrolyse the (1,4)-β-linkages of the XGA and RG-I backbones. The *bgl3A-C* and *est10A* have predicted activities as β-glucosidases and acetyl-xylan esterases, respectively, and are proposed to cleave the xylose and acetyl side groups from the pectic-oligosaccharides.

Similar to the xylan-grown cells, the abundance and high up-regulation of debranching enzymes and sugar SBP-dependent ABC transporter systems particularly in MB2003, corroborates the incomplete extracellular degradation of pectin and removal of side chain groups. The pectic-oligosaccharides must therefore be assimilated into the cytoplasm for further degradation by intracellular debranching enzymes, further supporting the proposed model for XGA and RG-I degradation.

6.7 Conclusions

The results of this thesis provide new insight into the types of enzymes that are predicted to play important roles in polysaccharide degradation by *Butyrivibrio* and *Pseudobutyrivibrio*. In particular, it highlights differences in both the capacity and capability of the xylan and pectin degrading enzymatic repertoires of *B. proteoclasticus* and *B. hungatei* that enables them to degrade these complex polysaccharides in concert within the rumen. Comparison of the *Butyrivibrio* and *Pseudobutyrivibrio* pan-genome has revealed distinct species and inter-species grouping based on both their genome sequence similarities but also on their polysaccharide degrading enzyme complements.

Several *Butyrivibrio* strains have been characterised that are genetically closely related to both *B. proteoclasticus* and *B. hungatei*, suggesting that the commensal interactions identified between these strains may occur for xylan and pectin degradation throughout the *Butyrivibrio*/*Pseudobutyrivibrio* genus. Investigation of the transcriptome components of the xylan and pectin degrading systems of two distinct and relatively abundant *Butyrivibrio* species has provided insight into the biological and metabolic processes involved in plant cell-wall degradation by the lignocellulolytic rumen microbiota. Furthermore, this project has shown that the growth conditions and substrates encountered by these bacteria evoke gene expression responses that result in increased expression of several secreted and cytosolic polysaccharidases and substrate-binding proteins proposed to mediate the assimilation of soluble oligosaccharides. The clear differences observed in gene expression between *B. proteoclasticus* and *B. hungatei* allows one to speculate that both bacteria have sensory mechanisms that enable cells to identify the polysaccharides present in the external environment, and respond accordingly to utilise the available growth substrate.

This project is the first pan-genome analysis of the fibre degrading enzyme capabilities of the polysaccharolytic rumen bacteria, *Butyrivibrio* and *Pseudobutyrivibrio*. The pan-genome results have identified the large capacity and diversity of polysaccharidases synthesised by these abundant fibre degrading rumen bacteria. This project has demonstrated that whole-genome sequence-based bioinformatics and 16*S* rRNA marker gene sequencing approaches can be used in combination to produce a detailed and complementary reconstruction of the phylogeny of these organism. Furthermore, the comprehensive analysis across all of the CAZy categories has allowed us to gain insight into the enzyme classes among the *Butyrivibrio* and *Pseudobutyrivibrio*

lineages and the results of this project will be a valuable blueprint for comparative genome analyses in the future.

The transcriptome analyses have highlighted the differences in the relative amounts of up-regulated genes in *B. proteoclasticus* B316 and *B. hungatei* MB2003 grown in mono- and co-culture that support the observed commensal interactions between the two rumen bacteria in the co-culture growth experiments. These analyses also confirmed the expression of the diverse repertoire of genes that MB2003 and B316 possess which encode polysaccharide-degrading enzymes and sugar ABC transport systems during xylan and pectin degradation. The SBP-mediated ABC-assimilation of polysaccharide-derived soluble sugars, such as xylo- and pectic-oligosaccharides, has been shown to be an essential component for xylan and pectin degradation and utilisation by B316 and MB2003. For future work, biochemical characterisation of the precise substrate specificities of the SBPs for the various xylo- and pectic-oligosaccharides and characterisation of the carbon catabolite repression mechanisms will be essential to defining the sugar transporting systems of B316 and MB2003.

The results from this thesis further support previous studies into the mechanisms of plant cell wall degradation for B316 (Kelly *et al.*, 2010). B316 possesses a highly developed glycobiome that is dedicated to the breakdown and reassembly of complex polysaccharides primarily by secretion of a large range of polysaccharide-degrading enzymes and its diverse intracellular enzymes to breakdown oligosaccharides. It is proposed, that *Butyrivibrio* initiate xylan and pectin breakdown extracellularly via secretion of polysaccharide-degrading enzymes and the SBPs are responsible for transporting xylo- and pectic-oligosaccharides into the cell, where they are further debranched by intracellular polysaccharide-degrading enzymes and the resulting monosaccharides are metabolised. In the future, this information will contribute to understanding digestion of lignocellulose in the rumen, provide insight for discovery of new opportunities to improve forage degradation and manipulate rumen function to maximise productivity in forage fed ruminants.

6.8 Future perspectives

Future analyses directed at understanding the xylan and pectin degrading enzyme systems of *B. proteoclasticus*, *B. hungatei* and other members of *Butyrivibrio* and *Pseudobutyrivibrio* species groups, should be focused on the catalytic properties of the polysaccharidases and transporter proteins identified in this project. In particular, the array of DEGs encoding secreted polysaccharidases identified in *B. hungatei* and *B. proteoclasticus* cells grown in xylan (Kong, 2007) and pectin supplemented culture medium should be investigated. The genomes of *B. proteoclasticus* and *B. hungatei* identified the large capacities and diverse capabilities of polysaccharidases which signifies their importance in the rumen for degradation and assimilation of xylan and pectin. Also, important targets for future analysis should include the cytosolic polysaccharidases that direct hydrolysis of the backbone and substituent groups of xylo- and pectic-oligosaccharides discussed above.

To gain a more complete understanding of the metabolic capacity of *Butyrivibrio* species, a comparison of their growth on a range of soluble substrates, end-product and metabolite analyses must be compared in parallel with whole genome sequences, as well as gene and protein expression profiling. Such investigations will improve current knowledge of the contributions that *Butyrivibrio* make to the digestion of plant fibre, that lead to the production of fermentation products and supply of microbial cells that provide the host animal with its main sources of energy and protein, with the ultimate goal of improving ruminant production. This thesis contributes insight into the mechanisms by which *Butyrivibrio* degrade and use polysaccharides that can provide new biotechnological tools and improve a number of well-established technologies used to improve plant fibre degradation in the rumen prior to feed consumption (Hu *et al.*, 1991; Berger *et al.*, 1994; Allen, 1996), and for the conversion of lignocellulosic biomass into monosaccharides.

The proposed model of GAX, XGA and RG-I degradation, assimilation and intracellular processing by *Butyrivibrio* was in the context of the oat-spelt xylan and apple derived pectin growth substrates. A recent study compared the transcriptome of *B. proteoclasticus* grown on oat-spelt or wheat-bran xylan (Kong, 2007), and found clear differences in the relative abundance of some polysaccharidases evoked by growth on different substrates. The most appropriate substrates for investigating the plant fibre-degrading capabilities of *Butyrivibrio* within the rumen would have been xylans and pectins derived from different forms of grasses that are abundant in New Zealand pasture,

such as ryegrass and clover, as well as other herbaceous species used as forage crop such as alfalfa. However, due to the high variation in hemicellulose concentrations among different crop species (Martens and Loeffelmann, 2002), and the technical difficulties around sourcing a large amount of such material of uniform composition (Van Soest *et al.*, 1991), such substrates were unavailable for this work. Therefore, an analysis of the xylan- and pectin-degrading enzyme profiles as well as the various cell-surface bound SBPs and their target sugars, of *B. proteoclasticus*, *B. hungatei* and other members of *Butyrivibrio* cells post cultivation on chemically diverse xylans and pectins warrants further investigation. Furthermore, additional experimental validation of the proposed models should be addressed by measuring global enzymatic activity or by obtaining global metabolomics data to be used in conjunction with the transcriptomic data.

The production of large amounts of exopolysaccharides is a common feature of *Butyrivibrio*/*Pseudobutyrivibrio* and in particular *B. fibrisolvens* and *B. proteoclasticus* cells grown in liquid media. Previous work has identified 363 genes in the *B. proteoclasticus* genome that encode putative proteins involved in exopolysaccharide synthesis (Kelly *et al.*, 2010). Although there was no evidence for the differential expression of genes involved in exo-polysaccharide synthesis in the data generated by this thesis, analysis of the transcriptomes of a variety of *Butyrivibrio* and *Pseudobutyrivibrio* strains grown on a range of xylan and pectin substrates would be meaningful in understanding exopolysaccharide production and regulation.

The rumen is a substantial source of fibrolytic capacity and potentially novel fibrolytic enzymes that are likely to have applications outside the rumen. Examples of polysaccharide-degrading enzymes identified from *Butyrivibrio* and *Pseudobutyrivibrio* include well characterised arabinosidases and feruloyl esterases that have industrial and biotechnological applications (Tabka *et al.*, 2006; Topakas *et al.*, 2007). A potential application of this project is in the development of pre-treatment systems for the depolymerisation of lignocellulosic material to provide a fermentable feedstock for the generation of renewable energy sources (Mielenz, 2001; Alvira *et al.*, 2010). At present, bioethanol is largely produced from corn grain starch and is the most sought after renewable fuel. However, due to the increasing demands of raw materials for large-scale production other sources are called for (Gray *et al.*, 2006). These economic limitations necessitate the exploitation of low value, lignocellulose feedstocks such as agricultural crops residues and wastes (Sarkar *et al.*, 2012).

Recently, investigations have been conducted on the use of co-cultivation of cellulolytic organisms as a possible means of increasing the rate of lignocellulosic bioconversion (Zuroff and Curtis, 2012). For example, a system was tested in which one strain of *E. coli* is genetically engineered to secrete fibrolytic enzymes and degrade xylan into xylo-oligosaccharides, while another *E. coli* strain was modified to partially uptake and utilise the released substrates (Shin *et al.*, 2010). Although this engineered system has potential, a system that is based on natural commensal interactions between bacteria in co-culture may be a more efficient approach to convert lignocellulosic materials into biofuels. Therefore co-cultivation of the relatively abundant lignocellulosic *Butyrivibrio* and *Pseudobutyrivibrio* may be a practical and effective means of increasing the levels of enzymatic components produced by these rumen bacteria. For instance, the strategy used in the co-culture experiment of pairing a xylan/pectin degrading *Butyrivibrio* (for example *B. proteoclasticus* or *B. fibrisolvens*) with another or multiple strains that are not able to utilise these complex substrates in mono-culture (such as *B. hungatei* or *Butyrivibrio* sp. 2 strains), may evoke similar commensal interactions and result in substantial production of fibrolytic enzymes. This is a valid approach as the co-culture system does not call for additional improvements and genetic engineering of these rumen bacteria.

The application of externally implemented fibrolytic/lignocellulosic enzymes as feed additives with the purpose of improving forage utilisation and advancing ruminant productivity, either by direct delivery to ruminants or via pre-treatment of forage, is a promising area of research (Beauchemin *et al.*, 2003; Yang and Xie, 2010). Studies investigating the effects of pre-treatment of rumen forage with fibrolytic enzymes pre-consumption have shown that increased forage digestibility results in improved voluntary intake, milk production and average daily weight gain in ruminants (Beauchemin *et al.*, 1995; Schingoethe *et al.*, 1999; Kung *et al.*, 2000; Yang *et al.*, 2000; Pinos-Rodríguez *et al.*, 2002; Titi and Lubbadeh, 2004; Cruywagen and Van Zyl, 2008). *In vivo* sheep trials investigating the effects of direct administration of fibrolytic enzyme mixtures have shown positive effects such as improved forage degradation and stimulation of cellulolytic bacterial growth (Giraldo *et al.*, 2008). However, these results have been difficult to replicate and adverse effects have also been observed in similar studies (ZoBell *et al.*, 2000; Peters *et al.*, 2010).

The fibrolytic enzymes currently available for commercial use in ruminant applications are developed based on their ability to breakdown plant cell wall polysaccharides *in vitro*, and are raw extracts composed of mixtures containing xylanases, cellulases or β-glucanases (Colombatto *et al.*, 2003; Yang and Xie, 2010). Despite these advancements there is a need for improvement and this can be achieved by focusing on the specific formulations and purpose-driven selection of fibrolytic enzymes. Examples of viable enzyme candidates for commercial development should incorporate the potentially rate-limiting hydrolysis of side-chains, such as ferulic acid esterases, acetyl-xylan esterases and arabinofuranosidases. The results presented in this thesis provide a description of the diverse classes of CAZymes that are utilised by the abundant polysaccharide-degrading *Butyrivibrio* and *Pseudobutyrivibrio* rumen bacteria and are available for future mining/development of targeted multi-enzyme products for enhancing forage cell wall degradation in ruminants.

Appendices

Appendix I

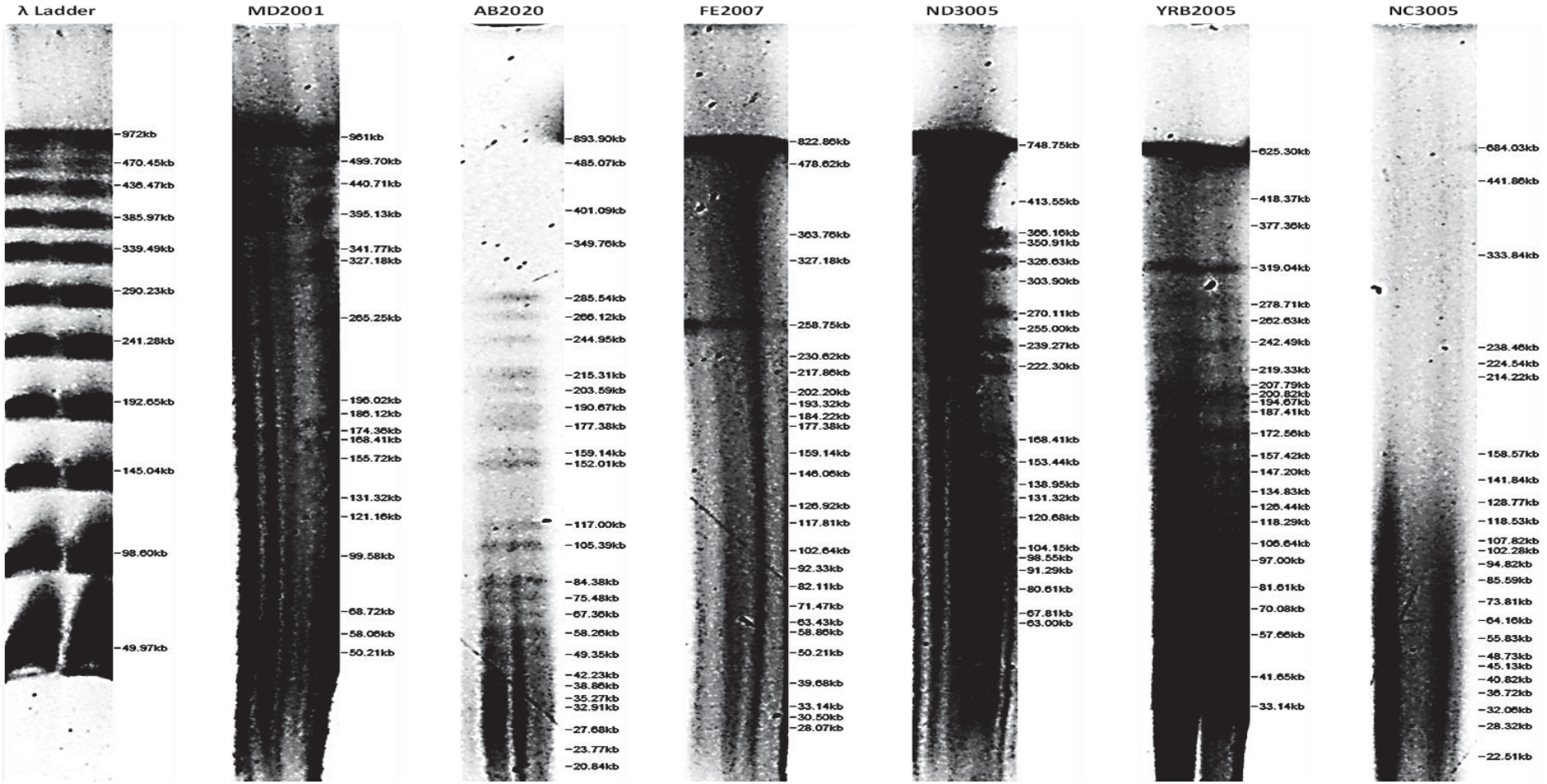


Figure A1.1. PFGE *Apa*I restriction profiles of *B. fibrisolvens* genomic DNAs. Genomic DNAs were digested with *Apa*I and separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

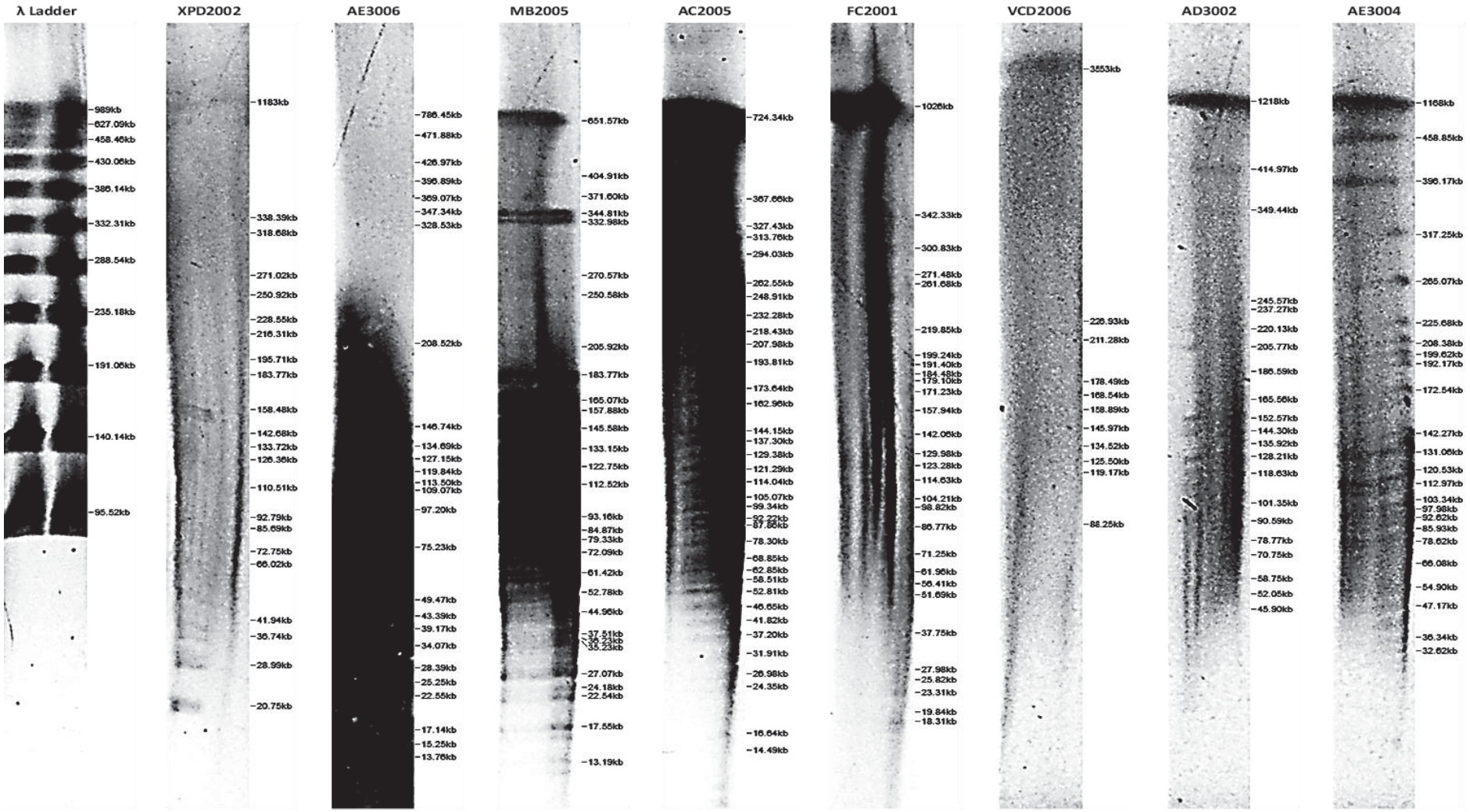


Figure A1.2. PFGE *Apa*I restriction profiles of *Butyrivibrio* sp. 2 genomic DNAs. Genomic DNAs were digested with *Apa*I and separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

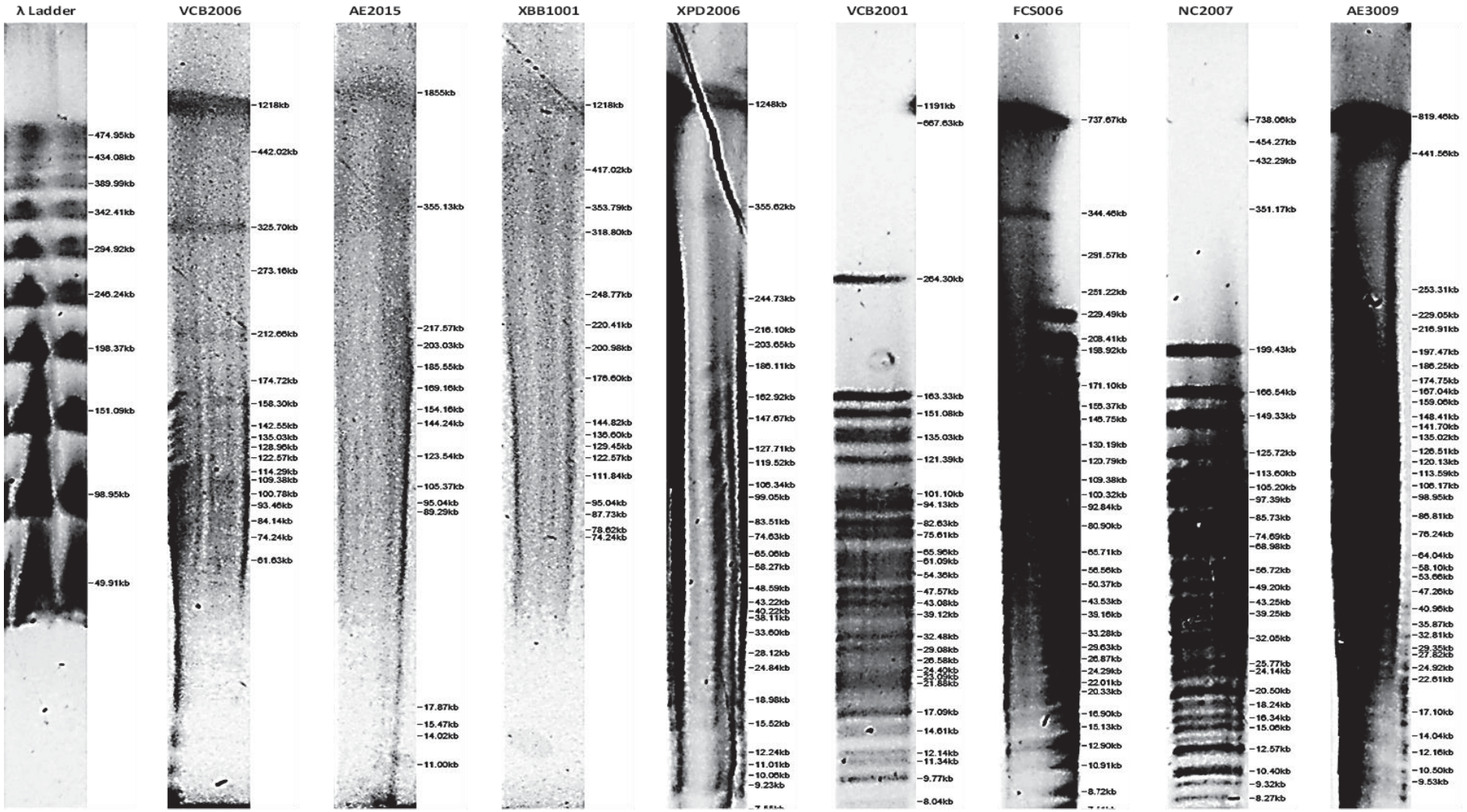


Figure A1.3. PFGE *Apa*I restriction profiles of *Butyrivibrio* sp. 3 genomic DNAs. Genomic DNAs were digested with *Apa*I and separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

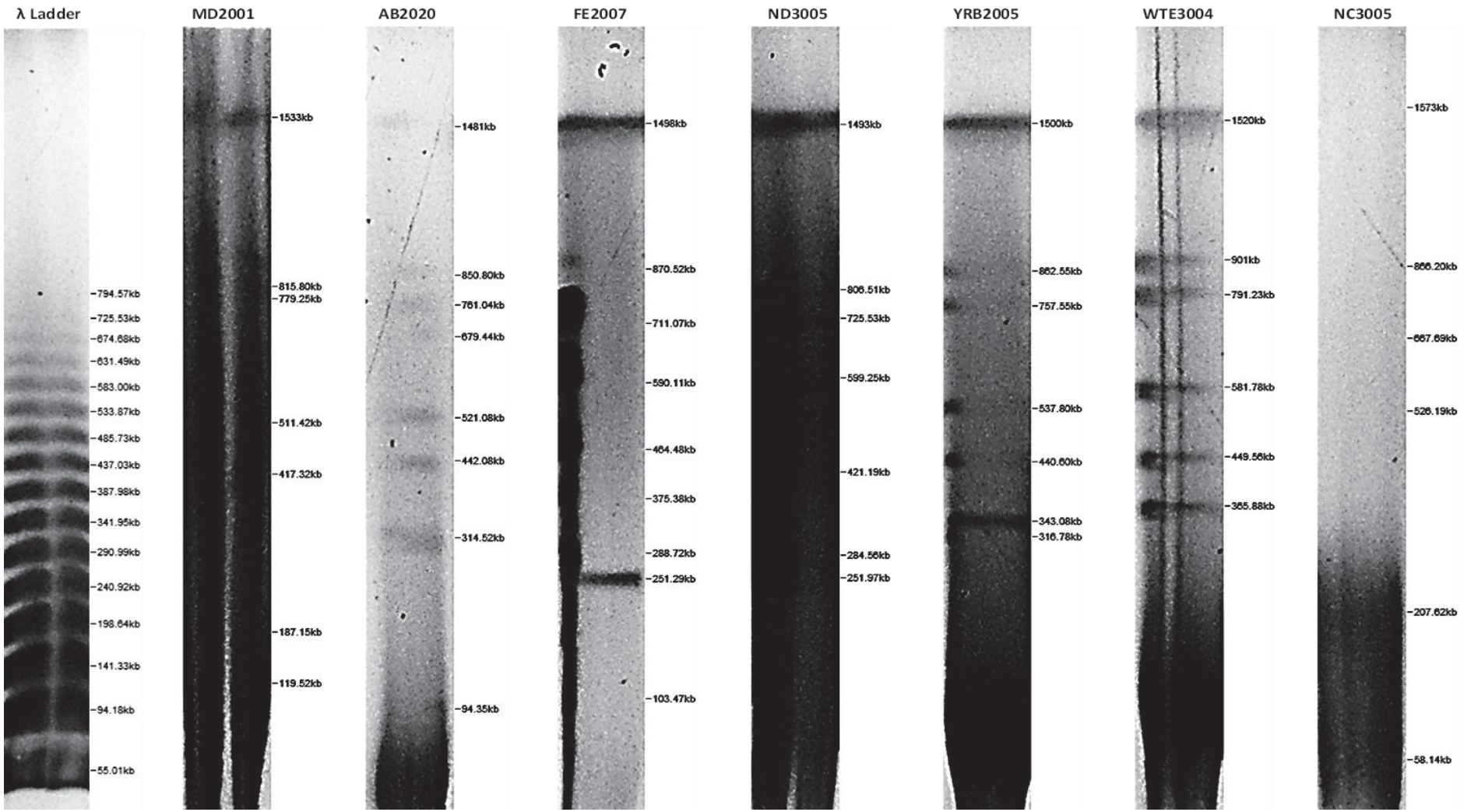


Figure A1.4. PFGE I-*Ceu*I restriction profiles of *B. fibrisolvens* genomic DNAs. Genomic DNAs were digested with I-*Ceu*I and separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

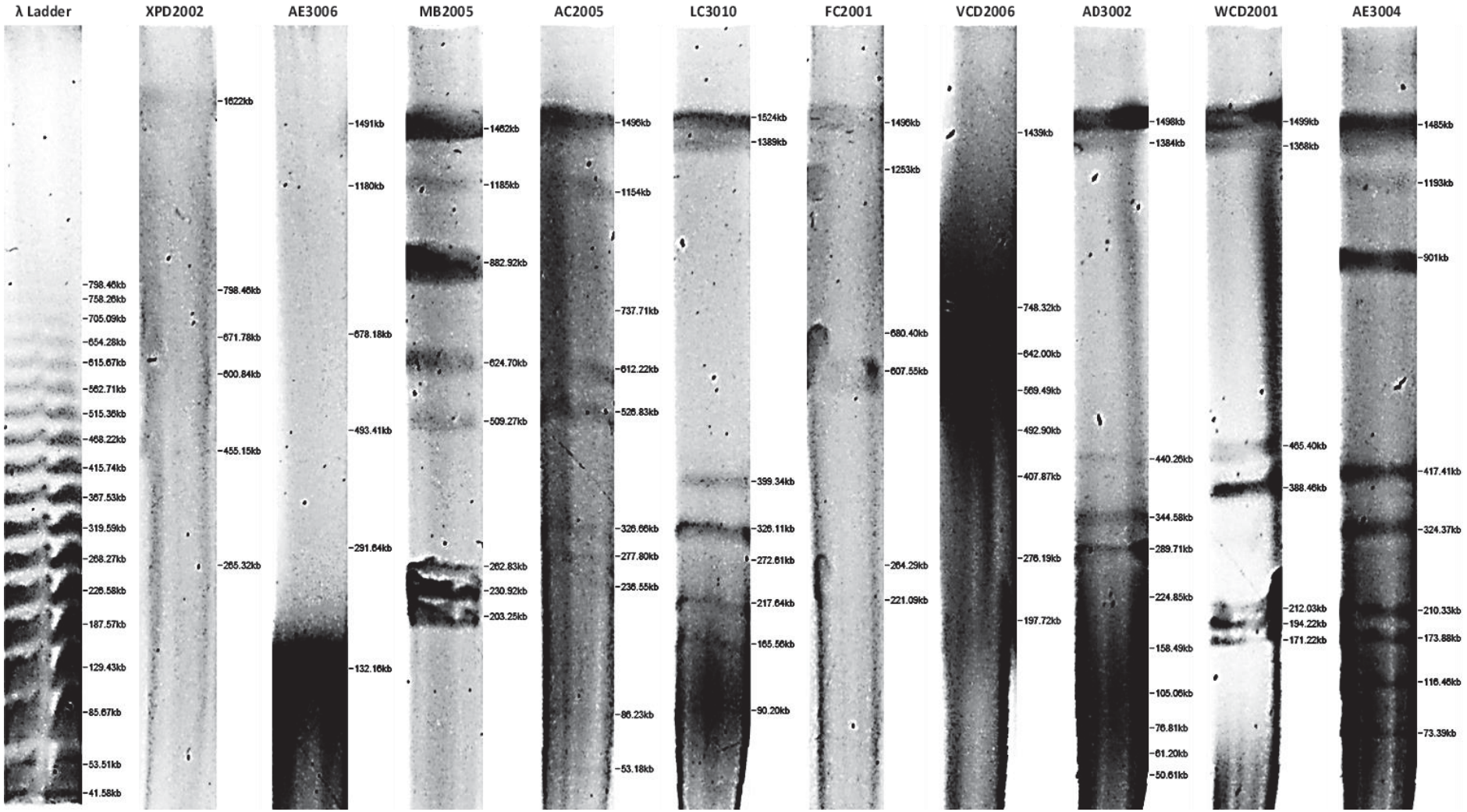


Figure A1.5. PFGE I-*Ceu*I restriction profiles of *Butyrivibrio* sp. 2 genomic DNAs. Genomic DNAs were digested with I-*Ceu*I and separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

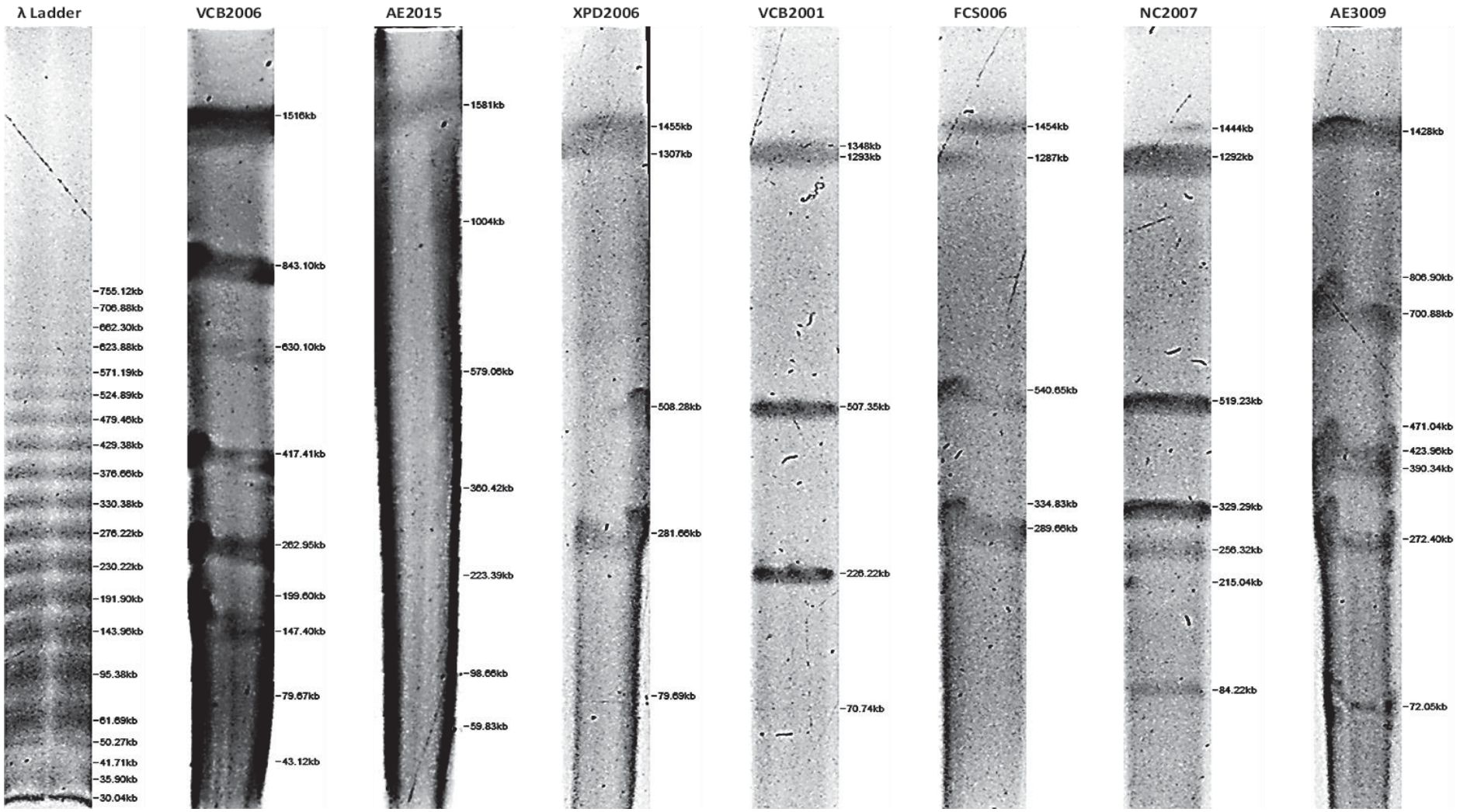


Figure A1.6. PFGE I-*Ceu*I restriction profiles of *Butyrivibrio* sp. 3 genomic DNAs. Genomic DNAs were digested with I-*Ceu*I and separated by PFGE in a 1% (w/v) agarose gel for 20 h, with the pulse time ramped from 1 to 30 sec.

Table A1.1. Summary of the genomic properties based on PFGE.

		Estimated sizes of replicons from undigested genomic DNAs (Kb)					Estimated genome sizes (Kb)	
Species	**Strain**	**Chromosome**	**Extrachromosomal elements**				***Apa*I digest**	**I-*Ceu*I digest**
B. fibrisolvens	MD2001	1,309	522	415	391		4,640	4,364
	AB2020	934	328				4,803	5,144
	FE2007	996	268				4,659	5,153
	ND3005	1,046	264	234			4,715	4,582
	YRB2005	934	339				4,879	4,758
	WTE3004	764	349				N/A	4,610
	MC2013	N/A	N/A				N/A	N/A
	NC3005	N/A	N/A				3,523	3,899
Butyrivibrio sp. 2	XPD2002	1,721	869				4,304	4,414
	AE3006	N/A	99				4,547	4,266
	MB2005	869	139				4,556	5,354
	AC2005	932	507	327	285	107	5,320	5,652
	LC3010	1,070	407	338	105		N/A	4,385
	FB3011	N/A	N/A				N/A	N/A
	FC2001	992	N/A				4,699	4,522
	WCD3002	731	N/A				N/A	N/A
	VCD2006	2,639	356				5,112	4,774
	AD3002	817	370				4,421	4,634
	WCD2001	976	411				N/A	4,298
	AE3004	847	433	350			4,806	4,895
Butyrivibrio sp. 3	VCB2006	898	287	179			3,972	4,139
	AE2015	3,365	N/A				3,756	3,906
	XBB1001	1,158	390	165			4,135	N/A
	XPD2006	1,201	329				3,840	3,632
	VCB2001	1,016	N/A				3,589	3,445
	FCS006	1,081	329				3,862	3,906
	NC2007	1,108	311	151			3,551	4,140
	FCS014	N/A	N/A				N/A	N/A
	AE2032	N/A	N/A				N/A	N/A
	AE3009	1,644	548				4,499	4,566

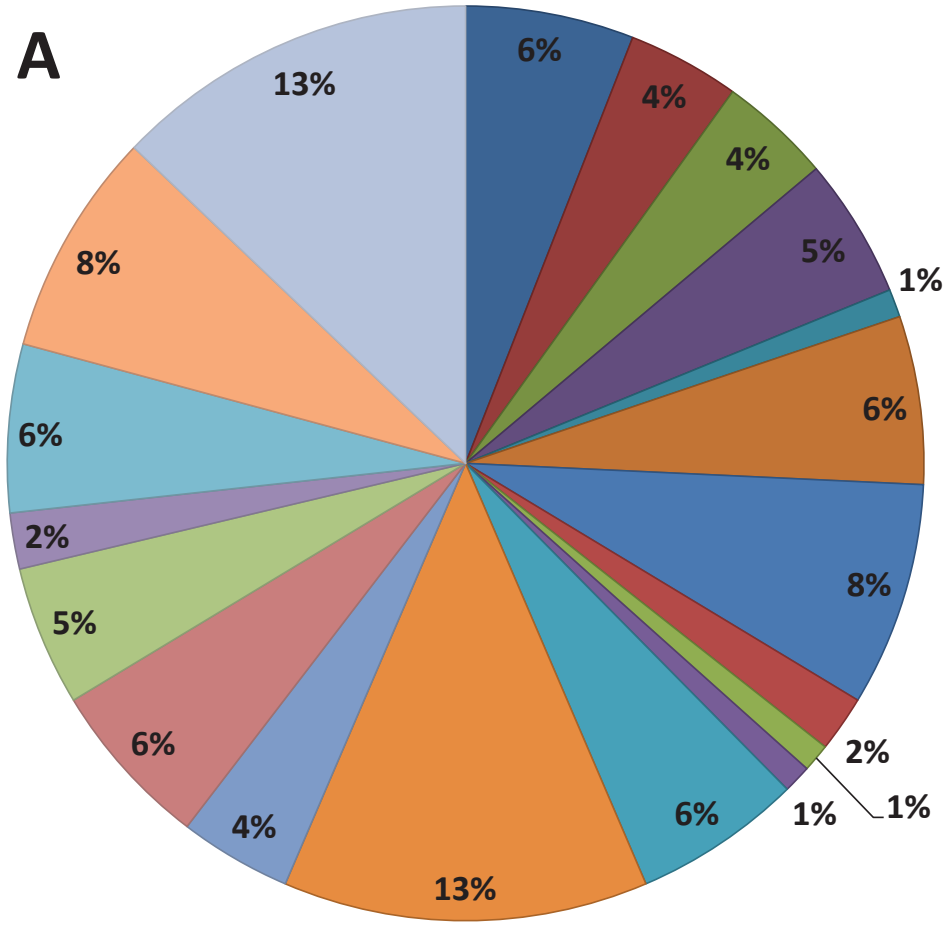
Appendix II

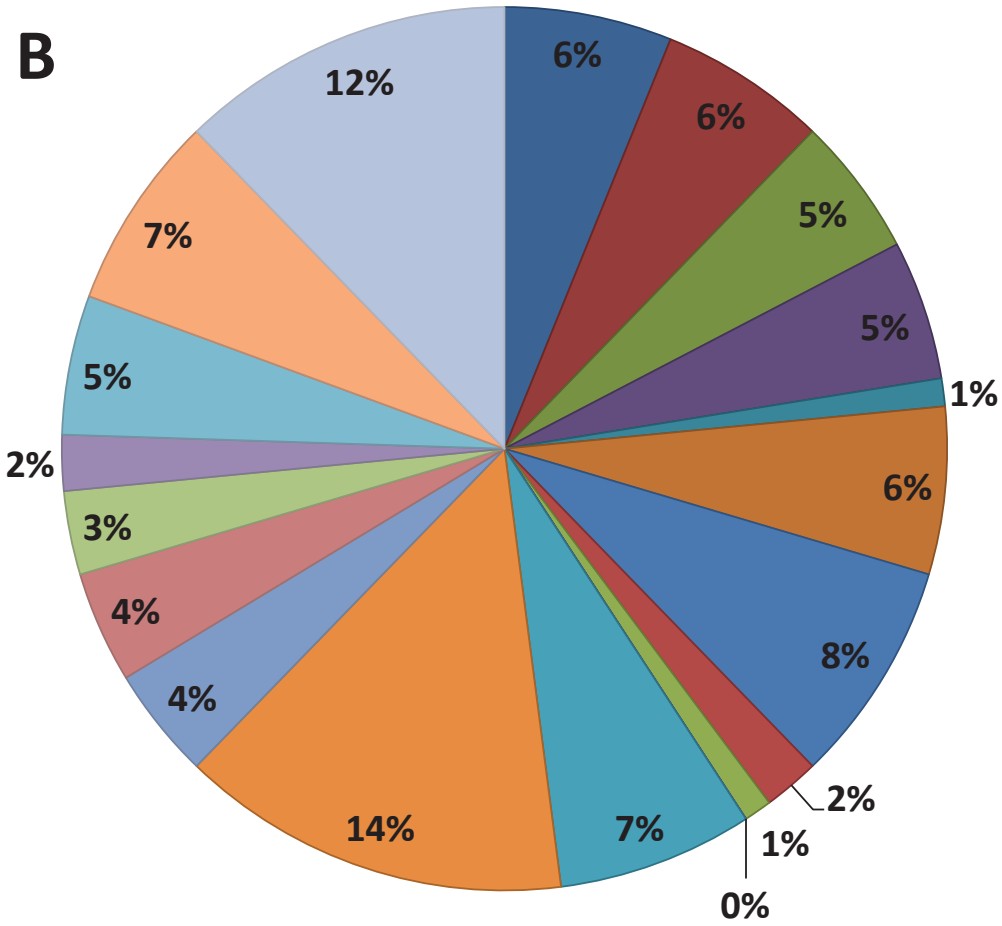
Table A2.1. Genome annotation features of *Butyrivibrio* and *Pseudobutyrivibrio* genomes.

Species	Strain	COG count	COG (%)	Pfam count	Pfam (%)	TIGRfam count	TIGRfam (%)	Enzyme count	Enzyme (%)	SP count [A]	SP (%) [A]	TM count [B]	TM (%) [B]
B. fibrisolvens	MD2001	2,304	57	3,041	76	965	24	791	20	281	7	1,128	28
	AB2020	2,301	56	3,065	75	996	24	806	20	270	7	1,098	27
	FE2007	2,112	57	2,780	74	940	25	759	20	243	7	1,013	27
	ND3005	2,089	54	2,796	72	945	24	757	20	249	6	1,042	27
	YRB2005	2,278	57	2,995	75	981	25	789	20	262	7	1,100	28
	WTE3004	2,310	58	3,025	76	982	25	785	20	276	7	1,084	27
	MC2013	1,938	59	2,534	77	863	26	676	21	218	7	845	26
	NC3005	1,998	59	2,621	77	893	26	709	21	188	6	890	26
Butyrivibrio sp. 2	AE3006	2,172	59	2,898	79	951	26	743	20	288	8	1,002	27
	MB2005	2,171	59	2,923	79	957	26	751	20	278	8	1,009	27
	WCD3002	2,166	58	2,930	78	939	25	728	19	306	8	1,037	28
	AD3002	2,196	56	2,990	76	956	24	724	18	366	9	1,132	29
	WCD2001	2,200	56	2,952	76	958	25	730	19	366	9	1,138	29
	VCD2006	2,328	57	3,173	77	1,004	24	753	18	350	9	1,171	28
	AE3004	2,239	55	3,048	75	978	24	750	19	308	8	1,108	27
	LC3010	2,277	56	3,076	75	949	23	752	18	317	8	1,144	28
	WCE2006	2,278	57	3,030	76	933	23	742	19	215	5	1,117	28
	AC2005	2,494	52	3,456	73	1040	22	806	17	373	8	1,275	27
	FC2001	1,723	41	3,143	76	993	24	766	18	383	9	1,225	29
	XPD2002	2,289	57	3,050	75	975	24	752	19	373	9	1,168	29
	NC2002	1,839	61	2,364	78	854	28	665	22	137	5	823	27

Butyrivibrio sp. 3	NK4A153	1,875	59	2,395	76	875	28	675	21	235	7	872	28
	AE2005	1,944	56	2,537	73	901	26	677	20	266	8	969	28
	AE3003	1,880	58	2,437	75	873	27	674	21	243	7	892	27
	LB2008	1,978	57	2,569	74	905	26	696	20	260	7	995	29
	MB2003 *	1,842	60	2,353	76	858	28	685	22	148	5	881	29
	AE2015	2,022	62	2,560	78	885	27	711	22	214	7	918	28
	P6B7	1,986	56	2,601	73	895	25	715	20	175	5	981	28
	B316 *	2,075	53	2,789	72	945	24	783	20	269	7	1,061	27
	FD2007	2,001	60	2,604	78	914	27	728	22	242	7	919	28
	VCB2006	2,059	59	2,688	77	906	26	725	21	248	7	1,002	29
	XBB1001	2,002	58	2,623	76	905	26	726	21	246	7	963	28
	AE3009	2,098	58	2,758	76	920	25	726	20	305	8	1,000	28
	MC2021	2,109	58	2,823	77	925	25	729	20	321	9	1,034	28
	XPD2006	2,148	61	2,772	79	941	27	746	21	277	8	1,041	30
	VCB2001	2,191	60	2,842	78	940	26	756	21	265	7	1,020	28
	FCS006	2,111	63	2,645	79	935	28	736	22	257	8	1,015	30
	NC2007	2,157	59	2,793	77	938	26	751	21	272	8	1,046	29
	FCS014	2,053	56	2,780	75	971	26	708	19	240	6	978	26
	AE2032	1,994	59	2,562	76	884	26	673	20	213	6	898	27
Pseudobutyrivibrio	MA3014	1,879	58	2,491	78	864	27	716	22	153	5	804	25
	HUN009	1,656	64	2,066	80	821	32	648	25	122	5	700	27
	CF1b	1,867	62	2,390	80	885	30	713	24	170	6	811	27
	AD2017	1,672	63	2,115	80	834	32	642	24	137	5	694	26
	LB2011	1,712	65	2,110	80	847	32	666	25	120	5	708	27
	MD2005	1,744	61	2,180	76	871	30	679	24	129	5	729	25

* indicates strains that have closed genomes. The genomes of all other strains are in draft status. [A], SP represents signal peptides; [B], TM represents transmembrane.


A
6%
4%
4%
5%
1%
6%
8%
2%
1%
1%
6%
13%
4%
6%
5%
2%
6%
8%


B
6%
6%
5%
5%
1%
6%
8%
2%
1%
0%
7%
14%
4%
4%
3%
2%
5%
7%
12%

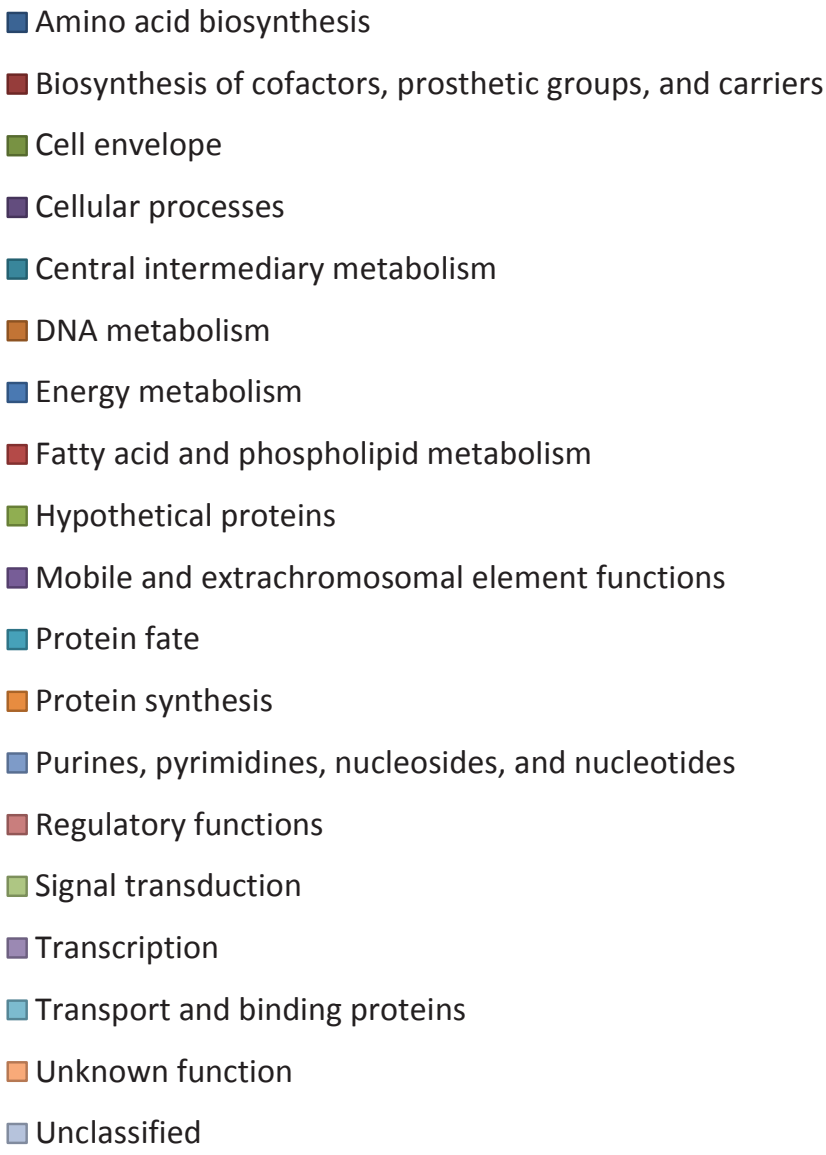
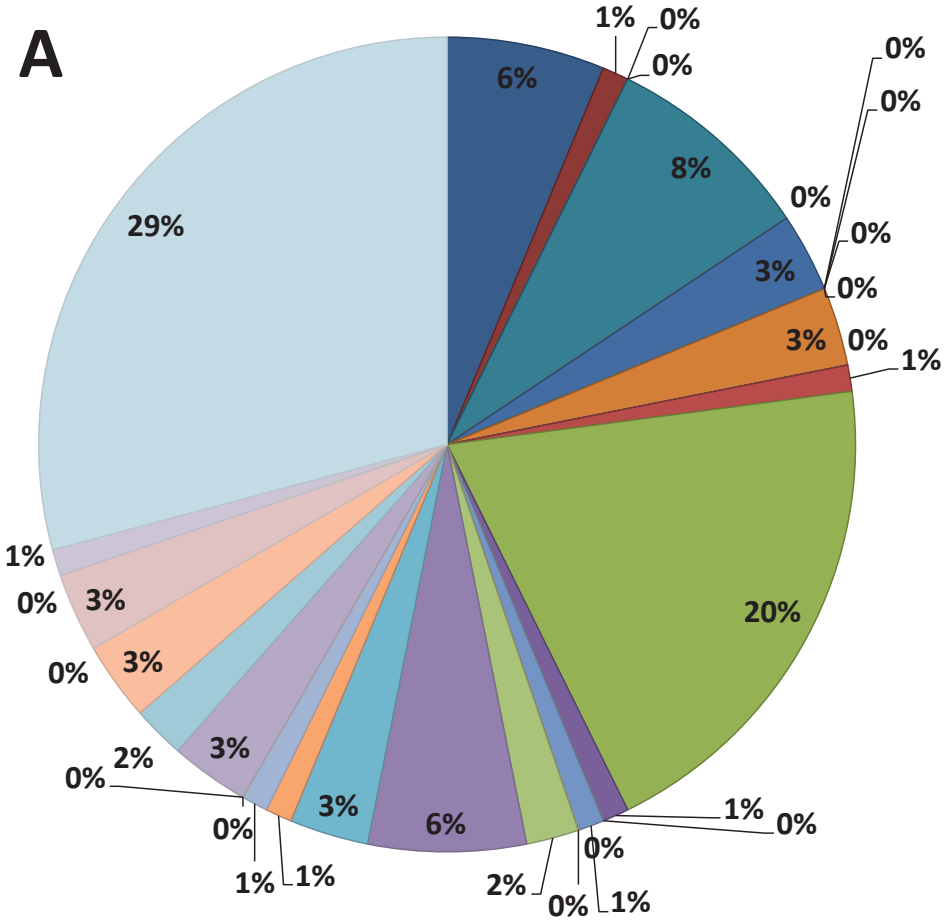
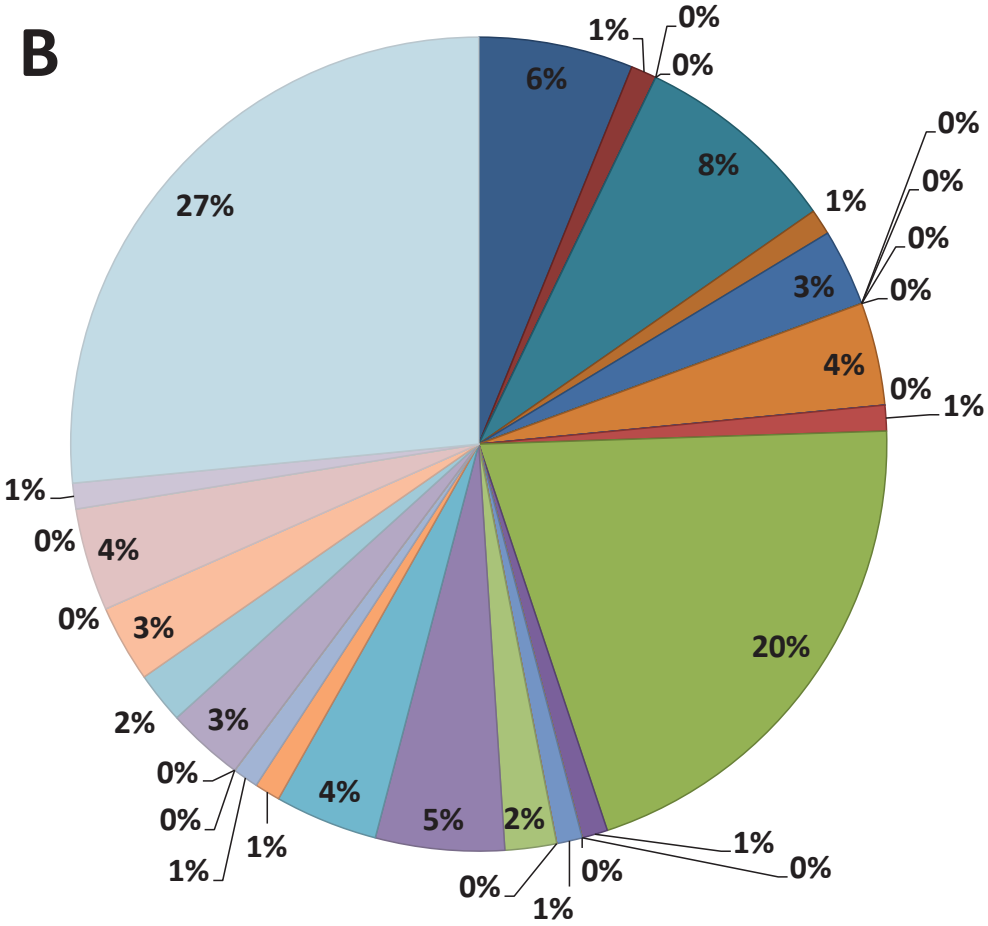


Figure A2.1. TIGRfam analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes. A, *Butyrivibrio* genomes; B, *Pseudobutyrivibrio* genomes.


A
6%
1%
0%
0%
8%
0%
0%
0%
3%
0%
3%
0%
1%
20%
1%
0%
3%
3%
0%
2%
3%
0%
3%
6%
1%
1%
2%
0%
1%
0%
0%
1%
0%


B
6%
1%
0%
0%
8%
0%
0%
1%
0%
0%
3%
4%
0%
1%
20%
1%
0%
4%
0%
3%
2%
3%
0%
4%
0%
1%
1%
5%
2%
0%
1%
1%
0%
0%

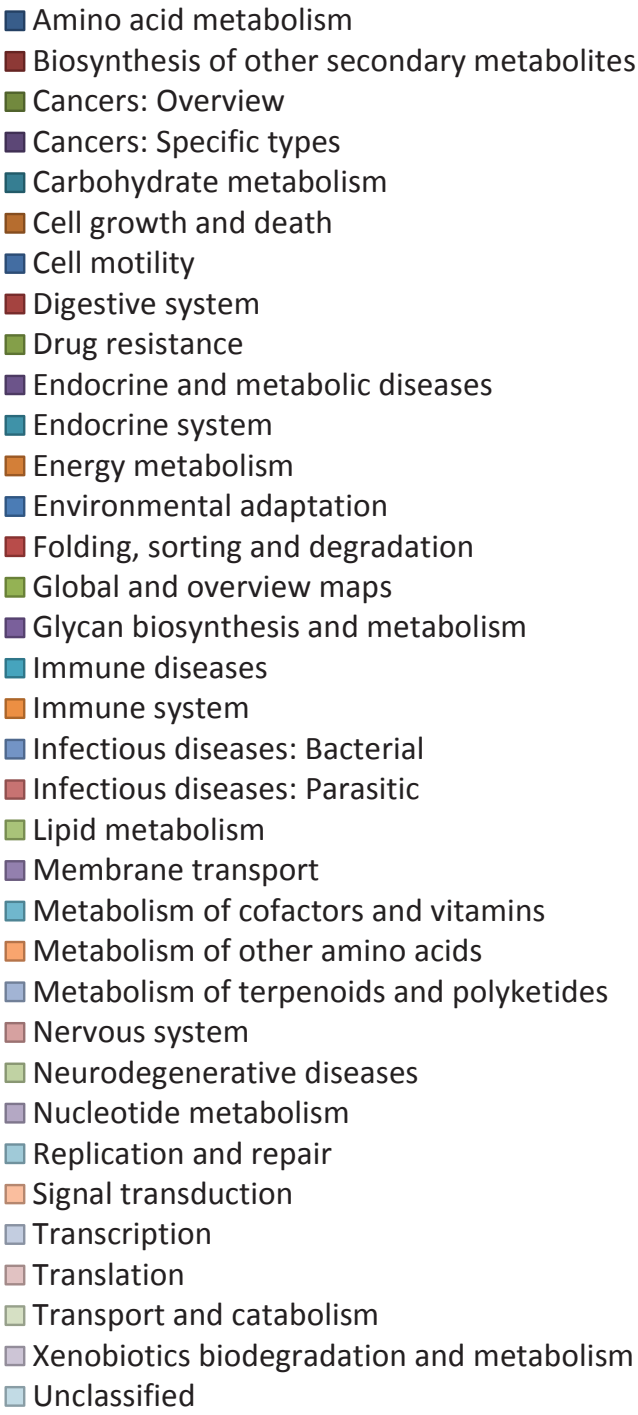


Figure A2.2. KEGG pathway analyses of *Butyrivibrio* and *Pseudobutyrivibrio* genomes. A, *Butyrivibrio* genomes; B, *Pseudobutyrivibrio* genomes.

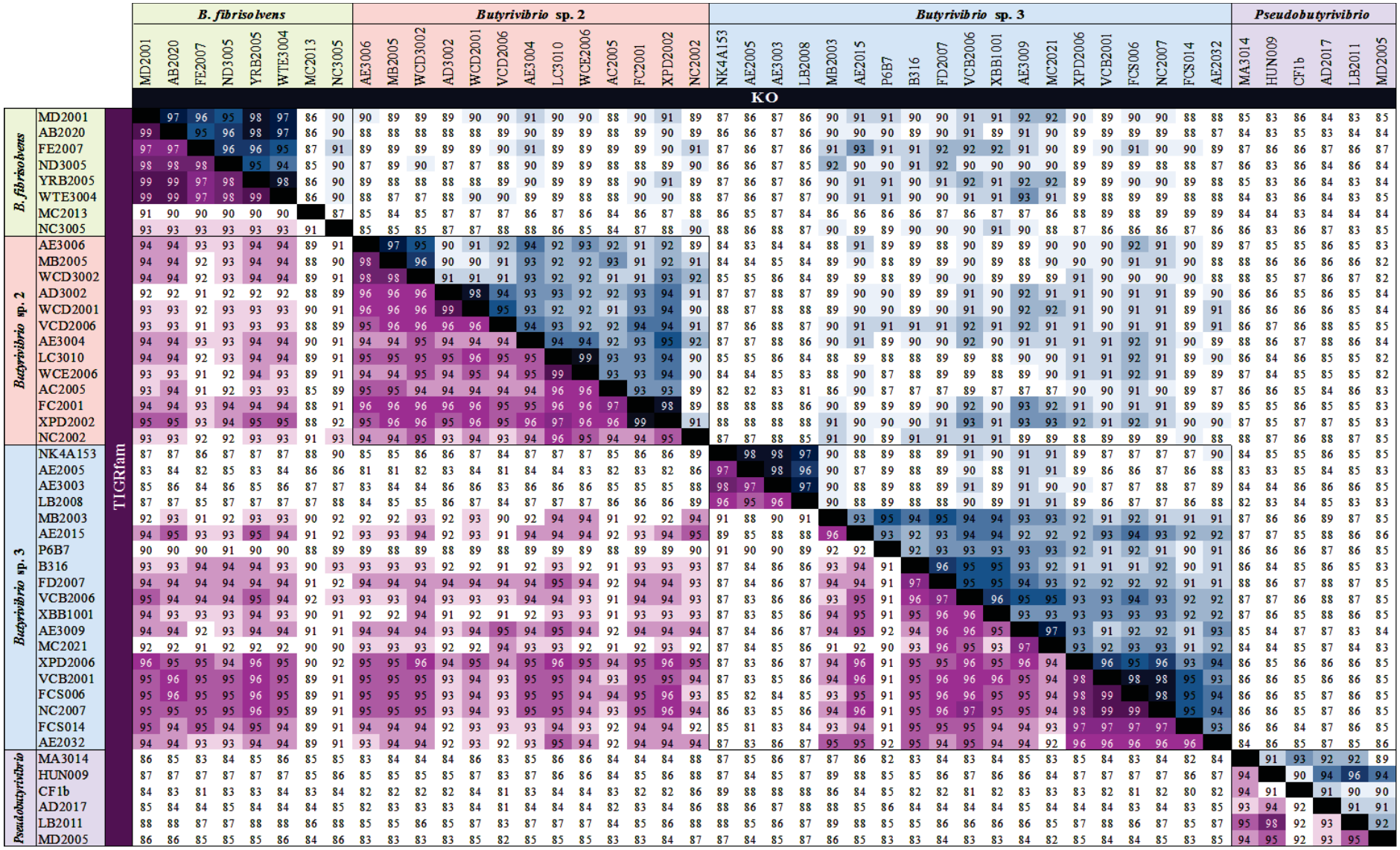

B. fibrisolvens
Butyrivibrio sp. 2
Butyrivibrio sp. 3
Pseudobutyrivibrio
KO
TIGRfam
MD2001 AB2020 FE2007 ND3005 YRB2005 WTE3004 MC2013 NC3005
AE3006 MB2005 WCD3002 AD3002 WCD2001 VCD2006 AE3004 LC3010 WCE2006 AC2005 FC2001 XPD2002 NC2002
NK4A153 AE2005 AE3003 LB2008 MB2003 AE2015 P6B7 B316 FD2007 VCB2006 XBB1001 AE3009 MC2021 XPD2006 VCB2001 FCS006 NC2007 FCS014 AE2032
MA3014 HUN009 CF1b AD2017 LB2011 MD2005

Figure A2.3. Genome clustering comparison of TIGRfam and KO domains from *Butyrivibrio* and *Pseudobutyrivibrio*. Each cell of the matrix displays the correlation coefficient between the genomes on the corresponding row and column. The correlation coefficient is computed based on the similarity of the functional characterization of the genomes. For this analysis the correlation coefficients have been converted to percentages. The diagonal correlations of the genomes with themselves are always 1.00 or 100%, and are thus left as black squares. The 46 genomes have been arranged in the order in which they are grouped in the FGD analysis to provide reference points for comparisons. The three *Butyrivibrio* and one *Pseudobutyrivibrio* designated cluster groups have been outlined.

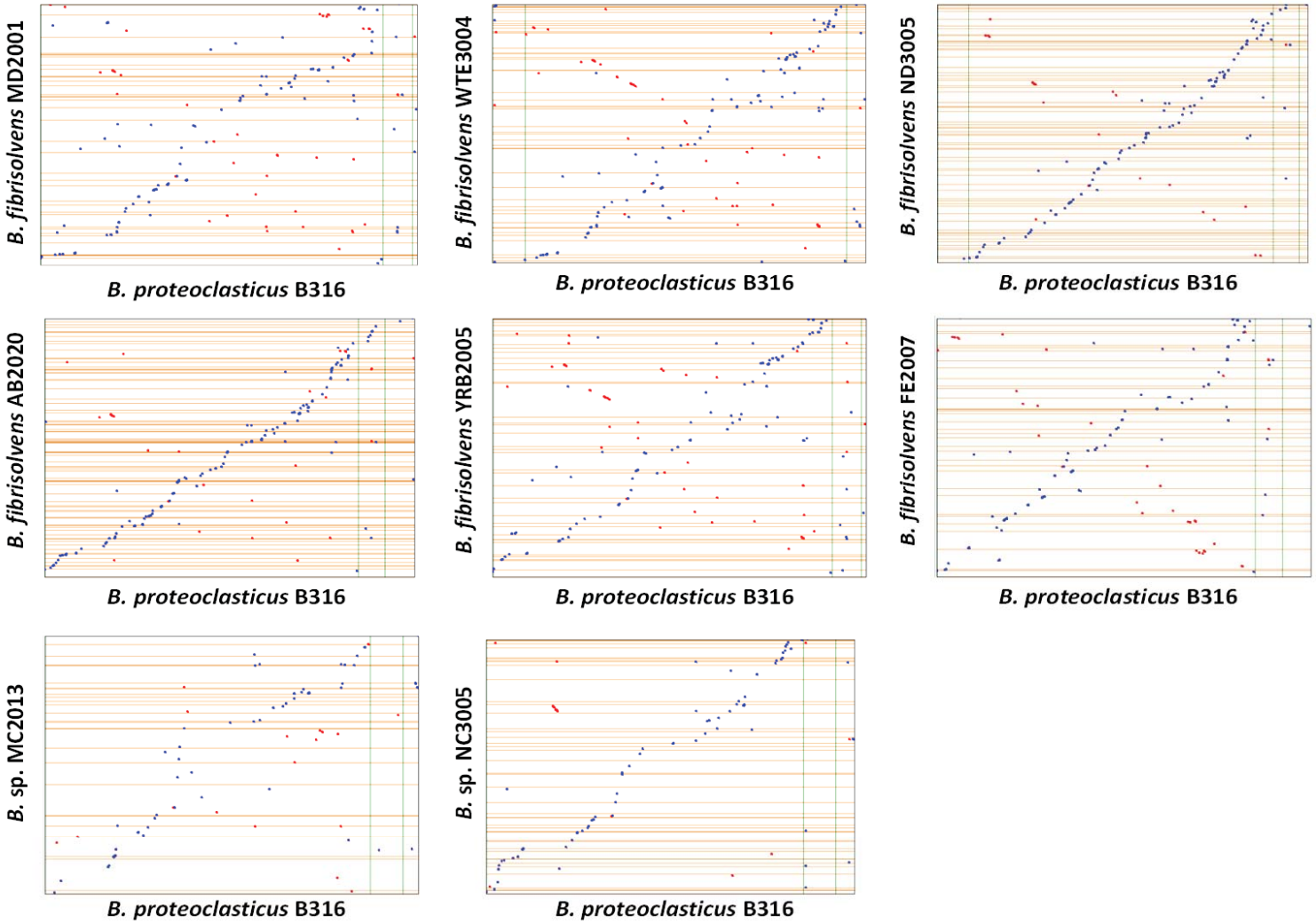


Figure A2.4. MUMmer plots of *B. fibrisolvens* draft genomes versus *B. proteoclasticus* B316 as the reference genome. Colour codes: blue, forward sequence, red, reverse sequence.

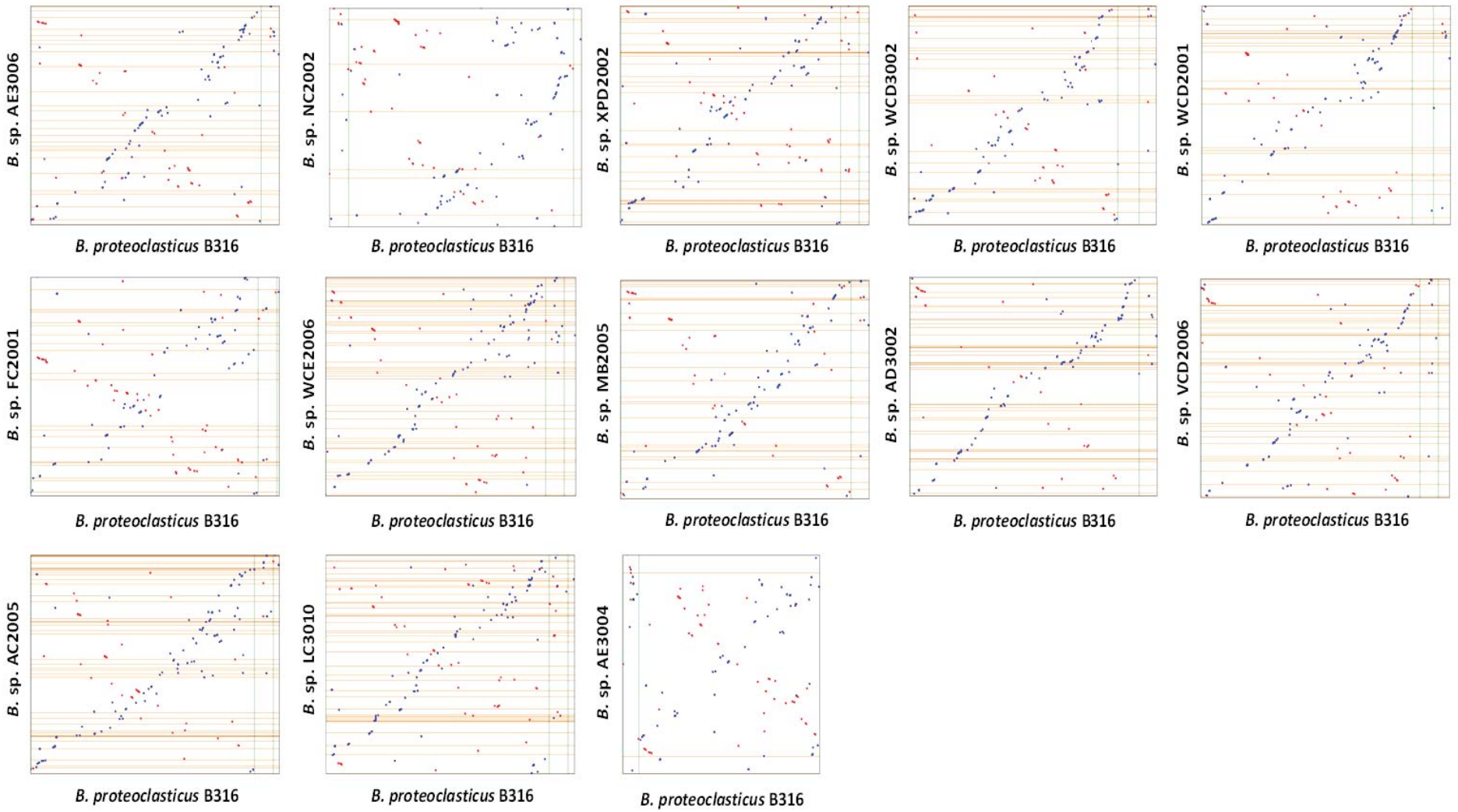


Figure A2.5. MUMmer plots of *Butyrivibrio* sp. 2 draft genomes versus *B. proteoclasticus* B316 as the reference genome. Colour codes: blue, forward sequence, red, reverse sequence.

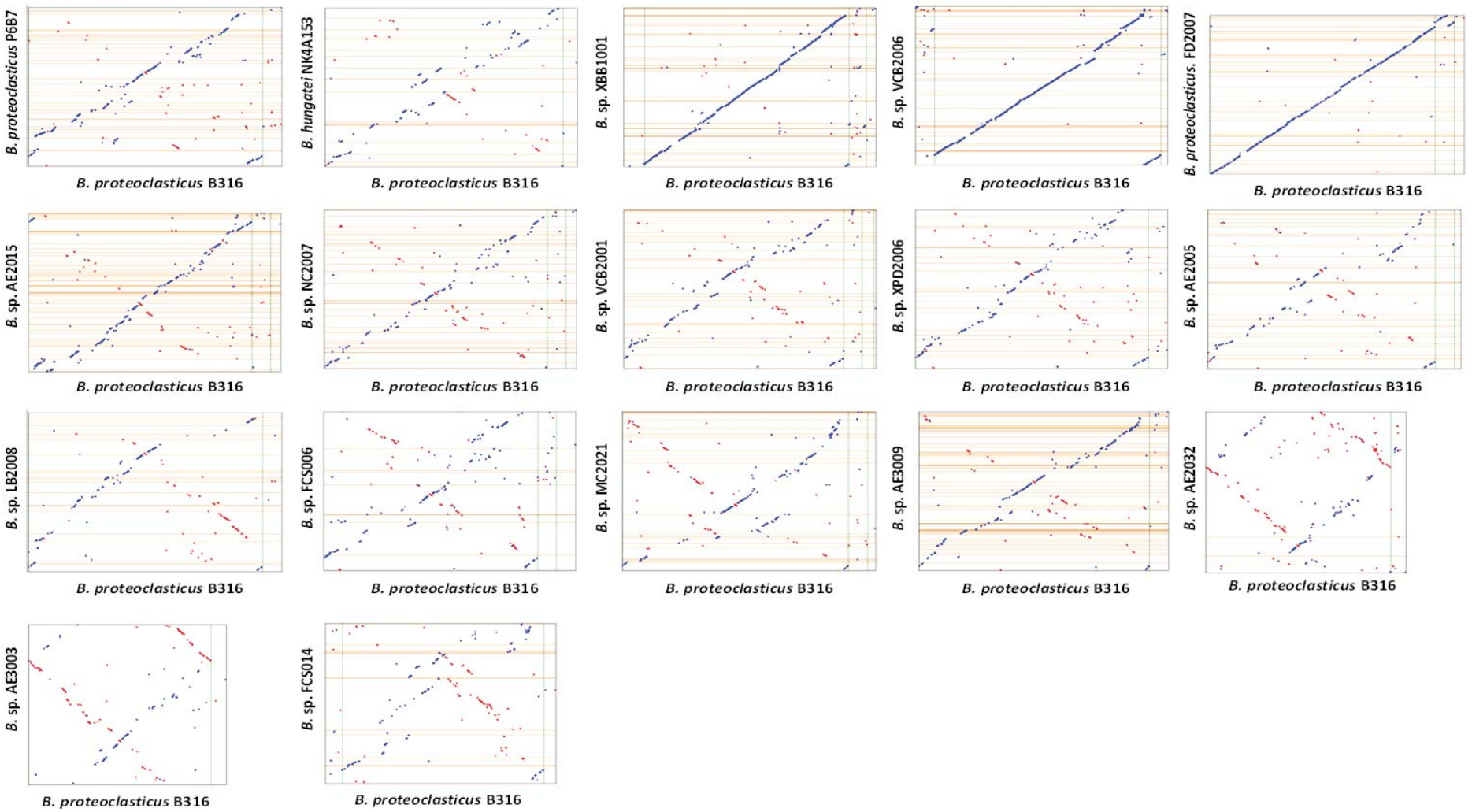


Figure A2.6. MUMmer plots of *Butyrivibrio* sp. 3 draft genomes versus *B. proteoclasticus* B316 as the reference genome. Colour codes: blue, forward sequence, red, reverse sequence.

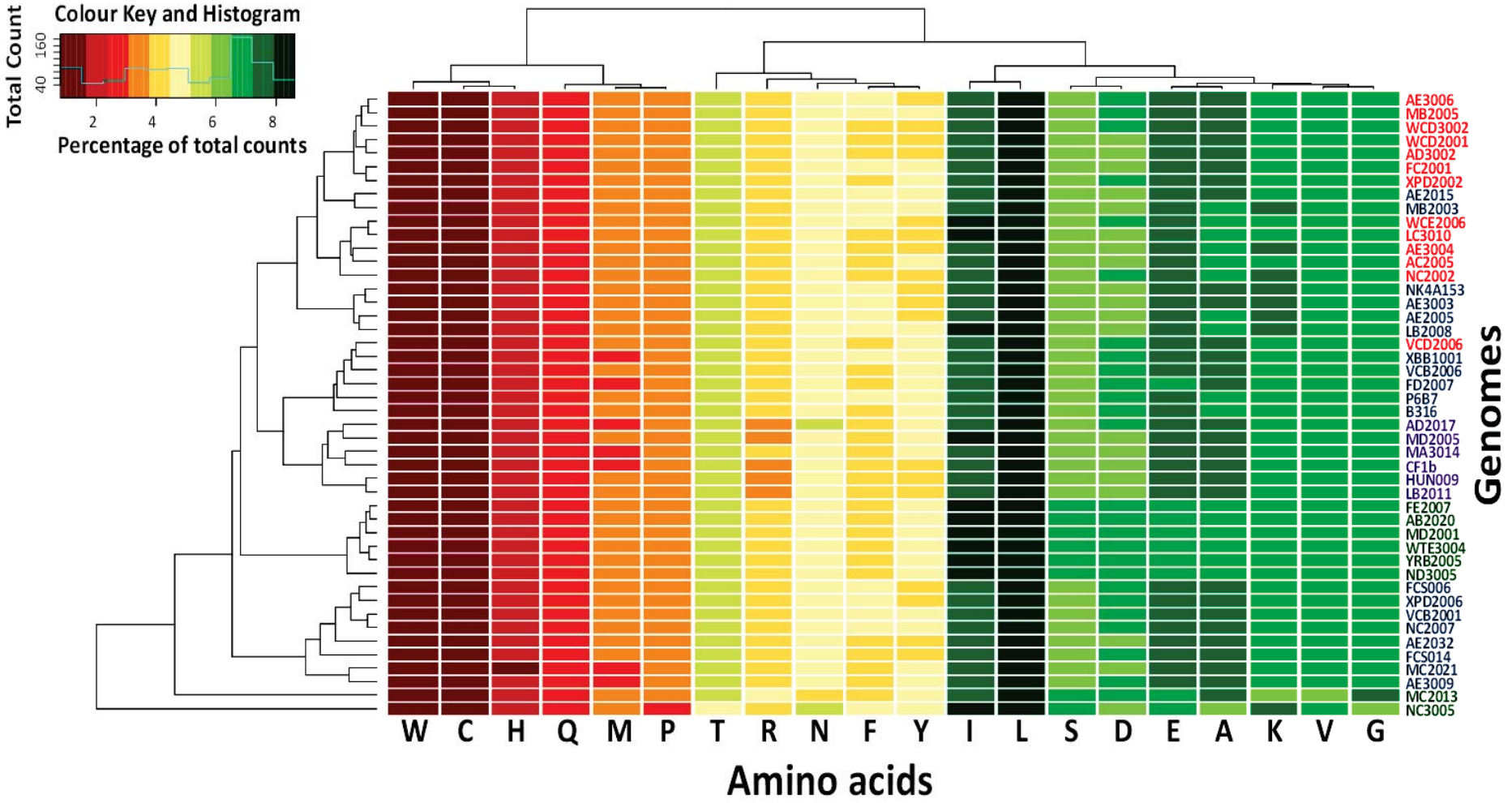


Figure A2.7. Amino acid usage heatmap of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* based on their protein content. Cluster grouping was based on comparison of the ORFeomes of 46 genomes by FGD analysis. Genomes are coloured to represent *B. fibrisolvens* in green, *Butyrivibrio* sp. 2 in red, *Butyrivibrio* sp. 3 in blue and *Pseudobutyrivibrio* in purple.

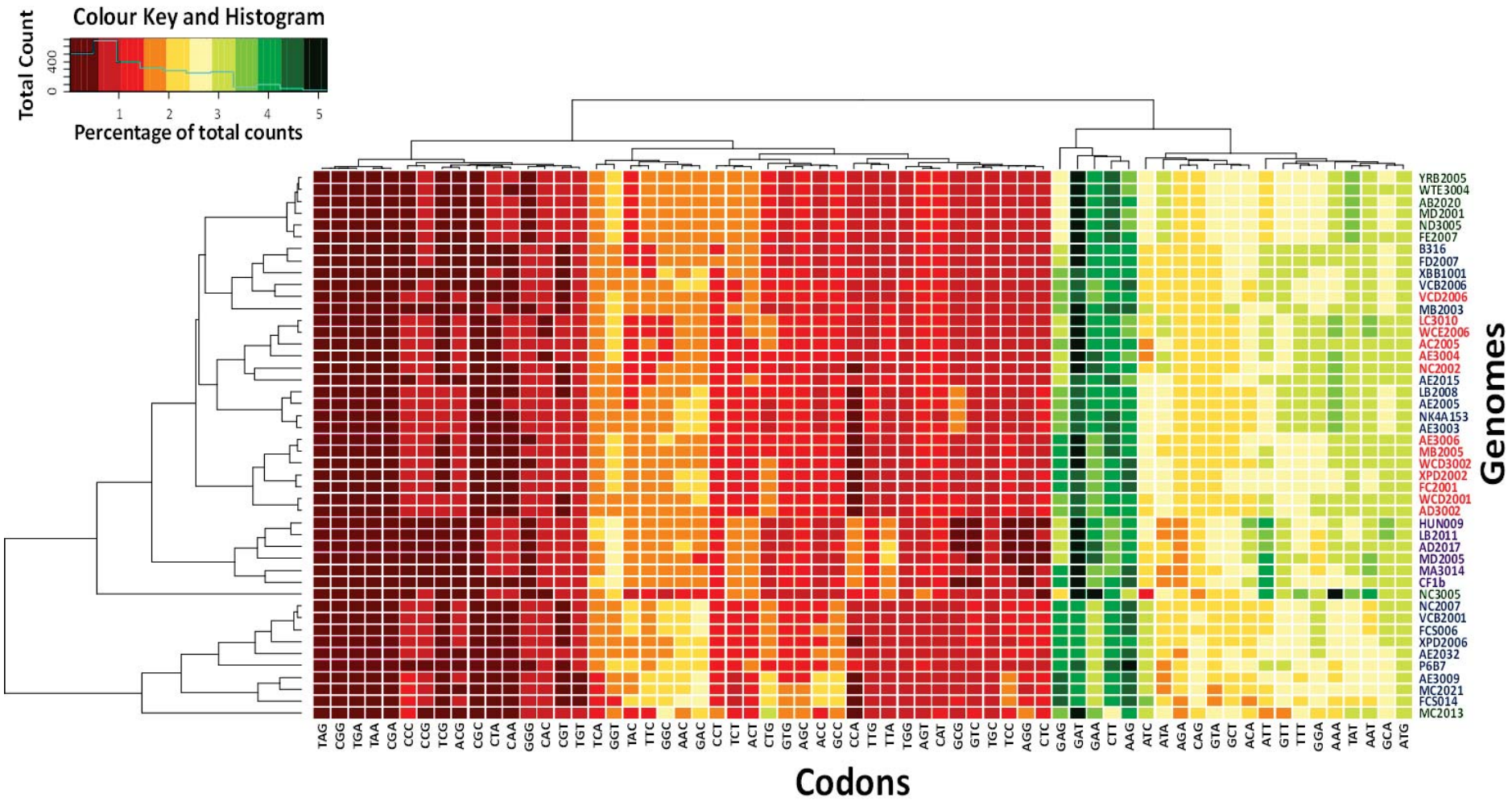


Figure A2.8. Codon usage heatmap of the 40 *Butyrivibrio* and 6 *Pseudobutyrivibrio* based on their protein content. Cluster grouping was based on comparison of the ORFeomes of 46 genomes by FGD analysis. Genomes are coloured to represent *B. fibrisolvens* in green, *Butyrivibrio* sp. 2 in red, *Butyrivibrio* sp. 3 in blue and *Pseudobutyrivibrio* in purple.

Table A2.2. CAZy families analysed and their known activities.

Family	Domain *	CAZy Activities
GH1	Pfam00232	β-glucosidase (EC 3.2.1.21); β-galactosidase (EC 3.2.1.23); β-mannosidase (EC 3.2.1.25); β-glucuronidase (EC 3.2.1.31); β-D-fucosidase (EC 3.2.1.38); phlorizin hydrolase (EC 3.2.1.62); exo-β-1,4-glucanase (EC 3.2.1.74); 6-phospho-β-galactosidase (EC 3.2.1.85); 6-phospho-β-glucosidase (EC 3.2.1.86); strictosidine β-glucosidase (EC 3.2.1.105); lactase (EC 3.2.1.108); amygdalin β-glucosidase (EC 3.2.1.117); prunasin β-glucosidase (EC 3.2.1.118); raucaffricine β-glucosidase (EC 3.2.1.125); thioglucosidase (EC 3.2.1.147); β-primeverosidase (EC 3.2.1.149); isoflavonoid 7-O-β-apiosyl-β-glucosidase (EC 3.2.1.161); hydroxyisourate hydrolase (EC 3.-.-.-); β-glycosidase (EC 3.2.1.-).
GH2	Pfam00703	β-galactosidase (EC 3.2.1.23) ; β-mannosidase (EC 3.2.1.25); β-glucuronidase (EC 3.2.1.31); mannosylglycoprotein endo-β-mannosidase (EC 3.2.1.152); exo-β-glucosaminidase (EC 3.2.1.165).
GH3	Pfam00933	β-glucosidase (EC 3.2.1.21); xylan 1,4-β-xylosidase (EC 3.2.1.37); β-N-acetylhexosaminidase (EC 3.2.1.52); glucan 1,3-β-glucosidase (EC 3.2.1.58); glucan 1,4-β-glucosidase (EC 3.2.1.74); exo-1,3-1,4-glucanase (EC 3.2.1.-); αlpha-L-arabinofuranosidase (EC 3.2.1.55).
GH4	Pfam02056	maltose-6-phosphate glucosidase (EC 3.2.1.122); α-glucosidase (EC 3.2.1.20); α-galactosidase (EC 3.2.1.22); 6-phospho-β-glucosidase (EC 3.2.1.86); α-glucuronidase (EC 3.2.1.139).
GH5	Pfam00150	chitosanase (EC 3.2.1.132); β-mannosidase (EC 3.2.1.25); Cellulase (EC 3.2.1.4); glucan 1,3-β-glucosidase (EC 3.2.1.58); licheninase (EC 3.2.1.73); glucan endo-1,6-β-glucosidase (EC 3.2.1.75); mannan endo-β-1,4-mannosidase (EC 3.2.1.78); endo-β-1,4-xylanase (EC 3.2.1.8); cellulose β-1,4-cellobiosidase (EC 3.2.1.91); β-1,3-mannanase (EC 3.2.1.-); xyloglucan-specific endo-β-1,4-glucanase (EC 3.2.1.151); mannan transglycosylase (EC 2.4.1.-); endo-β-1,6-galactanase (EC 3.2.1.164).
GH8	Pfam01270	chitosanase (EC 3.2.1.132); cellulase (EC 3.2.1.4); licheninase (EC 3.2.1.73); endo-1,4-β-xylanase (EC 3.2.1.8); reducing-end-xylose releasing exo-oligoxylanase (EC 3.2.1.156).
GH9	Pfam00759	endoglucanase (EC 3.2.1.4); cellobiohydrolase (EC 3.2.1.91); β-glucosidase (EC 3.2.1.21); exo-β-glucosaminidase (EC 3.2.1.165).
GH10	Pfam00331	endo-1,4-β-xylanase (EC 3.2.1.8); endo-1,3-β-xylanase (EC 3.2.1.32).
GH11	Pfam00457	xylanase (EC 3.2.1.8).
GH12	Pfam01670	endoglucanase (EC 3.2.1.4); xyloglucan hydrolase (EC 3.2.1.151); β-1,3-1,4-glucanase (EC 3.2.1.73); xyloglucan endotransglycosylase (EC 2.4.1.207).
GH13	Pfam00128	α-amylase (EC 3.2.1.1); pullulanase (EC 3.2.1.41); cyclomaltodextrin glucanotransferase (EC 2.4.1.19); cyclomaltodextrinase (EC 3.2.1.54); trehalose-6-phosphate hydrolase (EC 3.2.1.93); oligo-α-glucosidase (EC 3.2.1.10); maltogenic amylase (EC 3.2.1.133); neopullulanase (EC 3.2.1.135); α-glucosidase (EC 3.2.1.20); maltotetraose-forming α-amylase (EC 3.2.1.60); isoamylase (EC 3.2.1.68); glucodextranase (EC 3.2.1.70); maltohexaose-forming α-amylase (EC 3.2.1.98); maltotriose-forming α-amylase (EC 3.2.1.116); branching enzyme (EC 2.4.1.18); trehalose synthase (EC 5.4.99.16); 4-α-glucanotransferase (EC 2.4.1.25); maltopentaose-forming α-amylase (EC 3.2.1.-) ; amylosucrase (EC 2.4.1.4) ; sucrose phosphorylase (EC 2.4.1.7); malto-oligosyltrehalose trehalohydrolase (EC 3.2.1.141); isomaltulose synthase (EC 5.4.99.11); amino acid transporter.
GH15	Pfam00723	glucoamylase (EC 3.2.1.3); glucodextranase (EC 3.2.1.70); α,α-trehalase (EC 3.2.1.28).
GH16	cd00413	xyloglucan:xyloglucosyltransferase (EC 2.4.1.207); keratan-sulfate endo-1,4-β-galactosidase (EC 3.2.1.103); endo-1,3-β-glucanase (EC 3.2.1.39); endo-1,3(4)-β-glucanase (EC 3.2.1.6); licheninase (EC 3.2.1.73); β-agarase (EC 3.2.1.81); kappa;-carrageenase (EC 3.2.1.83); xyloglucanase (EC 3.2.1.151).
GH18	Pfam00704	chitinase (EC 3.2.1.14); endo-β-N-acetylglucosaminidase (EC 3.2.1.96); xylanase inhibitor; concanavalin B; narbonin.

GH20	Pfam00728	β-hexosaminidase (EC 3.2.1.52); lacto-N-biosidase (EC 3.2.1.140); β-1,6-N-acetylglucosaminidase) (EC 3.2.1.-); β-6-SO3-N-acetylglucosaminidase (EC 3.2.1.-).
GH23	cd00254	lysozyme type G (EC 3.2.1.17); peptidoglycan lyase (EC 4.2.2.-) also known in the literature as peptidoglycan lytic transglycosylase.
GH24	Pfam05838	lysozyme (EC 3.2.1.17).
GH25	Pfam01183	lysozyme (EC 3.2.1.17).
GH26	Pfam02156	β-mannanase (EC 3.2.1.78); β-1,3-xylanase (EC 3.2.1.32).
GH27	Pfam02065	α-galactosidase (EC 3.2.1.22); α-N-acetylgalactosaminidase (EC 3.2.1.49); isomalto-dextranase (EC 3.2.1.94); β-L-arabinopyranosidase (EC 3.2.1.88).
GH28	Pfam00295	polygalacturonase (EC 3.2.1.15); exo-polygalacturonase (EC 3.2.1.67); exo-polygalacturonosidase (EC 3.2.1.82); rhamnogalacturonase (EC 3.2.1.-); endo-xylogalacturonan hydrolase (EC 3.2.1.-); rhamnogalacturonan α-L-rhamnopyranohydrolase (EC 3.2.1.40).
GH29	Pfam01120	α-L-fucosidase (EC 3.2.1.51).
GH30	COG5520	glucosylceramidase (EC 3.2.1.45); β-1,6-glucanase (EC 3.2.1.75); β-xylosidase (EC 3.2.1.37); β-fucosidase (EC 3.2.1.38); β-glucosidase (3.2.1.21); endo-β-1,6-galactanase (EC:3.2.1.164).
GH31	Pfam01055	α-glucosidase (EC 3.2.1.20); α-1,3-glucosidase (EC 3.2.1.84); sucrase-isomaltase (EC 3.2.1.48) (EC 3.2.1.10); α-xylosidase (EC 3.2.1.-); α-glucan lyase (EC 4.2.2.13); isomaltosyltransferase (EC 2.4.1.-).
GH32	Pfam00251	invertase (EC 3.2.1.26); endo-inulinase (EC 3.2.1.7); β-2,6-fructan 6-levanbiohydrolase (EC 3.2.1.64); endo-levanase (EC 3.2.1.65); exo-inulinase (EC 3.2.1.80); fructan β-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153); fructan β-(2,6)-fructosidase/6-exohydrolase (EC 3.2.1.154); sucrose:sucrose 1-fructosyltransferase (EC 2.4.1.99); fructan:fructan 1-fructosyltransferase (EC 2.4.1.100); sucrose:fructan 6-fructosyltransferase (EC 2.4.1.10); fructan:fructan 6G-fructosyltransferase (EC 2.4.1.243).
GH35	Pfam01301	β-galactosidase (EC 3.2.1.23); exo-β-glucosaminidase (EC 3.2.1.165).
GH36	COG3345	α-galactosidase (EC 3.2.1.22); α-N-acetylgalactosaminidase (EC 3.2.1.49); stachyose synthase (EC 2.4.1.67); raffinose synthase (EC 2.4.1.82).
GH37	Pfam01204	α-trehalase (EC 3.2.1.28).
GH38	Pfam01074	α-mannosidase (EC 3.2.1.24) ; mannosyl-oligosaccharide α-1,3-1,6-mannosidase (EC 3.2.1.114); mannosyl-oligosaccharide α-1,3-mannosidase (EC 3.2.1.-).
GH39	self-built	α-L-iduronidase (EC 3.2.1.76); β-xylosidase (EC 3.2.1.37).
GH42	Pfam02449	β-galactosidase (EC 3.2.1.23).
GH43	Pfam04616	β-xylosidase (EC 3.2.1.37); β-1,3-xylosidase (EC 3.2.1.-); α-L-arabinofuranosidase (EC 3.2.1.55); arabinanase (EC 3.2.1.99); xylanase (EC 3.2.1.8); galactan 1,3-β-galactosidase (EC 3.2.1.145).
GH46	cd00978	chitosanase (EC 3.2.1.132).
GH51	Pfam06964	α-L-arabinofuranosidase (EC 3.2.1.55); endoglucanase (EC 3.2.1.4).
GH53	Pfam07745	endo-β-1,4-galactanase (EC 3.2.1.89).
GH55	self-built	exo-β-1,3-glucanase (EC 3.2.1.58); endo-β-1,3-glucanase (EC 3.2.1.39).
GH64	self-built	β-1,3-glucanase (EC 3.2.1.39).
GH65	Pfam03632	α-trehalase (EC 3.2.1.28); maltose phosphorylase (EC 2.4.1.8); trehalose phosphorylase (EC 2.4.1.64); kojibiose phosphorylase (EC 2.4.1.230); trehalose-6-phosphate phosphorylase (EC 2.4.1.-).
GH67	Pfam07488	α-glucuronidase (EC 3.2.1.139); xylan α-1,2-glucuronidase (EC 3.2.1.131).
GH73	Pfam01832	peptidoglycan hydrolase with endo-β-N-acetylglucosaminidase specificity (EC 3.2.1.-).
GH74	smart00602	endoglucanase (EC 3.2.1.4); oligoxyloglucan reducing end-specific cellobiohydrolase (EC 3.2.1.150); xyloglucanase (EC 3.2.1.151).
GH76	Pfam03663	α-1,6-mannanase (EC 3.2.1.101).
GH77	Pfam02446	amylomaltase or 4-α-glucanotransferase (EC 2.4.1.25).
GH78	Pfam05592	α-L-rhamnosidase (EC 3.2.1.40).
GH81	Pfam03639	endo-β-1,3-glucanase (EC 3.2.1.39).
GH85	Pfam03644	endo-β-N-acetylglucosaminidase (EC 3.2.1.96).
GH87	self-built	mycodextranase (EC 3.2.1.61); α-1,3-glucanase (EC 3.2.1.59).

GH88	Pfam07470	d-4,5 unsaturated β-glucuronyl hydrolase (EC 3.2.1.-).
GH94	COG3459	cellobiose phosphorylase (EC 2.4.1.20); cellodextrin phosphorylase (EC 2.4.1.49); chitobiose phosphorylase (EC 2.4.1.-); cyclic β-1,2-glucan synthase (EC 2.4.1.-).
GH95	self-built	α-1,2-L-fucosidase (EC 3.2.1.63); α-L-fucosidase (EC 3.2.1.51).
GH99	self-built	glycoprotein endo-α-1,2-mannosidase (EC 3.2.1.130).
GH105	Pfam07470	unsaturated rhamnogalacturonyl hydrolase (EC 3.2.1.-).
GH106	self-built	α-L-rhamnosidase (EC 3.2.1.40).
GH109	Pfam01408	α-N-acetylgalactosaminidase (EC 3.2.1.49).
GH112	Pfam09508	lacto-N-biose phosphorylase or galacto-N-biose phosphorylase (EC 2.4.1.211); D-galactosyl-1,4-L-rhamnose phosphorylase (EC 2.4.1.-).
GH114	Pfam03537	endo-α-1,4-polygalactosaminidase (EC 3.2.1.109).
GH115	Pfam15979	xylan α-1,2-glucuronidase (3.2.1.131); α-(4-O-methyl)-glucuronidase (3.2.1.-).
GH117	Pfam04616	α-1,3-L-neoagarooligosaccharide hydrolase (EC 3.2.1.-); α-1,3-L-neoagarobiase / neoagarobiose hydrolase (EC 3.2.1.-).
GH120	Pfam07602	β-xylosidase (EC 3.2.1.37).
GH125	Pfam06824	exo-α-1,6-mannosidase (EC 3.2.1.-).
GH127	N/A	β-L-arabinofuranosidase (EC 3.2.1.185).
GH128	N/A	β-1,3-glucanase (EC 3.2.1.39).
GH129	N/A	α-N-acetylgalactosaminidase (EC 3.2.1.49).
GH130	N/A	1-β-D-mannopyranosyl-4-D-glucopyranose:phosphate α-D-mannosyltransferase (EC 2.4.1.281); β-1,4-mannooligosaccharide phosphorylase (EC 2.4.1.319); β-1,2-mannobiose phosphorylase (EC 2.4.1.-); β-1,2-oligomannan phosphorylase (EC 2.4.1.-).
CE1	Pfam00756	acetyl xylan esterase (EC 3.1.1.72); cinnamoyl esterase (EC 3.1.1.-); feruloyl esterase (EC 3.1.1.73); carboxylesterase (EC 3.1.1.1); S-formylglutathione hydrolase (EC 3.1.2.12).
CE2	cd01831	acetyl xylan esterase (EC 3.1.1.72).
CE3	cd01833	acetyl xylan esterase (EC 3.1.1.72).
CE4	Pfam01522	acetyl xylan esterase (EC 3.1.1.72); chitin deacetylase (EC 3.5.1.41); chitooligosaccharide deacetylase (EC 3.5.1.-); peptidoglycan GlcNAc deacetylase (EC 3.5.1.-); peptidoglycan N-acetylmuramic acid deacetylase (EC 3.5.1.-).
CE5	Pfam01083	acetyl xylan esterase (EC 3.1.1.72); cutinase (EC 3.1.1.74).
CE6	Pfam03629	acetyl xylan esterase (EC 3.1.1.72).
CE7	Pfam05448	acetyl xylan esterase (EC 3.1.1.72); cephalosporin-C deacetylase (EC 3.1.1.41).
CE8	Pfam01095	pectin methylesterase (EC 3.1.1.11).
CE9	cd00854	N-acetylglucosamine 6-phosphate deacetylase (EC 3.5.1.25); N-acetylgalactosamine-6-phosphate deacetylase (EC 3.5.1.80).
CE10	COG0657	arylesterase (EC 3.1.1.-); carboxyl esterase (EC 3.1.1.3); acetylcholinesterase (EC 3.1.1.7); cholinesterase (EC 3.1.1.8); sterol esterase (EC 3.1.1.13); brefeldin A esterase (EC 3.1.1.-).
CE12	COG2755	pectin acetylesterase (EC 3.1.1.-); rhamnogalacturonan acetylesterase (EC 3.1.1.-); acetyl xylan esterase (EC 3.1.1.72).
CE13	Pfam03283	pectin acetylesterase (EC 3.1.1.-).
CE14	Pfam02585	N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (EC 3.5.1.89); diacetylchitobiose deacetylase (EC 3.5.1.-); mycothiol S-conjugate amidase (EC 3.5.1.-).
CE15	self-built	4-O-methyl-glucuronoyl methylesterase (EC 3.1.1.-).
PL1	smart00656	pectate lyase (EC 4.2.2.2); exo-pectate lyase (EC 4.2.2.9); pectin lyase (EC 4.2.2.10).
PL3	Pfam12708	pectate lyase (EC 4.2.2.-)
PL6	self-built	alginate lyase (EC 4.2.2.3); chondroitinase B (EC 4.2.2.4).
PL9	self-built	pectate lyase (EC 4.2.2.2); exopolygalacturonate lyase (EC 4.2.2.9); thiopeptidoglycan lyase (EC 4.2.2.-).
PL10	Pfam09492	pectate lyase (EC 4.2.2.2).

PL11	COG14111	rhamnogalacturonan lyase (EC 4.2.2.-); exo-unsaturated rhamnogalacturonan lyase (EC 4.2.2.-).
PL12	Pfam07940	heparin-sulfate lyase (EC 4.2.2.8).
PL21	Pfam07940	heparin lyase (EC 4.2.2.7); heparin-sulfate lyase (EC 4.2.2.8); acharan-sulfate lyase (EC 4.2.2.-).
GT1	COG1819	UDP-glucuronosyltransferase (EC 2.4.1.17); 2-hydroxyacylsphingosine 1-β-galactosyltransferase (EC 2.4.1.45); N-acylsphingosine galactosyltransferase (EC 2.4.1.47); flavonol 3-O-glucosyltransferase (EC 2.4.1.91); indole-3-acetate β-glucosyltransferase (EC 2.4.1.121); sterol glucosyltransferase (EC 2.4.1.173); ecdysteroid UDP-glucosyltransferase (EC 2.4.1.-); zeaxanthin glucosyltransferase (EC 2.4.1.-); zeatin O-β-glucosyltransferase (EC 2.4.1.203); zeatin O-β-xylosyltransferase (EC 2.4.2.40); limonoid glucosyltransferase (EC 2.4.1.210); sinapate 1-glucosyltransferase (EC 2.4.1.120); anthocyanin 3-O-galactosyltransferase (EC 2.4.1.-); anthocyanin 5-O-glucosyltransferase (EC 2.4.1.-); anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115); dTDP-β-2-deoxy-L-fucose: α-L-2-deoxyfucosyltransferase (EC 2.4.1.-); UDP-β-L-rhamnose: α-L-rhamnosyltransferase (EC 2.4.1.-); UDP-glucose: 4-hydroxybenzoate 4-O-β-glucosyltransferase (EC 2.4.1.194); flavonol L-rhamnosyltransferase (EC 2.4.1.159); salicylic acid β-glucosyltransferase (EC 2.4.1.-).
GT2	Pfam00535	cellulose synthase (EC 2.4.1.12); chitin synthase (EC 2.4.1.16); dolichyl-phosphate β-D-mannosyltransferase (EC 2.4.1.83); dolichyl-phosphate β-glucosyltransferase (EC 2.4.1.117); N-acetylglucosaminyltransferase (EC 2.4.1.-); N-acetylgalactosaminyltransferase (EC 2.4.1.-); hyaluronan synthase (EC 2.4.1.212); chitin oligosaccharide synthase (EC 2.4.1.-); β-1,3-glucan synthase (EC 2.4.1.34); β-1,4-mannan synthase (EC 2.4.1.-); β-mannosylphosphodecaprenol mannooligosaccharide α-1,6-mannosyltransferase (EC 2.4.1.199); α-1,3-L-rhamnosyltransferase (EC 2.4.1.-).
GT4	Pfam00534	sucrose synthase (EC 2.4.1.13); sucrose-phosphate synthase (EC 2.4.1.14); α-glucosyltransferase (EC 2.4.1.52); lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56); GDP-Man α-mannosyltransferase (EC 2.4.1.-); 1,2-diacylglycerol 3-glucosyltransferase (EC 2.4.1.157); diglucosyl diacylglycerol synthase (EC 2.4.1.208); digalactosyldiacylglycerol synthase (EC 2.4.1.141); trehalose phosphorylase (EC 2.4.1.231); phosphatidylinositol α-mannosyltransferase (EC 2.4.1.57); UDP-Gal α-galactosyltransferase (EC 2.4.1.-); UDP-Xyl α-xylosyltransferase (EC 2.4.2.-).
GT5	cd03791	UDP-Glc: glycogen glucosyltransferase (EC 2.4.1.11); ADP-Glc: starch glucosyltransferase (EC 2.4.1.21); NDP-Glc: starch glucosyltransferase (EC 2.4.1.242); UDP-Glc: α-1,3-glucan synthase (EC 2.4.1.183) UDP-Glc: α-1,4-glucan synthase (EC 2.4.1.-).
GT7	Pfam02709	lactose synthase (EC 2.4.1.22); β-N-acetylglucosaminyl-glycopeptide β-1,4-galactosyltransferase (EC 2.4.1.38); N-acetyllactosamine synthase (EC 2.4.1.90); β-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.-); xylosylprotein β-4-galactosyltransferase (EC 2.4.1.133).
GT8	Pfam01501	lipopolysaccharide α-1,3-galactosyltransferase (EC 2.4.1.44); UDP-Glc: (glucosyl)lipopolysaccharide α-1,2-glucosyltransferase (EC 2.4.1.-); lipopolysaccharide glucosyltransferase 1 (EC 2.4.1.58); glycogenin glucosyltransferase (EC 2.4.1.186); inositol 1-α-galactosyltransferase (galactinol synthase) (EC 2.4.1.123); homogalacturonan α-1,4-galacturonosyltransferase (EC 2.4.1.43); UDP-GlcA: xylan α-glucuronyltransferase (EC 2.4.1.-).
GT9	Pfam01075	lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56); heptosyltransferase (EC 2.4.-.-).
GT10	Pfam00852	galactoside α-1,3/1,4-L-fucosyltransferase (EC 2.4.1.65); galactoside α-1,3-L-fucosyltransferase (EC 2.4.1.152); glycoprotein α-1,3-L-fucosyltransferase (EC 2.4.1.214).
GT11	Pfam01531	GDP-L-Fuc: galactoside α-1,2-L-fucosyltransferase (EC 2.4.1.69); GDP-L-Fuc: β-LacNac α-1,3-L-fucosyltransferase (EC 2.4.1.-).

GT12	Pfam00535	[N-acetylneuraminyl]-galactosylglucosylceramide N-acetylgalactosaminyltransferase (EC 2.4.1.92).
GT13	Pfam03071	α-1,3-mannosyl-glycoprotein β-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.101).
GT14	Pfam02485	β-1,3-galactosyl-O-glycosyl-glycoprotein β-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.102); N-acetyllactosaminide β-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.150); protein O-β-xylosyltransferase (EC 2.4.2.26).
GT17	Pfam04724	β-1,4-mannosyl-glycoprotein β-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144).
GT19	Pfam02684	lipid-A-disaccharide synthase (EC 2.4.1.182).
GT22	Pfam03901	Dol-P-Man: Man6GlcNAc2-PP-Dol α-1,2-mannosyltransferase (EC 2.4.1.259); Dol-P-Man: Man8GlcNAc2-PP-Dol α-1,2-mannosyltransferase (EC 2.4.1.261); Dol-P-Man: Man2-GlcNAc-phosphatidylinositol α-1,2-mannosyltransferase (EC 2.4.1.-); Dol-P-Man: Man3-GlcNAc-phosphatidylinositol α-1,2-mannosyltransferase (EC 2.4.1.-).
GT23	self-built	N-acetyl-β-D-glucosaminide α-1,6-L-fucosyltransferase (EC 2.4.1.68).
GT26	Pfam03808	UDP-ManNAcA: β-N-acetyl mannosaminuronyltransferase (EC 2.4.1.-); UDP-ManNAc: β-N-acetyl-mannosaminyltransferase (EC 2.4.1.-); UDP-Glc: β-1,4-glucosyltransferase (EC 2.4.1.-).
GT27	cd02510	polypeptide α-N-acetylgalactosaminyltransferase (EC 2.4.1.41).
GT28	Pfam04101	1,2-diacylglycerol 3-β-galactosyltransferase (EC 2.4.1.46); 1,2-diacylglycerol 3-β-glucosyltransferase (EC 2.4.1.157); UDP-GlcNAc: Und-PP-MurAc-pentapeptide β-N-acetylglucosaminyltransferase (EC 2.4.1.227).
GT30	Pfam04413	CMP-β-KDO: α-3-deoxy-D-manno-octulosonic-acid (KDO) transferase (EC 2.4.99.-).
GT32	Pfam04488	α-1,6-mannosyltransferase (EC 2.4.1.-); α-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.-); α-1,4-N-acetylgalactosaminyltransferase (EC 2.4.1.-).
GT35	Pfam00343	glycogen or starch phosphorylase (EC 2.4.1.1).
GT39	Pfam02366	Dol-P-Man: protein α-mannosyltransferase (EC 2.4.1.109).
GT45	Pfam00535	α-N-acteylglucosaminyltransferase (EC 2.4.1.-).
GT46	self-built	Currently uncharacterised.
GT51	Pfam00912	murein polymerase (EC 2.4.1.129).
GT56	Pfam07429	TDP-Fuc4NAc: lipid II Fuc4NAc transferase (EC 2.4.1.-)
GT60	Pfam11397	UDP-GlcNAc: polypeptide α-N-acetylglucosaminyltransferase (EC 2.4.1.-); UDP-GlcNAc: hydroxyproline polypeptide α-N-acetylglucosaminyltransferase (EC 2.4.1.-).
GT73	PRK09822	CMP-β-KDO: α-3-deoxy-D-manno-octulosonic-acid (KDO) transferase (EC 2.4.99.-).
GT76	Pfam04188	Dol-P-Man: α-1,6-mannosyltransferase (EC 2.4.1.-).
GT81	PRK13915	NDP-Glc: glucosyl-3-phosphoglycerate synthase (EC 2.4.1.-); NDP-Man: mannosyl-3-phosphoglycerate synthase (EC 2.4.1.-).
GT83	COG1807	undecaprenyl phosphate-α-L-Ara4N: 4-amino-4-deoxy-β-L-arabinosyltransferase (EC 2.4.2.-); dodecaprenyl phosphate-β-galacturonic acid: lipopolysaccharide core α-galacturonosyl transferase (EC 2.4.1.-).
GT87	Pfam09594	polyprenol-P-Man: α-1,2-mannosyltransferase (EC 2.4.1.-).
GT90	smart00672	UDP-Xyl: (mannosyl) glucuronoxylomannan/galactoxylomannan β-1,2-xylosyltransferase (EC 2.4.2.-).
GT92	Pfam01697	UDP-Gal: N-glycan core α-1,6-fucoside β-1,4-galactosyltransferase (EC 2.4.1.-); UDP-Gal: β-galactoside β-1,4-galactosyltransferase (EC 2.4.1.-).
GT94	Pfam06165	GDP-Man: GlcA-β-1,2-Man-α-1,3-Glc-β-1,4-Glc-α-1-PP-undecaprenol β-1,4-mannosyltransferase (2.4.1.251).
CBM2	Pfam00553	Modules of approx. 100 residues and which are found in a large number of bacterial enzymes. The cellulose-binding function has been demonstrated in many cases. Several of these modules have been shown to also bind chitin or xylan.

CBM3	Pfam00942	Modules of approx. 150 residues found in bacterial enzymes. The cellulose-binding function has been demonstrated in many cases. In one instance binding to chitin has been reported.
CBM4	Pfam02018	Modules of approx. 150 residues found in bacterial enzymes. Binding of these modules has been demonstrated with xylan, β-1,3-glucan, β-1,3-1,4-glucan, β-1,6-glucan and amorphous cellulose but not with crystalline cellulose.
CBM5	Pfam02839	Modules of approx. 60 residues found in bacterial enzymes. Chitin-binding described in several cases. Distantly related to the CBM12 family.
CBM6	Pfam03422	Modules of approx. 120 residues. The cellulose-binding function has been demonstrated in cellulose and β-1,4-xylan. Some of these modules also bind β-1,3-glucan, β-1,3-1,4-glucan, and β-1,4-glucan.
CBM9	Pfam06452	Modules of approx. 170 residues found so far only in xylanases. The cellulose-binding function has been demonstrated in one case.
CBM12	Pfam02839	Modules of approx. 40-60 residues. The majority of these modules is found among chitinases where the function is chitin-binding. Distantly related to the CBM5 family.
CBM13	Pfam00652	Modules of approx. 150 residues which always appear as a threefold internal repeat. These modules have been found in a number of other proteins of various functions including glycoside hydrolases and glycosyltransferases.
CBM16	Pfam02018	Carbohydrate-binding module 16. Binding to cellulose and glucomannan.
CBM20	Pfam00686	The granular starch-binding function has been demonstrated in several cases. Interact strongly with cyclodextrins. Often designated as starch-binding domains (SBD).
CBM21	Pfam03370	Modules of approx. 100 residues. The granular starch-binding function has been demonstrated in one case. Sometimes designated as starch-binding domains (SBD).
CBM22	Pfam02018	A xylan binding function has been demonstrated in several cases and affinity with mixed β-1,3/β-1,4-glucans in one. In several cases a thermostabilizing effect has also been seen.
CBM23	Pfam03425	Mannan-binding function demonstrated in one case.
CBM26	self-built	Starch-binding function demonstrated in two cases.
CBM27	Pfam09212	Mannan-binding function demonstrated in two cases
CBM28	Pfam03424	The module from the endo-1,4-glucanase binds to non-crystalline cellulose, cellooligosaccharides, and β-(1,3)(1,4)-glucans.
CBM30	Pfam02927	Binding to cellulose has been demonstrated for the N-terminal module of Fibrobacter succinogenes CelF.
CBM32	Pfam00754	Binding to galactose, lactose, polygalacturonic acid, LacNAc (β-D-galactosyl-1,4-β-D-N-acetylglucosamine).
CBM34	cd02857	Modules of approx. 120 residues. Granular starch-binding function in α-amylase.
CBM35	Pfam03422	Modules of approx. 130 residues. A module that is conserved in xylan-degrading enzymes binds to xylan and the interaction is calcium dependent, while a module from a mannanase binds to decorated soluble mannans and mannooligosaccharides.
CBM36	Pfam03422	Modules of approx. 120-130 residues displaying structural similarities to CBM6 modules. The only CBM36 currently characterised is from xylanase 43A, shows calcium-dependent binding of xylans and xylooligosaccharides. X-ray crystallography shows that there is a direct interaction between calcium and ligand.
CBM37	smart00060	Modules of approx. 100 residues, conserved in numerous polysaccharide-degrading enzymes and other proteins. Several members of CBM37 have been shown to exhibit rather broad binding specificity to xylan, chitin, microcrystalline and phosphoric-acid swollen cellulose, as well as more heterogeneous substrates, such as alfalfa cell walls, banana stem and wheat straw.
CBM39	self-built	Modules generally found at the N-terminus of a GH16 module (itself frequently lacking a catalytic machinery) and more seldomly in isolation. The β-1,3-glucan binding function has been demonstrated, along with binding to lipopolysaccharide and lipoteichoic acid.

CBM40	Pfam02973	Modules of approx. 200 residues, found at the N-terminus of GH33 sialidases. Can also be found inserted in the β-propeller of GH33 sialidases. The sialic acid binding function has been demonstrated for the N-terminal CBM40.
CBM41	Pfam03714	Modules of approx. 100 residues found in primarily in bacterial pullulanases. The N-terminal module from Pul13 has been shown to bind to the α-glucans amylose, amylopectin, pullulan, and oligosaccharide fragments derived from these polysaccharides.
CBM42	Pfam05270	Modules of approx. 160 residues found mostly at the C-terminus of GH54 catalytic domains. Binding to arabinofuranose (present in arabinoxylan) has been demonstrated.
CBM46	Pfam03442	Modules of approx. 100 residues, found at the C-terminus of several GH5 cellulases. Cellulose-binding function demonstrated in one case.
CBM47	Pfam00754	Modules of approx 150 residues. Fucose-binding activity demonstrated.
CBM48	Pfam02922	Modules of approx. 100 residues with glycogen-binding function, appended to GH13 modules. Also found in the β subunit (glycogen-binding) of AMP-activated protein kinases (AMPK).
CBM50	Pfam01476	Modules of approx. 50 residues found attached to various enzymes from families GH18, GH19, GH23, GH24, GH25 and GH73, i.e. enzymes cleaving either chitin or peptidoglycan. Binding to chitopentaose demonstrated in chitinase A. CBM50 modules are also found in a multitude of other enzymes targeting the petidoglycan such as peptidases and amidases.
CBM51	Pfam08305	Modules of approx. 150 residues found attached to various enzymes from families GH2, GH27, GH31, GH95, GH98 and GH101. Binding to galactose and to blood group A/B-antigens demonstrated in GH95/CBM51 and GH98/CBM51 respectively.
CBM56	self-built	β-1,3-glucan binding function.
CBM57	Pfam11721	Related domains attached to various glycosidases.
CBM59	self-built	Binding to mannan, xylan, and cellulose demonstrated for the CBM59 of ManF-X10 xylanase.
CBM61	self-built	Modules of approx. 150 residues found appended to GH16, GH30, GH31, GH43, GH53 and GH66 catalytic domains. A β-1,4-galactan binding function has been demonstrated for a GH53 galactanase.
CBM65	N/A	CBM65A and CBM65B, derived from endoglucanase EcCel5A, bind to a range of β-glucans but, uniquely, display significant preference for xyloglucan.
CBM66	N/A	The CBM66 module, derived from exo-acting β-fructosidase SacC, targets the terminal fructoside residue of fructans.
CBM67	N/A	L-rhamnose binding activity and 3-D structure of the CBM67 of α-L-rhamnosidase (SaRha78A).
AA3	N/A	cellobiose dehydrogenase (EC 1.1.99.18); glucose 1-oxidase (EC 1.1.3.4); aryl alcohol oxidase (EC 1.1.3.7); alcohol oxidase (EC 1.1.3.13); pyranose oxidase (EC 1.1.3.10).
AA6	N/A	1,4-benzoquinone reductase (EC. 1.6.5.6).
cohesin	Pfam00963	The cohesin-dockerin interaction is the crucial interaction for complex formation in the cellulosome.
SLH	Pfam00395	S-layer homology domain, anchoring cellulosome onto the bacterial cell surfaces.

* Domain: CDD signature domain model if any (Marchler-Bauer *et al.*, 2012), otherwise the self-built model used as described by Yin *et al.* (2012).

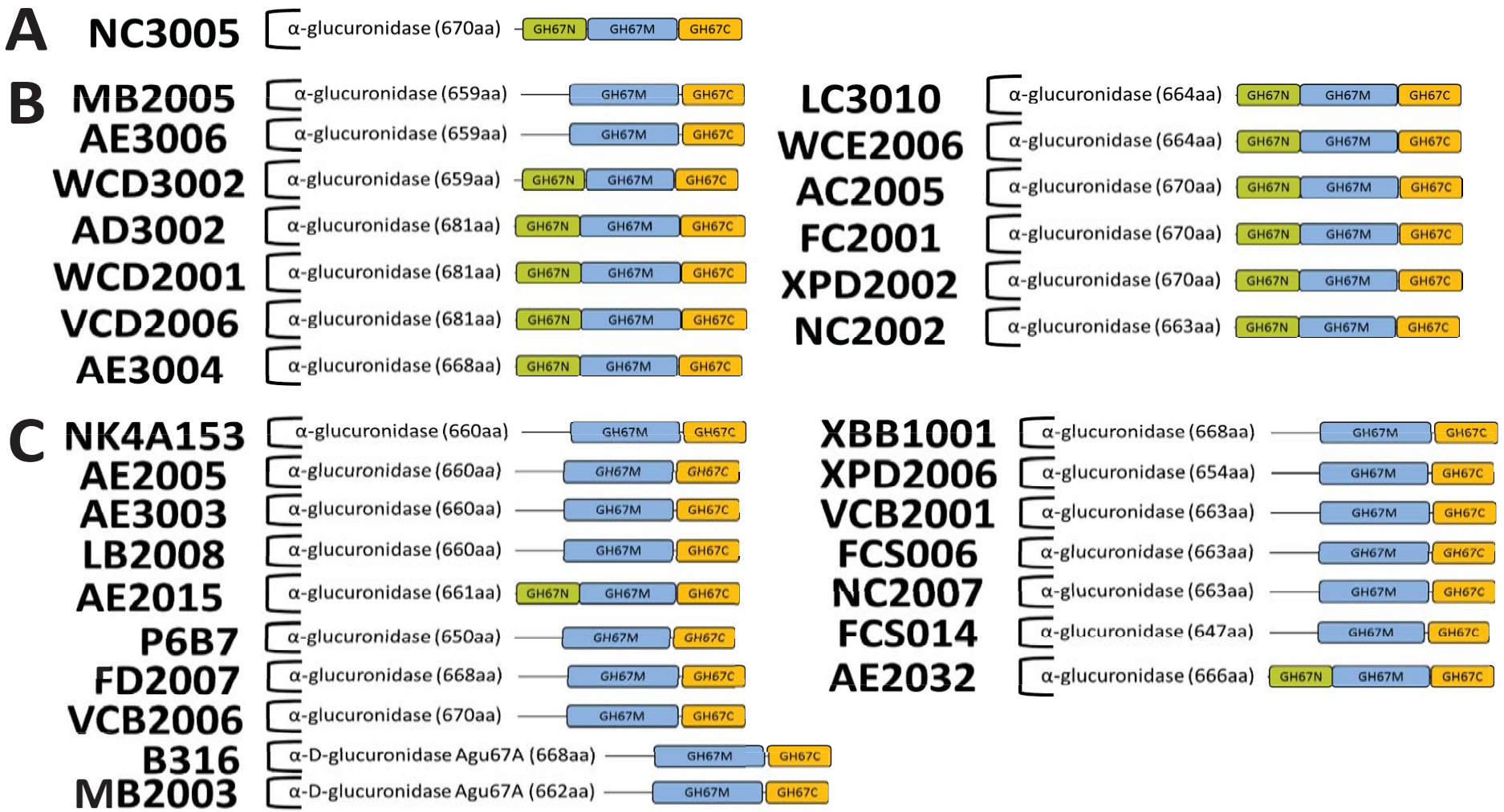


Figure A2.9. Predicted Pfam domains of GH67 α-glucuronidases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2 and C, *Butyrivibrio* sp. 3. Abbreviations: GH67N, glycoside hydrolase family 67 N-terminus (Pfam03648); GH67M, glycoside hydrolase family 67 middle domain (Pfam07488); GH67C, glycoside hydrolase family 67 C-terminus (Pfam07477).

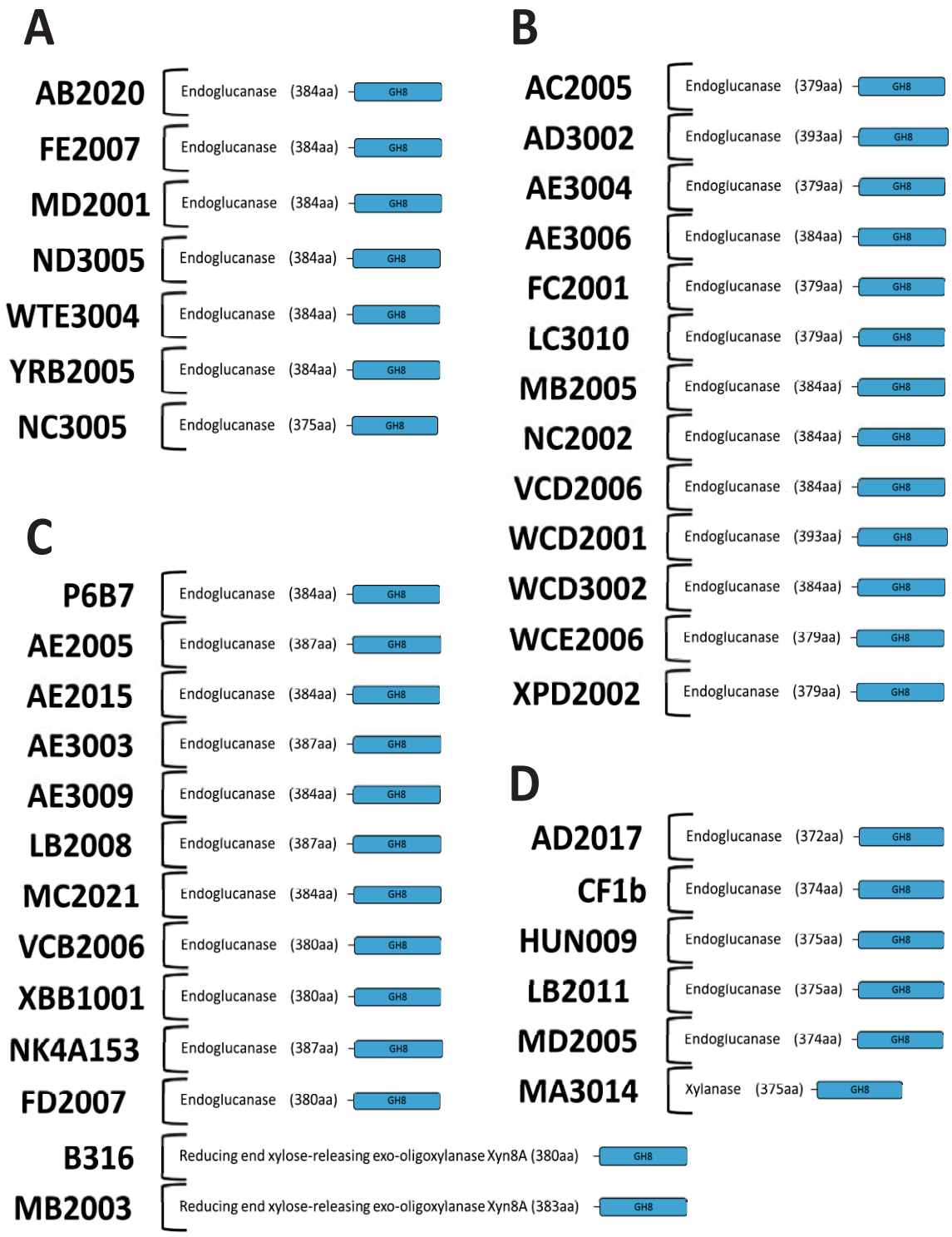


Figure A2.10. Predicted Pfam domain of GH8 oligosaccharide reducing-end xylanases. A, *B. fibrisolvens*, B, *Butyrivibrio* sp. 2, C, *Butyrivibrio* sp. 3 and D, *Pseudobutyrivibrio*. Abbreviations: GH8, glycoside hydrolases family 8 (Pfam01270).

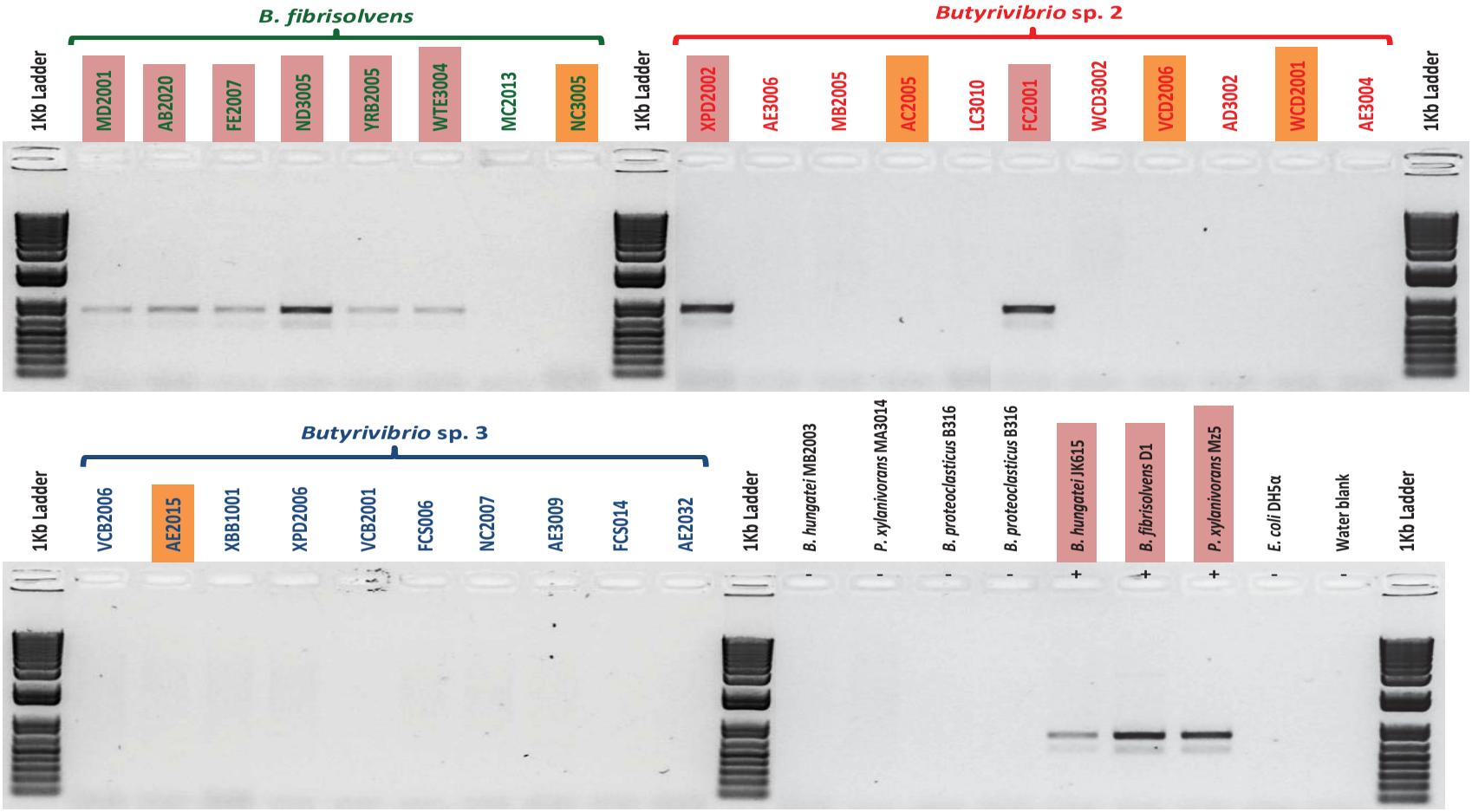


Figure A2.11. PCR amplification of enolase genes from *Butyrivibrio* and *Pseudobutyrivibrio* strains. Colour codes: red label, indicates strains that possess enolase gene and screened positive, orange label, represents strains that possess enolase gene but screened negative. The positive controls used were *B. hungatei* JK615^{T} (DSM 14810) (Kopecny *et al.*, 2003), *B. fibrisolvens* D1^{T} (DSM 3071) (Noel, 2013) and *P. xylanivorans* Mz5^{T} (DSM 14809) (Kopecny *et al.*, 2003), and negative controls were *B. hungatei* MB2003, *P. xylanivorans* MA3014, and *B. proteoclasticus* B316. The *E. coli* DH5α strain (Scheutz and Strockbine, 2005) was used as a negative control to demonstrate specificity of the primers for *Butyrivibrio*.

Appendix III

Table A3.1. Classification and general features of *B. hungatei* MB2003.

MIGS ID	Property	Term	Evidence code *
	Current classification	Domain: Bacteria	IDA
		Phylum: Firmicutes	
		Class: Clostridia	
		Order: Clostridiales	
		Family: Lachnospiraceae	
		Genus: *Butyrivibrio*	
		Species: *hungatei*	
		Type strain: No	
		Strain: MB2003	
	Gram stain	Positive	IDA
	Cell shape	Short/medium rod	IDA
	Motility	No	IDA
	Sporulation	No	IDA
	Temperature range	37-39 °C	NAS
	Optimum temperature	38 °C	NAS
	Carbon source	Variety of carbohydrates	IDA
	Energy metabolism	Fermentative metabolism	IDA
MIGS-6	Habitat	Bovine rumen	IDA
MIGS-6.3	Salinity	not reported	
MIGS-22	Oxygen requirement	Strict anaerobe	IDA
MIGS-15	Biotic relationship	Free-living	NAS
MIGS-14	Pathogenicity	Not known as a pathogen	NAS
MIGS-4	Geographic location	Ruakura, Hamilton, New Zealand	IDA
MIGS-5	Sample collection time	May and August 2009 and May 2010	
MIGS-4.1	Latitude	37.77	IDA
MIGS-4.2	Longitude	175.31	
MIGS-4.3	Depth	Not reported	
MIGS-4.4	Altitude	40	

*, Evidence codes: IDA, Inferred from Firect Assay, NAS, Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). Evidence codes are from the Gene Ontology project (Ashburner *et al.*, 2000).

Table A3.2. Project information.

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality, closed genome
MIGS-28	Libraries used	3 Kb mate paired-end library
MIGS-29	Sequencing platforms	454 GS FLX Titanium
MIGS-31.2	Fold coverage	234×
MIGS-30	Assemblers	Newbler version 2.3
MIGS-32	Gene calling method	Glimmer and BLASTX
	Genome Database release	On publication
	Taxon ID	2619618998
	IMG Submission ID	64390
	NCBI Taxon ID	185008
	GOLD ID in IMG Database	Study ID: Gs0114730 Project ID: Gp0115331
	GOLD ID	Ga0074201
	Genbank ID	N/A
	Genbank Date of Release	On publication
	BioProject ID	Gb0114882
	Project relevance	Ruminant plant-fibre degradation

Table A3.3. Genome statistics of *B. hungatei* MB2003.

Attribute	Genome (total)	
	Value	% Total [A]
Genome size (bp)	3,386,314	100
DNA G+C content (bp)	1,344,683	39.71
DNA coding region (bp)	3,071,277	90.69
Number of replicons	4	-
Total genes	3,072	100
Protein-coding genes	3,007	97.89
RNA genes	65	2.11
rRNA operons	4	0.39
tRNA genes	48	1.55
Pseudogenes	66	2.14
Genes assigned to COGs	1,842	59.96
Genes with signal peptides	148	4.79
Genes with transmembrane helices	881	28.54

[A], The total is based on either the size of the genome in base pairs or the total number of genes or protein-coding genes in the annotated genome.

Table A3.4. Assignment of MB2003 protein coding genes to COG functional categories.

COG	COG descriptor	Number of CDS	% Total
A	RNA processing and modification	0	0
B	Chromatin structure and dynamics	0	0
C	Energy production and conversion	69	3.39
D	Cell cycle control, cell division, chromosome partitioning	32	1.57
E	Amino acid transport and metabolism	177	8.69
F	Nucleotide transport and metabolism	80	3.93
G	Carbohydrate transport and metabolism	243	11.93
H	Coenzyme transport and metabolism	79	3.88
I	Lipid transport and metabolism	72	3.53
J	Translation, ribosomal structure and biogenesis	194	9.52
K	Transcription	149	7.31
L	Replication, recombination and repair	88	4.32
M	Cell wall/membrane/envelope biogenesis	155	7.61
N	Cell motility	61	2.99
O	Posttranslational modification, protein turnover, chaperones	78	3.83
P	Inorganic ion transport and metabolism	79	3.88
Q	Secondary metabolites biosynthesis, transport and catabolism	16	0.79
R	General function prediction only	158	7.76
S	Function unknown	73	3.58
T	Signal transduction mechanisms	139	6.82
U	Intracellular trafficking, secretion, and vesicular transport	22	1.08
V	Defense mechanisms	65	3.19
W	Extracellular structures	4	0.2
X	Mobilome: prophages, transposons	3	0.15
Z	Cytoskeleton	1	0.05
	No COG category assigned	1245	40.33

Table A3.5. Comparison of MB2003 and B316 protein coding gene percentages to COG functional categories.

COG	COG descriptor	MB2003 (%)	B316 (%)
A	RNA processing and modification	0	0
B	Chromatin structure and dynamics	0	0
C	Energy production and conversion	3.39	3.72
D	Cell cycle control, cell division, chromosome partitioning	1.57	1.44
E	Amino acid transport and metabolism	8.69	7.91
F	Nucleotide transport and metabolism	3.93	3.98
G	Carbohydrate transport and metabolism	11.93	12.15
H	Coenzyme transport and metabolism	3.88	3.23
I	Lipid transport and metabolism	3.53	2.8
J	Translation, ribosomal structure and biogenesis	9.52	8.96
K	Transcription	7.31	7.3
L	Replication, recombination and repair	4.32	4.63
M	Cell wall/membrane/envelope biogenesis	7.61	8.52
N	Cell motility	2.99	2.75
O	Posttranslational modification, protein turnover, chaperones	3.83	3.89
P	Inorganic ion transport and metabolism	3.88	2.75
Q	Secondary metabolites biosynthesis, transport and catabolism	0.79	0.79
R	General function prediction only	7.76	7.43
S	Function unknown	3.58	4.11
T	Signal transduction mechanisms	6.82	7.47
U	Intracellular trafficking, secretion, and vesicular transport	1.08	1.14
V	Defense mechanisms	3.19	3.19
W	Extracellular structures	0.2	0.31
X	Mobilome: prophages, transposons	0.15	1.49
Z	Cytoskeleton	0.05	0.04
	No COG category assigned	40.33	46.51

Table A3.6. List of top 20 candidate gene targets for the universal mono- and co-culture qPCR assays based on DBA.

ORFs conserved in <u>both</u> B316 and MB2003			**BLASTP hits to MB2003 genome**				
Candidate gene annotation	**Locus tag**	**Size (aa)**	**Gene annotation**	**Locus tag**	**Size (aa)**	**Alignment length**	**E-value**
butyrate kinase, *buk* *	**bpr_I2323**	**354**	**butyrate kinase, *Buk***	**bhn_I0586**	**354**	**354**	**0**
anthranilate synthase component I *trpE*	bpr_I2335	503	anthranilate synthase component I, *trpE*	bhn_I1059	496	494	1.00E-175
cobalamin biosynthesis protein, *cbiM*	bpr_I2478	342	cobalamin biosynthesis protein, *cbiM*	bhn_I2248	334	333	1.00E-169
erythromycin esterase	bpr_I2892	431	erythromycin esterase	bhn_I2030	435	436	1.00E-158
5,10-methylenetetrahydrofolate reductase, *metF*	bpr_I2211	287	5,10-methylenetetrahydrofolate reductase, *metF*	bhn_I2112	287	287	1.00E-149
enoyl-CoA hydratase, *paaB*	bpr_I2487	264	enoyl-CoA hydratase, *paaB*	bhn_I2255	264	264	1.00E-147
RNA-binding S4 domain-containing protein	bpr_I2933	271	RNA-binding S4 domain-containing protein	bhn_I2760	269	268	1.00E-137
chromosome partitioning protein, *parB*	bpr_I2952	300	chromosome partitioning protein, *parB*	bhn_I2795	300	296	1.00E-134
hypothetical protein	bpr_I2147	263	hypothetical protein	bhn_I0706	265	259	1.00E-133
glutamate racemase, *racE*	bpr_I2919	272	glutamate racemase, *murI*	bhn_I2751	272	271	1.00E-124
transcription antitermination protein, *nusG*	bpr_I2301	213	transcription antitermination protein, *nusG*	bhn_I0608	207	214	1.00E-96
hypothetical protein	bpr_I2915	378	hypothetical protein	bhn_I2597	356	380	1.00E-91
ribosomal protein S12, *rpsL*	bpr_I2367	139	ribosomal protein S12, *rpsL*	bhn_I0616	139	139	2.00E-77
phosphodiesterase, *yfcE*	bpr_I2469	160	phosphodiesterase, *yfcE*	bhn_I2240	160	160	3.00E-77
molecular chaperone, *grpE*	bpr_I2626	233	molecular chaperone, *grpE*	bhn_I2522	214	226	9.00E-77
hypothetical protein	bpr_I2725	255	hypothetical protein	bhn_I1398	266	250	3.00E-76
cobalamin adenosyltransferase, *cobA*	bpr_I2401	177	cobalamin adenosyltransferase, *cobO*	bhn_I2192	381	172	2.00E-67
hypothetical protein	bpr_I2785	160	hypothetical protein	bhn_I2643	150	146	3.00E-51
cold shock domain protein, *cspD1*	bpr_I2451	66	cold shock domain protein, *cspD*	bhn_I2230	66	66	5.00E-34
preprotein translocase, *secE*	bpr_I2302	87	preprotein translocase, *secE*	bhn_I0607	92	92	2.00E-25

*, represents the gene selected for qPCR assay design.

Table A3.7. List of top 20 candidate gene targets for the *B. proteoclasticus* B316 specific mono- and co-culture qPCR assays based on DBA.

ORFs conserved in B316 but not in MB2003			**BLASTP hit to MB2003 genome**				
Candidate gene annotation	**Locus tag**	**Size (aa)**	**Gene annotation**	**Locus tag**	**Size (aa)**	**Alignment length**	**E-value**
cell wall binding domain-containing protein CBD *	**bpr_I0264**	**1802**	**lysozyme, *lyc25C***	**bhn_I0194**	**561**	**220**	**3.00E-07**
pectin methylesterase, *pme8B*	bpr_I2473	2732	lysozyme, *lyc25C*	bhn_I0194	561	237	1.00E-14
preprotein translocase, *secG*	bpr_I1295	80	preprotein translocase subunit, *secG*	bhn_I1525	80	80	4.00E-33
tryptophanyl-tRNA synthetase *trpS*	bpr_I1256	364	tryptophanyl-tRNA synthetase, *trpS*	bhn_I1319	333	345	3.00E-34
preprotein translocase, *yajC*	bpr_I1335	117	preprotein translocase, *yajC*	bhn_I1490	120	120	4.00E-41
carbohydrate esterase family 12, *est12B*	bpr_I1204	2676	hypothetical protein	bhn_I0030	746	463	3.00E-57
hypothetical protein	bpr_I1233	218	hypothetical protein	bhn_I1173	201	198	3.00E-61
flagellar basal-body rod protein, *flgB*	bpr_I1355	131	flagellar basal-body rod protein, *flgB*	bhn_I1470	130	130	1.00E-62
lysozyme, *lyc25B*	bpr_I1326	486	lysozyme, *lyc25E*	bhn_I0536	1213	194	5.00E-78
SsrA-binding protein, *smpB*	bpr_I1297	159	SsrA-binding protein, *smpB*	bhn_I1523	156	156	1.00E-84
adenine phosphoribosyl transferase, *apt*	bpr_I1345	174	adenine phosphoribosyltransferase, *apt*	bhn_I1480	174	174	5.00E-88
integrase/recombinase, *xerD*	bpr_I1245	292	phage integrase family protein, *xerD*	bhn_I1187	297	298	1.00E-91
guanylate kinase, *gmk2*	bpr_I1282	207	guanylate kinase, *gmk*	bhn_I1539	207	207	2.00E-92
DNA repair protein, *recO*	bpr_I1218	217	DNA repair protein, *recO*	bhn_I1158	215	215	2.00E-96
chromosome segregation protein, *scpB*	bpr_I1225	189	chromosome segregation protein, *scpB*	bhn_I1166	189	189	9.00E-99
ribose 5-phosphate isomerase, *rpiB2*	bpr_I1289	213	sugar-phosphate isomerase, *rpiB*	bhn_I1531	214	212	1.00E-118
uracil-DNA glycosylase, *ung*	bpr_I1242	226	uracil-DNA glycosylase, *ung*	bhn_I1184	225	225	1.00E-120
cobalamin biosynthesis protein, *cbiD*	bpr_I1255	401	cobalamin biosynthesis protein, *cbiD*	bhn_I1197	395	392	1.00E-130
glutamate 5-kinase, *proB*	bpr_I1299	266	glutamate 5-kinase, *proB*	bhn_I1521	267	264	1.00E-139
oligopeptide ABC transporter permease, *oppC1*	bpr_I1278	324	nickel/peptide ABC transporter permease protein	bhn_I2721	325	323	1.00E-163

*, represents the gene selected for qPCR assay design.

Table A3.8. List of top 20 candidate gene targets for the *B. hungatei* MB2003 specific mono- and co-culture qPCR assays based on DBA.

ORFs conserved in MB2003 but not in B316			**BLASTP hit to B316 genome**				
Candidate gene annotation	**Locus tag**	**Size (aa)**	**Gene annotation**	**Locus tag**	**Size (aa)**	**Alignment length**	**E-value**
α-L-arabinofuranosidase, *arf51C* *	**bhn_I1532**	**638**	**α-L-arabinofuranosidase, *arf51A***	**bpr_I0329**	**501**	**393**	**1.00E-11**
hypothetical protein	bhn_I1019	266	conserved hypothetical transmembrane protein	bpr_I0857	270	246	5.00E-19
S4 domain protein, *yaaA*	bhn_I0003	71	S4 domain protein, *yaaA*	bpr_I0003	71	71	1.00E-30
two component system response regulator *pdsS*	bhn_I0970	531	HAMP/SpoIIE domain-containing protein	bpr_I0245	557	162	5.00E-32
β-galactosidase, *bgl3C*	bhn_I0182	671	β-glucosidase, *bgl1A*	bpr_I1685	434	452	3.00E-35
flagellar protein, *fliS2*	bhn_I0056	158	flagellar protein, *fliS2*	bpr_I0063	158	158	4.00E-55
hypothetical protein	bhn_I2784	214	YkuD domain-containing protein	bpr_I1849	211	211	1.00E-61
cell division protein, *ftsL*	bhn_I1740	146	cell division protein, *ftsL*	bpr_I1868	146	142	2.00E-64
hypothetical protein	bhn_I0779	215	conserved hypothetical protein	bpr_I2951	173	173	8.00E-71
FeS assembly protein, *sufE*	bhn_I0049	147	FeS assembly protein, *sufE*	bpr_I0055	149	146	2.00E-73
hypothetical protein	bhn_I0079	146	conserved hypothetical protein	bpr_I0085	146	146	8.00E-74
MraZ protein, *mraZ*	bhn_I1742	146	MraZ protein, *mraZ*	bpr_I1870	146	146	5.00E-75
fumarate hydratase, *fumB*	bhn_I0072	185	fumarate hydratase beta subunit, *fumB*	bpr_I0079	187	176	3.00E-80
S-ribosylhomocysteinase, *luxS*	bhn_I0025	159	S-ribosylhomocysteinase, *luxS*	bpr_I0033	159	159	2.00E-89
hypothetical protein	bhn_I0044	296	conserved hypothetical protein	bpr_I0071	306	298	9.00E-94
cell division protein, *ftsQ*	bhn_I1735	269	cell division protein, *ftsQ*	bpr_I1863	283	244	1.00E-104
glutamate racemase, *murI*	bhn_I2751	272	glutamate racemase, *murI*	bpr_I2919	272	271	1.00E-124
Flp pilus assembly protein, *tadC*	bhn_I0061	485	Flp pilus assembly protein, *tadC*	bpr_I0068	486	485	1.00E-130
fumarate hydratase subunit alpha	bhn_I0071	280	fumarate hydratase alpha subunit, *fumA*	bpr_I0078	280	280	1.00E-141
cell division protein, *ftsW*	bhn_I1736	380	cell division protein, *ftsW*	bpr_I1864	386	380	1.00E-163

*, represents the gene selected for qPCR assay design.

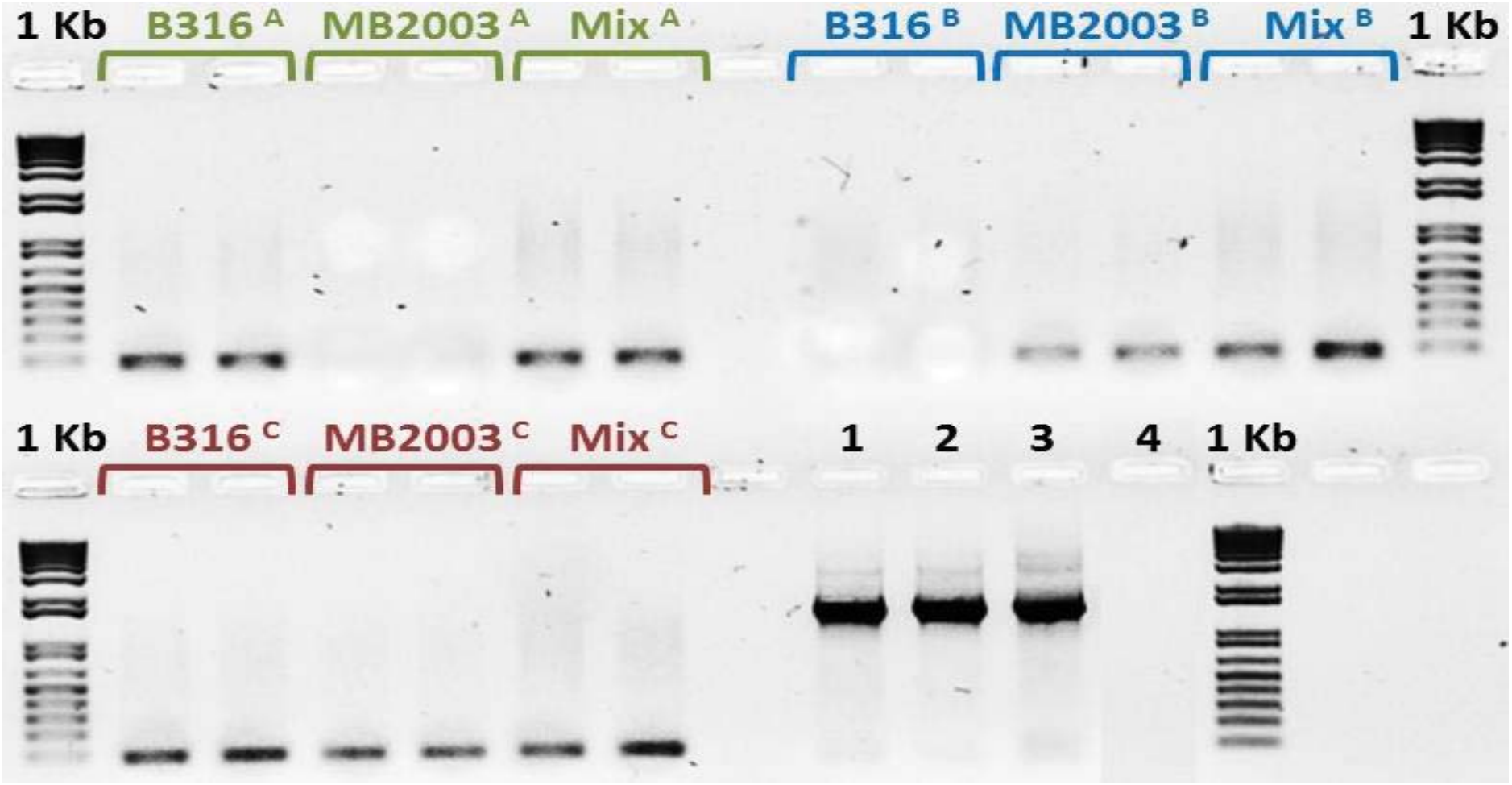


Figure A3.1. PCR amplification of *Butyrivibrio* strain DNAs using *Taq*Man primer/probe assay oligonucleotides. Standard PCRs performed using specific primers for: [A], *CBD*, *B. proteoclasticus* B316 specific gene, [B], *arf51C*, *B. hungatei* MB2003 specific gene, [C], *buk* universal gene. Mix, represents co-culture of *B. proteoclasticus* B316 and *B. hungatei* MB2003. Positive (1, B316, 2, MB2003 and 3, co-culture) and negative (4, no gDNA) controls were amplified using fD1and rD1primers (Table 2.7). Samples were run on a 1% (w/v) agarose gel for 40 min.

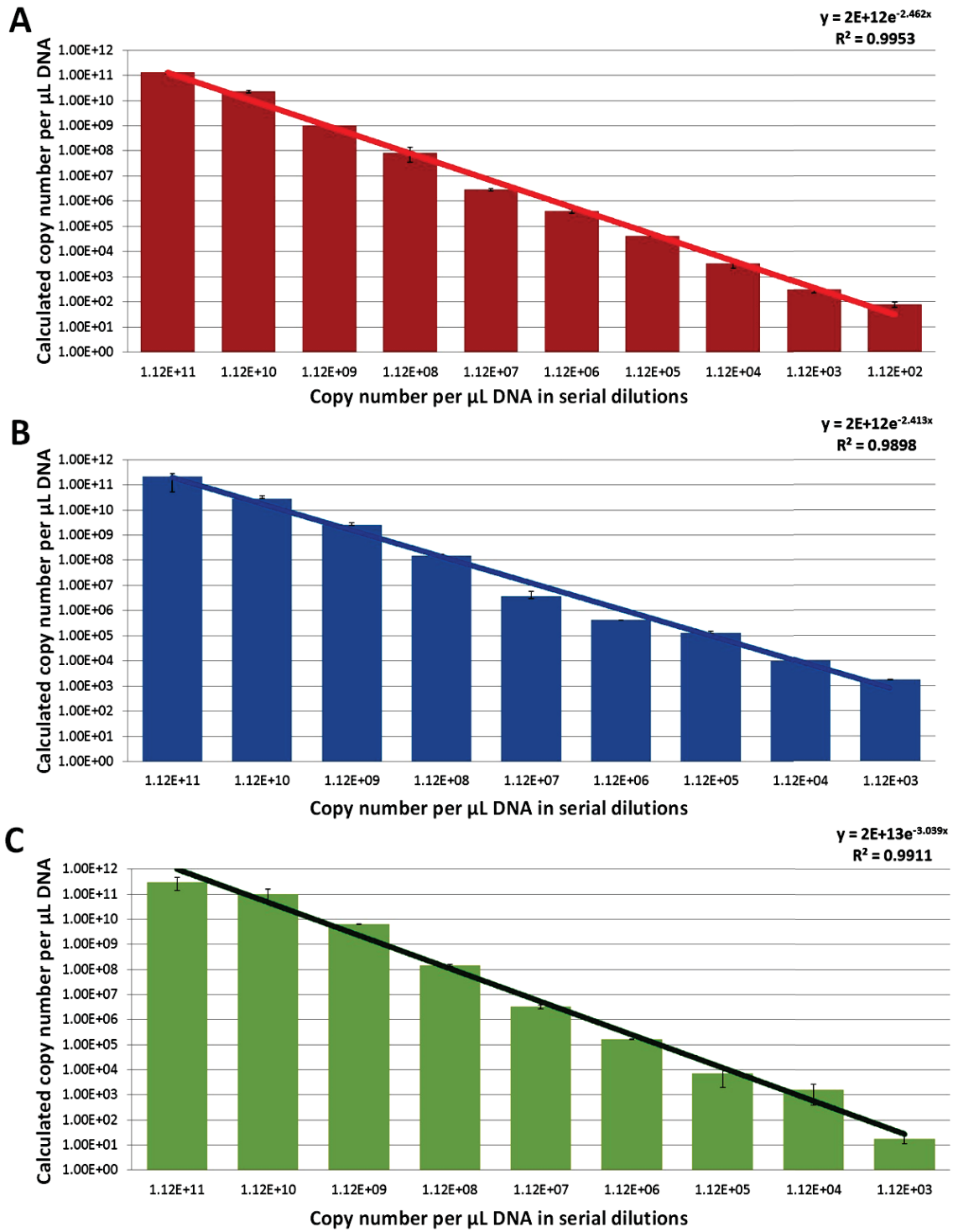


Figure A3.2. Standard curves derived from simplex real time PCR reactions. A, *buk* universal assay, B, *arf51C B. hungatei* MB2003 specific assay, C, *CBD B. proteoclasticus* B316 specific assay. Curves were generated using 10-fold series dilution of gDNAs extracted from plasmids containing the cloned PCR product of the respective primer/probe assay. Error bars represent ±1 standard deviation of the mean.

Table A3.9. Quality of raw reads and FastQC results of trimmed reads.

Sample	Total Raw Sequences	Filtered Sequences	Sequence Length (bp)	GC (%)	Sequence Duplication Level (%)	Lowest Quality score across all bases (Phred64)
MB2003-Monoculture-1-Xylan_F	12,690,950	0	9-90	42	>90.25	31
MB2003-Monoculture-1-Xylan_R	12,690,950	0	7-90	41	>92.42	30
MB2003-Monoculture-2-Xylan_F	12,718,791	0	8-90	42	>90.4	31
MB2003-Monoculture-2-Xylan_R	12,718,791	0	6-90	41	>92.74	30
MB2003-Monoculture-3-Xylan_F	12,704,236	0	8-90	42	>90.31	31
MB2003-Monoculture-3-Xylan_R	12,704,236	0	7-90	42	>92.57	30
B316-Monoculture-1-Xylan_F	12,341,173	0	8-90	43	>87.72	30
B316-Monoculture-1-Xylan_R	12,341,173	0	6-90	43	>88.08	26
B316-Monoculture-2-Xylan_F	12,444,544	0	7-90	41	>85.86	30
B316-Monoculture-2-Xylan_R	12,444,544	0	6-90	40	>86.38	27
B316-Monoculture-3-Xylan_F	6,511,651	0	8-90	42	>92.19	31
B316-Monoculture-3-Xylan_R	6,511,651	0	7-90	41	>91.85	30
Co-culture-1-Xylan_F	12,357,812	0	8-90	41	>85.21	30
Co-culture-1-Xylan_R	12,357,812	0	6-90	40	>85.36	26
Co-culture-2-Xylan_F	12,391,049	0	8-90	42	>85.08	30
Co-culture-2-Xylan_R	12,391,049	0	6-90	41	>85.4	26
Co-culture-3-Xylan_F	12,307,514	0	8-90	41	>83.85	30
Co-culture-3-Xylan_R	12,307,514	0	6-90	40	>84.1	27
MB2003-Monoculture-1-Pectin_F	12,172,558	0	7-90	40	>79.98	30
MB2003-Monoculture-1-Pectin_R	12,172,558	0	6-90	40	>78.76	27
MB2003-Monoculture-2-Pectin_F	12,549,560	0	8-90	40	>91.09	31
MB2003-Monoculture-2-Pectin_R	12,549,560	0	6-90	40	>91.75	30
MB2003-Monoculture-3-Pectin_F	12,482,716	0	8-90	40	>78.86	30
MB2003-Monoculture-3-Pectin_R	12,482,716	0	6-90	40	>77.05	27

Sample	Total Raw Sequences	Filtered Sequences	Sequence Length (bp)	GC (%)	Sequence Duplication Level (%)	Lowest Quality score across all bases (Phred64)
B316-Monoculture-1-Pectin_F	12,522,784	0	7-90	40	>96.01	30
B316-Monoculture-1-Pectin_R	12,522,784	0	6-90	39	>96.52	30
B316-Monoculture-2-Pectin_F	12,626,949	0	7-90	41	>96.24	30
B316-Monoculture-2-Pectin_R	12,626,949	0	6-90	40	>96.8	30
B316-Monoculture-3-Pectin_F	12,593,585	0	7-90	41	>96.54	30
B316-Monoculture-3-Pectin_R	12,593,585	0	6-90	41	>97.01	30
Co-culture-1-Pectin_F	12,643,059	0	8-90	41	>94.71	30
Co-culture-1-Pectin_R	12,643,059	0	6-90	40	>95.49	30
Co-culture-2-Pectin_F	12,642,031	0	8-90	40	>94.15	30
Co-culture-2-Pectin_R	12,642,031	0	7-90	38	>95.01	30
Co-culture-3-Pectin_F	12,612,627	0	7-90	41	>95.2	30
Co-culture-3-Pectin_R	12,612,627	0	6-90	41	>95.95	30

Table A3.10. Alignment details of trimmed reads to the *B. proteoclasticus* B316 genome.

Sample	Replicon	Total reads	Successfully aligned reads	Alignment rate	Aligning to protein coding genes	Aligning to rRNAs & tRNAs	Aligning to unannotated regions
B316-1-Xylan	chromosome	12,341,173	10,786,591	87%	55%	1%	44%
	chromid		512,623	4%	88%	5%	7%
	pCY360		333,254	3%	65%	50%	35%
	pCY186		111,792	1%	69%	0%	31%
B316-2-Xylan	chromosome	12,444,544	10,806,468	87%	60%	0%	40%
	chromid		568,124	5%	91%	2%	7%
	pCY360		369,116	3%	75%	0%	25%
	pCY186		99,308	1%	83%	0%	17%
B316-3-Xylan	chromosome	6,511,651	4,777,269	73%	59%	17%	24%
	chromid		226,087	3%	41%	57%	2%
	pCY360		126,437	2%	74%	0%	26%
	pCY186		40,020	1%	84%	0%	16%
Co-culture-1-Xylan	chromosome	12,357,812	8,482,270	69%	55%	1%	44%
	chromid		365,395	3%	88%	5%	7%
	pCY360		261,592	2%	71%	0%	28%
	pCY186		51,856	0%	84%	0%	16%
Co-culture-2-Xylan	chromosome	12,391,049	8,886,985	72%	58%	3%	39%
	chromid		388,832	3%	74%	21%	5%
	pCY360		269,751	2%	72%	0%	28%
	pCY186		57,841	0%	81%	0%	19%
Co-culture-3-Xylan	chromosome	12,307,514	8,632,051	70%	59%	0%	41%
	chromid		378,957	3%	90%	3%	7%
	pCY360		307,095	2%	72%	0%	28%
	pCY186		70,070	1%	80%	0%	20%
B316-1-Pectin	chromosome	12,522,784	11,510,831	92%	17%	0%	83%
	chromid		84,840	1%	88%	6%	6%
	pCY360		78,697	1%	65%	0%	35%
	pCY186		237,326	2%	66%	0%	34%
B316-2-Pectin	chromosome	12,626,949	11,628,625	92%	16%	0%	84%
	chromid		73,536	1%	87%	8%	5%
	pCY360		67,209	1%	64%	0%	36%
	pCY186		178,754	1%	64%	0%	36%

Sample	Replicon	Total reads	Successfully aligned reads	Alignment rate	Aligning to protein coding genes	Aligning to rRNAs & tRNAs	Aligning to unannotated regions
B316-3-Pectin	chromosome	12,593,585	11,637,074	92%	17%	0%	83%
	chromid		59,333	0%	80%	15%	5%
	pCY360		58,749	0%	64%	0%	36%
	pCY186		161,925	1%	64%	0%	36%
Co-culture-1-Pectin	chromosome	12,643,059	4,342,452	34%	18%	1%	81%
	chromid		41,169	0%	80%	15%	5%
	pCY360		30,593	0%	61%	0%	39%
	pCY186		74,127	1%	63%	0%	37%
Co-culture-2-Pectin	chromosome	12,642,031	6,963,420	55%	17%	0%	83%
	chromid		100,092	1%	90%	4%	6%
	pCY360		74,974	1%	68%	0%	32%
	pCY186		177,888	1%	69%	0%	31%
Co-culture-3-Pectin	chromosome	12,612,627	3,218,381	26%	23%	1%	76%
	chromid		19,351	0%	63%	33%	4%
	pCY360		22,256	0%	60%	0%	40%
	pCY186		24,528	0%	62%	0%	38%

Table A3.11. Alignment details of trimmed reads to the *B. hungatei* MB2003 genome.

Sample	Replicon	Total reads	Successfully aligned reads	Alignment rate	Aligning to protein coding genes	Aligning to rRNAs & tRNAs	Aligning to unannotated regions
MB2003-1-Xylan	chromosome	12,690,950	11,444,411	90%	61%	0%	39%
	chromid		311,602	2%	92%	0%	8%
	pNP144		305,824	2%	56%	0%	44%
	pNP6		0	0%	0%	0%	0%
MB2003-2-Xylan	chromosome	12,718,791	11,526,634	91%	62%	0%	38%
	chromid		271,463	2%	92%	0%	8%
	pNP144		289,566	2%	57%	0%	43%
	pNP6		0	0%	0%	0%	0%
MB2003-3-Xylan	chromosome	12,704,236	11,337,588	89%	57%	0%	43%
	chromid		291,213	2%	92%	0%	8%
	pNP144		326,613	3%	52%	0%	48%
	pNP6		0	0%	0%	0%	0%
Co-culture-1-Xylan	chromosome	12,357,812	2,503,217	20%	63%	0%	37%
	chromid		48,799	0%	95%	0%	5%
	pNP144		77,400	1%	59%	0%	41%
	pNP6		0	0%	0%	0%	0%
Co-culture-2-Xylan	chromosome	12,391,049	2,156,024	17%	58%	0%	42%
	chromid		35,529	0%	95%	0%	5%
	pNP144		59,704	0%	56%	0%	44%
	pNP6		0	0%	0%	0%	0%
Co-culture-3-Xylan	chromosome	12,307,514	2,256,725	18%	61%	0%	39%
	chromid		42,990	0%	94%	0%	6%
	pNP144		78,891	1%	56%	0%	44%
	pNP6		0	0%	0%	0%	0%
MB2003-1-Pectin	chromosome	12,172,558	10,183,368	84%	67%	0%	33%
	chromid		229,215	2%	91%	0%	9%
	pNP144		607,013	5%	81%	0%	19%
	pNP6		0	0%	0%	0%	0%
MB2003-2-Pectin	chromosome	12,549,560	11,277,678	90%	47%	0%	53%
	chromid		153,916	1%	90%	0%	10%
	pNP144		373,595	3%	64%	0%	36%
	pNP6		0	0%	0%	0%	0%

Sample	Replicon	Total reads	Successfully aligned reads	Alignment rate	Aligning to protein coding genes	Aligning to rRNAs & tRNAs	Aligning to unannotated regions
MB2003-3-Pectin	chromosome	12,482,716	10,838,763	87%	88%	0%	12%
	chromid		299,357	2%	93%	0%	7%
	pNP144		706,299	6%	87%	0%	13%
	pNP6		1	0%	0%	0%	0%
Co-culture-1-Pectin	chromosome	12,643,059	7,400,822	59%	27%	0%	73%
	chromid		30,196	0%	94%	0%	6%
	pNP144		69,630	1%	70%	0%	30%
	pNP6		0	0%	0%	0%	0%
Co-culture-2-Pectin	chromosome	12,642,031	4,566,428	36%	28%	0%	72%
	chromid		13,297	0%	94%	0%	6%
	pNP144		38,388	0%	68%	0%	32%
	pNP6		0	0%	0%	0%	0%
Co-culture-3-Pectin	chromosome	12,612,627	8,586,442	68%	26%	0%	74%
	chromid		28,553	0%	94%	0%	6%
	pNP144		69,241	1%	74%	0%	26%
	pNP6		0	0%	0%	0%	0%

Table A3.12. Transcript abundance for *B. proteoclasticus* B316 grown in mono- and co-culture on xylan and pectin.

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bpr_II101	-	IS200/IS605 family transposase	1847	1748	882	1228
bpr_I1733	*pyrI*	aspartate carbamoyltransferase	142	504	214	405
bpr_I2620	*thiI*	thiamine biosynthesis/tRNA modification protein	38	115	193	437
bpr_I2578	-	NAD-dependent epimerase/dehydratase	136	357	66	197
bpr_II236	-	DNA-binding protein	232	273	4	5
bpr_I0288	*dnaD*	primosome protein	101	205	79	101
bpr_I1083	*xyn10D*	endo-1,4-β-xylanase	185	86	69	95
bpr_II096	-	ISNCY family transposase	161	139	23	39
bpr_II097	-	ISNCY family transposase	161	139	23	39
bpr_I0750	-	two component system histidine kinase	79	47	69	131
bpr_I1144	*purC*	phosphoribosylaminoimidazole-succinocarboxamide synthase	49	106	65	78
bpr_I2692	-	LysR family transcriptional regulator	89	154	16	21
bpr_I0889	*dusA1*	tRNA dihydrouridine synthase A	126	95	6	10
bpr_I0242	-	molecular chaperone Hsp20	51	106	14	40
bpr_I2515	*ugl88B*	unsaturated glucuronyl hydrolase	71	56	33	50
bpr_I1020	-	ferric uptake regulator family protein	94	76	2	9
bpr_I1430	*dacA1*	D-alanyl-D-alanine carboxypeptidase	53	33	24	64
bpr_I2800	-	Rrf2 family transcriptional regulator	67	65	14	25
bpr_I2922	-	thioredoxin	85	82	1	2
bpr_III252	-	DNA-binding protein	88	79		1
bpr_I1684	-	sugar phosphate isomerase/epimerase	70	91	1	2
bpr_I2112	-	guanylate cyclase	29	50	24	47
bpr_I1496	-	UvrD/REP helicase	86	45	3	4
bpr_I1505	-	ACT domain-containing protein	75	53	1	1
bpr_I2332	-	IS200/IS605 family transposase	69	45	2	6
bpr_I0030	*murG*	UDP-diphospho- muramoylpentapeptide β-N-acetylglucosaminyltransferase	54	59	1	3
bpr_I1732	*pyrB*	aspartate carbamoyltransferase	27	36	14	36
bpr_III035	-	CorA-like Mg2+ transporter protein	53	54	1	5
bpr_I0468	-	molecular chaperone Hsp20	53	53	1	2
bpr_III248	-	sugar ABC transporter SBP	51	53	1	1
bpr_I1718	-	sugar ABC transporter permease	46	53	1	1
bpr_I1583	-	xylose isomerase domain-containing protein	44	51	1	4
bpr_I0489	*fliC2*	flagellin	36	50	4	9
bpr_I2923	*trmB*	tRNA (guanine-N1)-methyltransferase	49	48	1	1
bpr_I2908	*divIC*	septum formation initiator	56	40	1	1
bpr_I0245	-	HAMP/SpoIIE domain-containing protein	36	25	7	28
bpr_I2921	-	bacterial SH3 domain-containing protein	58	30	1	2
bpr_I2225	*fnt*	formate/nitrite transporter	23	18	16	30
bpr_I1816	-	response regulator domain-containing protein	50	34		
bpr_I1384	*cheB*	chemotaxis protein	34	39	1	3
bpr_I0637	-	IS110 family transposase	42	35		
bpr_I1125	-	IS110 family transposase	42	35		
bpr_I1132	-	IS110 family transposase	42	35		
bpr_I2849	-	IS110 family transposase	42	35		
bpr_III157	-	α/β fold family hydrolase	22	21	17	14
bpr_I0262	-	anti-sigma regulatory factor	39	32	1	2
bpr_I0974	*rpsO*	30S ribosomal protein S15	47	20		
bpr_III209	*bga2B*	β-galactosidase	26	32	4	3
bpr_I0751	-	IS200/IS605 family transposase	34	27	1	2
bpr_III092	-	IS110 family transposase	33	28		
bpr_IV010	-	IS110 family transposase	33	28		

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bpr_I0198	-	MatE efflux family protein	22	21	7	11
bpr_I2050	*gap*	glyceraldehyde-3-phosphate dehydrogenase	28	32		
bpr_I1740	*xyn10E*	endo-1,4-β-xylanase	16	21	11	11
bpr_I0823	-	GDSL-family lipase/acylhydrolase	33	20	2	4
bpr_I2243	-	ABC transporter ATP-binding protein	27	26	1	4
bpr_I1222	*rpsU*	30S ribosomal protein S21	39	18		
bpr_I1035	-	GNAT family acetyltransferase	31	23	1	1
bpr_I0796	-	bacteriocin signal sequence domain-containing protein	3	1	28	22
bpr_III009	-	MatE efflux family protein	16	29	3	5
bpr_III235	*pyrE2*	orotate phosphoribosyltransferase	27	24		1
bpr_I0104	-	phosphocarrier HPr family protein	47	5		
bpr_I1720	-	sugar ABC transporter SBP	35	15		1
bpr_I0467	*sodA*	superoxide dismutase	18	29	1	2
bpr_IV100	*hsdS2*	type I restriction modification system S subunit	24	18	3	4
bpr_I1466	*rpmB*	50S ribosomal protein L28	23	25		
bpr_I1429	-	cNMP binding domain-containing protein	23	17	3	5
bpr_I1734	-	SCP domain-containing protein	21	18	4	5
bpr_I2116	-	sugar ABC transporter SBP	19	29		
bpr_I2805	-	GNAT family acetyltransferase	19	23	2	2
bpr_III227	-	GntR family transcriptional regulator	26	19		1
bpr_I1283	*rpoZ*	DNA-directed RNA polymerase omega subunit	20	20	2	2
bpr_III074	-	TPR domain-containing protein	20	11	3	10
bpr_III019	-	AraC family transcriptional regulator	19	23		1
bpr_I2489	-	ribosomal subunit interface protein	22	20	1	
bpr_I0292	-	SpoVG family protein	16	21	3	3
bpr_II424	*nrdD3*	anaerobic ribonucleoside triphosphate reductase	28	10	1	2
bpr_I1149	-	degV family protein	12	24	2	3
bpr_I0031	*rbr1*	rubrerythrin	18	22		1
bpr_III139	-	TetR family transcriptional regulator	14	14	4	9
bpr_I0276	-	cell wall hydrolase	28	11		
bpr_I2903	*fbaA*	fructose-1,6-bisphosphate aldolase	21	16	1	
bpr_I0495	-	glycosyl transferase 2	14	16	2	6
bpr_I0509	-	LicD family protein	14	15	2	7
bpr_I0269	-	pyruvate:ferredoxin oxidoreductase	10	12	6	10
bpr_I2303	*rpmG*	50S ribosomal protein L33	26	10		
bpr_I0728	*cel5D*	endo-1,4-β-glucanase	15	18	1	2
bpr_I0571	-	glycoprotease	14	18	2	2
bpr_III034	-	acetyltransferase	17	14	1	3
bpr_I1944	-	diguanylate phosphodiesterase	17	16		2
bpr_I1360	*fliH*	flagellar assembly protein	18	15	1	1
bpr_I1949	-	NUDIX hydrolase	8	11	5	11
bpr_I2364	*tufA*	translation elongation factor Tu	17	18		
bpr_I1689	-	AraC family transcriptional regulator	15	10	4	5
bpr_I1883	-	GGDEF/PAS domain-containing protein	17	15	1	1
bpr_I0565	-	HD-GYP domain-containing protein	9	14	4	7
bpr_I1721	-	LysM domain-containing protein	15	14	1	4
bpr_II014	-	PIN domain-containing protein	15	18		1
bpr_I2104	-	PTS system IIABC fructose-specific family protein	7	27		
bpr_I0800	-	anti-sigma factor antagonist	14	13	2	4
bpr_III123	-	archaeal ATPase family protein	20	13		
bpr_I0197	*arcC*	carbamate kinase	5	12	6	10
bpr_III011	*pta*	phosphate acetyltransferase	13	17	1	2
bpr_I0510	-	ribulose-phosphate 3-epimerase	18	15		

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bpr_I2864	-	ABC transporter ATP-binding protein/permease	14	10	4	4
bpr_I0261	-	anti-sigma factor antagonist	17	15		
bpr_I0176	-	AraC family transcriptional regulator	10	7	7	8
bpr_I2059	-	response regulator domain-containing protein	16	16		
bpr_I0244	*vanX*	D-alanyl-D-alanine dipeptidase	11	19		1
bpr_I0888	*cysE*	serine acetyltransferase	18	12		1
bpr_I2048	*tpiA*	triosephosphate isomerase	11	20		
bpr_I0268	-	ferredoxin	14	9	4	3
bpr_I1425	-	iron-containing alcohol dehydrogenase	12	18		
bpr_I2242	-	radical SAM domain-containing protein	16	14		
bpr_I0207	-	citrate transporter	13	10	2	4
bpr_I0130	-	hydroxypyruvate reductase	11	10	3	5
bpr_I1564	-	RNA binding protein	19	10		
bpr_I0673	-	ABC transporter ATP-binding protein	9	16	1	2
bpr_I0167	*atpG1*	ATP synthase F1 subunit gamma	12	16		
bpr_I1560	-	Bmp family protein	11	17		
bpr_I2033	*cheR*	chemotaxis protein	15	13		
bpr_III028	*nagB*	glucosamine-6-phosphate isomerase	15	10	1	2
bpr_I2927	-	glutathione peroxidase	16	12		
bpr_I2687	-	TetR family transcriptional regulator	18	10		
bpr_I0038	-	two component system response regulator	14	14		
bpr_I0631	*rpsI*	30S ribosomal protein S9	15	12		
bpr_I1636	-	acyltransferase	14	11	1	1
bpr_I1783	-	AraC family transcriptional regulator	10	10	3	4
bpr_I0037	-	two component system histidine kinase	15	12		
bpr_III253	-	two component system histidine kinase/response regulator hybrid protein	8	13	2	4
bpr_I2249	-	ATPase AAA	14	10	1	1
bpr_I1388	*cheD*	chemotaxis protein	7	16	1	2
bpr_I1383	-	flagellar protein	12	13		1
bpr_I1756	-	GGDEF domain-containing protein	13	9	2	2
bpr_I0290	*glgC1*	glucose-1-phosphate adenylyltransferase	13	12		1
bpr_I2672	-	GNAT family acetyltransferase	5	1	4	16
bpr_I0129	-	GntP family transporter	7	12	2	5
bpr_I0937	-	sugar ABC transporter SBP	7	18		1
bpr_I1098	-	translation elongation factor EF-G-like protein	15	11		
bpr_I2631	-	aspartate/tyrosine/aromatic aminotransferase	5	3	8	9
bpr_I2484	*etfB*	electron transfer flavoprotein beta subunit	10	15		
bpr_I1374	*fliY*	flagellar motor switch protein	12	11	1	1
bpr_III244	-	sugar ABC transporter SBP	10	8	1	6
bpr_I2772	*plsC2*	1-acyl-sn-glycerol-3-phosphate acyltransferase	15	9		
bpr_I2863	-	CAAX amino terminal protease	2	5	6	11
bpr_I1237	*cheW*	chemotaxis protein	12	12		
bpr_I1382	*flhG*	flagellar biosynthesis protein	10	11	1	2
bpr_I0220	-	FRG domain-containing protein	12	10	1	1
bpr_I1047	-	HDOD domain-containing protein	10	14		
bpr_I2106	-	PTS system I PEP-phosphotransferase	5	19		
bpr_I1173	-	sugar ABC transporter SBP	18	5		1
bpr_I1162	*atpD2*	ATP synthase F1 subunit beta	11	12		
bpr_I1016	-	LrgB-like family protein	7	7	4	5
bpr_I2049	*pgk*	phosphoglycerate kinase	9	14		
bpr_I0591	-	RelA/SpoT domain-containing protein	7	4	3	9
bpr_I2010	-	sugar ABC transporter SBP	11	12		
bpr_I2443	-	sugar ABC transporter SBP	6	13	2	2

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bpr_I2693	-	4Fe-4S ferredoxin	14	8		
bpr_III083	-	α/β fold family hydrolase	11	9	1	1
bpr_I1160	*atpA2*	ATP synthase F1 subunit alpha	10	12		
bpr_III238	-	HTH domain-containing protein	10	7		5
bpr_III143	-	thioesterase family protein	13	8	1	
bpr_I1209	*alr*	alanine racemase	13	8		
bpr_I0168	*atpD1*	ATP synthase F1 subunit beta	6	15		
bpr_I0165	*atpH1*	ATP synthase F1 subunit delta	7	14		
bpr_I0743	-	iron ABC transporter ATP-binding protein	10	11		
bpr_I0511	-	LicD family protein	6	9	3	3
bpr_I0270	-	MecA family protein	8	9	2	2
bpr_I1759	-	PilZ domain-containing protein	12	8		1
bpr_I2824	-	SAM-dependent methyltransferase	1	2	12	6
bpr_I2133	*efp*	translation elongation factor P	13	8		
bpr_III002	-	AAA family ATPase	9	11		
bpr_I0793	*mcpE*	chemotaxis protein	8	12		
bpr_I2388	*dapF*	diaminopimelate epimerase	10	6	1	3
bpr_I1960	-	diguanylate phosphodiesterase	10	10		
bpr_I2474	-	HD-GYP/response regulator domain-containing protein	9	11		
bpr_I2103	*pfkB*	1-phosphofructokinase	7	12		
bpr_I2035	*rpsA*	50S ribosomal protein L9	12	7		
bpr_I0020	-	aldo/keto reductase	9	8	1	1
bpr_I2924	*dacA3*	D-alanyl-D-alanine carboxypeptidase	5	10	2	2
bpr_I2102	-	DeoR family transcriptional regulator	9	10		
bpr_I1070	-	DNA-binding protein	10	9		
bpr_I2405	-	GTP-binding protein	11	7		1
bpr_I1870	-	MraZ protein	14	5		
bpr_I1552	*lepB3*	signal peptidase I	10	9		
bpr_I2655	-	thioesterase	10	7	1	1
bpr_I0674	-	ABC transporter permease	5	9	1	3
bpr_I2865	-	ABC transporter permease	8	6	1	3
bpr_I2338	-	ANTAR domain-containing protein	9	7		2
bpr_I0928	-	cell wall hydrolase	8	8	1	1
bpr_I1961	-	GGDEF/cache/HAMP domain-containing protein	5	5	2	6
bpr_I1148	-	Hpt domain-containing protein	10	8		
bpr_I2274	-	ion transport protein	9	8		1
bpr_I0508	-	LicD family protein	10	7		1
bpr_I2579	-	SCP domain-containing protein	7	9		2
bpr_I0182	-	sugar ABC transporter SBP	7	11		
bpr_I0224	*pfkA1*	6-phosphofructokinase	7	9		1
bpr_I1331	-	6-phosphogluconolactonase	10	7		
bpr_I1158	*atpE2*	ATP synthase F0 subunit C	8	9		

Transcript abundance is reported as RPKM values. Genes are ranked based on the total number of transcripts per gene. The top 200 genes with the highest overall transcript abundance are included. Abbreviations: XC1, B316 grown in mono-culture on xylan; XC2, B316 grown in co-culture on xylan; PC1, B316 grown in mono-culture on pectin; PC2, B316 grown in co-culture on pectin. Genes that did not possess a single transcript or annotated as "hypothetical proteins" in any of the treatment groups have been excluded from the analysis. The degree of redness indicates relative transcript abundance across all treatments and black indicates an absence of transcripts for that gene.

Table A3.13. Transcript abundance for *B. hungatei* MB2003 grown in mono- and co-culture on xylan and pectin.

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bhn_III24	-	MatE efflux family protein	5370	7355	5379	10090
bhn_II106	-	DNA-binding protein	3531	12629	7766	1743
bhn_III79	-	ATP-binding protein	7522	2500	6597	6717
bhn_III90	-	MFS transporter	119	914	16108	328
bhn_III30	-	ATPase AAA family	571	5503	5930	4283
bhn_III16	-	DegV family protein	855	3624	4784	2819
bhn_III86	-	HicA family protein	28	644	10536	46
bhn_III18	-	anti-sigma factor antagonist	112	246	9786	52
bhn_III88	-	GNAT family acetyltransferase/ NUDIX domain- containing protein	13	145	9594	54
bhn_III34	-	DNA-binding protein	1041	1587	5091	970
bhn_III42	*accB*	acetyl-CoA carboxylase biotin protein			8643	
bhn_III21	-	phosphoribosyltransferase domain-containing protein	123	997	7162	351
bhn_III45	*accD*	acetyl-CoA carboxylase carboxyl transferase beta subunit	5		7764	
bhn_III47	*ascA*	acetyl-CoA synthetase	12	116	7233	355
bhn_III87	-	HicB family protein	52	559	6690	43
bhn_III44	*accC*	acetyl-CoA carboxylase biotin carboxylase	5	17	7209	
bhn_III70	-	HTH domain-containing protein	28	99	7063	16
bhn_III39	*fabD*	malonyl CoA-acyl transacylase	6	5	7145	
bhn_III51	-	ACT domain-containing protein	13		7041	20
bhn_III89	-	CrcB-like protein	12	117	6852	38
bhn_III41	*fabF*	3-oxoacyl-(acyl-carrier-protein) synthase	7	5	7004	
bhn_III43	*fabZ*	β-hydroxyacyl-(acyl-carrier-protein) dehydratase	1		6930	
bhn_III02	-	radical SAM domain-containing protein	25	112	6688	
bhn_III55	*rsmH*	16*S* rRNA (cytosine(1402)- N(4))-methyltransferase	10	29	6535	9
bhn_III73	-	histidine phosphatase superfamily protein	12	9	6528	
bhn_III40	*fabG*	3-oxoacyl-(acyl-carrier-protein) reductase	1	43	6367	
bhn_III28	*hypA*	hydrogenase nickel insertion protein			6378	
bhn_III46	*accA*	acetyl-CoA carboxylase carboxyl transferase alpha subunit		5	6266	
bhn_III83	-	transposase	7	22	5981	48
bhn_III84	-	AAA domain-containing protein	7		5891	3
bhn_II101	-	HTH domain-containing protein	4	43	5809	
bhn_III007	-	HTH domain-containing protein	3	21	5671	7
bhn_III001	-	TetR family transcriptional regulator	135	1509	3092	852
bhn_III80	-	sugar ABC transporter SBP	245	217	4922	158
bhn_III32	*parA*	plasmid partitioning protein	6	28	5431	19
bhn_III31	*repB*	replication initiation protein	8	10	5283	
bhn_III004	-	phage integrase family protein	11	55	5054	52
bhn_III003	-	SBP/EAL domain-containing protein	32	237	4816	34
bhn_III14	-	DNA-binding protein	918	961	2526	708
bhn_II049	-	transposase	40	229	4551	56
bhn_III49	-	sugar ABC transporter SBP	70	351	4021	353
bhn_III78	*est1A*	feruloyl esterase	120	83	4512	34
bhn_III10	*bga42A*	β-galactosidase	59	104	4454	105
bhn_II073	-	phage integrase family protein	130	408	3430	621
bhn_II020	-	sortase family protein	12	28	4538	

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bhn_III63	-	potassium uptake protein TrkH family	57	109	4068	155
bhn_III48	-	methyl-accepting chemotaxis protein	96	522	3262	494
bhn_II009	*parA*	partitioning protein	16	29	4289	31
bhn_III77	-	GNAT family acetyltransferase	7		4323	24
bhn_III11	-	PAP2 family protein	11	30	4266	
bhn_III76	*lyc25B*	lysozyme	7	8	4283	5
bhn_III64	*bgl3A*	β-glucosidase	2	4	4183	5
bhn_II147	*repB*	replication initiation protein	48	209	3879	57
bhn_III52	-	aminotransferase	12	22	4116	9
bhn_III66	-	Na+/H+ antiporter	16	296	3815	13
bhn_II125	-	phosphoadenosine phosphosulfate reductase family protein	1	7	4078	
bhn_II004	-	restriction endonuclease	48	55	3878	7
bhn_II092	-	Fic/DOC family protein	644	826	2096	409
bhn_III25	-	GGDEF domain-containing protein	47	203	3330	378
bhn_II143	-	HTH domain-containing protein	559	201	2184	934
bhn_III38	*fabK*	enoyl-(acyl-carrier-protein) reductase II	2	5	3836	
bhn_III23	*pta*	phosphate acetyltransferase	675	655	1846	603
bhn_II001	-	cell surface protein	4	10	3635	8
bhn_II022	-	signal peptidase I	8	14	3623	12
bhn_III65	-	UBA/TS-N domain-containing protein	14	240	3217	
bhn_II019	-	cell surface protein	3	15	3369	2
bhn_II008	-	TraG/TraD family protein	6	6	3369	1
bhn_III36	-	MerR family transcriptional regulator	18	5	3299	27
bhn_II036	-	type IV secretion system protein	3	8	3306	5
bhn_III12	-	acyltransferase	7	23	3274	13
bhn_III75	-	metallo-β-lactamase family protein	12	55	3183	38
bhn_II072	-	phage integrase family protein	10	35	3192	9
bhn_II086	*cas2*	CRISPR-associated protein	147	118	2920	36
bhn_III59	-	ATP-dependent DNA helicase UvrD/REP family	38	45	3131	5
bhn_III81	-	methyl-accepting chemotaxis protein	39	55	3020	64
bhn_II087	*cpf1*	CRISPR-associated protein	24	30	3020	28
bhn_III50	-	response regulator domain-containing protein	40	7	3007	41
bhn_III62	-	potassium uptake protein TrkA family	52	130	2680	145
bhn_II069	-	DNA polymerase III delta' subunit	56	75	2830	19
bhn_III74	-	TauE family protein	10		2933	
bhn_III57	-	SPFH domain/band 7 family protein	36	265	2599	37
bhn_III56	-	radical SAM domain-containing protein	1	6	2724	
bhn_III35	-	MatE efflux family protein	48	76	2470	114
bhn_II089	-	archaeal ATPase family protein	8	18	2590	7
bhn_III54	-	GIY-YIG domain-containing protein	12		2569	
bhn_II064	-	DNA polymerase III delta subunit	2	3	2541	4
bhn_II060	-	ThiF family protein	10		2429	9
bhn_II021	-	type II/IV secretion system protein	12	6	2281	5
bhn_I2353	*fliC*	flagellin	532	732	571	417
bhn_II006	-	DNA/RNA non-specific endonuclease	23	51	2160	11
bhn_II111	-	HD family phosphohydrolase	28	4	2085	18
bhn_II065	-	single-stranded DNA binding protein	12		2006	
bhn_II070	*dnaX*	DNA polymerase III gamma subunits	10	13	1975	15
bhn_III37	-	acyl carrier protein			1983	
bhn_I1037	-	YolD-like protein	819	558	300	193
bhn_III008	-	TetR family transcriptional regulator			1801	
bhn_III61	-	CorA-like Mg2+ transporter protein	2		1795	
bhn_II063	-	ATP-dependent DNA helicase UvrD/REP family	1	1	1607	1

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bhn_III72	*est4A*	polysaccharide deacetylase	1	9	1472	
bhn_I2379	-	voltage gated chloride channel protein	818	118	193	53
bhn_II094	-	DNA-binding protein	10	34	914	
bhn_I2774	-	GGDEF/PAS protein	17	305	94	167
bhn_I2657	-	Rrf2 family transcriptional regulator	50	235	109	51
bhn_I0225	-	GNAT family acetyltransferase			436	
bhn_I0870	-	DnaD domain-containing protein	114	73	115	19
bhn_I2545	-	two component system histidine kinase	1	207	113	
bhn_I2777	-	ABC transporter SBP	16	167	89	32
bhn_I0453	-	CTP:phosphocholine cytidylyltransferase	1		300	
bhn_I0228	-	MFS transporter	1	8	274	
bhn_I1799	-	MFS transporter	2	9	271	
bhn_I1162	*rpsU*	ribosomal protein S21	12	72	191	
bhn_I0497	-	glycosyl transferase GT4 family			260	
bhn_I0255	-	α/β hydrolase family protein			252	
bhn_I2222	-	sugar ABC transporter SBP	19	98	117	17
bhn_I1892	-	endonuclease V			249	
bhn_I1836	*rhaM*	L-rhamnose mutarotase			238	
bhn_I0292	-	4Fe-4S binding protein			232	
bhn_I1227	-	HNH endonuclease family protein			229	
bhn_I2080	*map*	methionine aminopeptidase	1	2	226	
bhn_I2636	-	NUDIX family hydrolase			222	
bhn_I2601	-	RNA polymerase sigma factor sigma-70 family	1	47	171	1
bhn_I0093	-	ROK family protein			219	
bhn_I0240	*fliC*	flagellin	8	146	60	4
bhn_I2354	-	SAM-dependent methyltransferase			218	
bhn_I2240	-	phosphodiesterase			215	
bhn_I2772	*xsa43E*	xylosidase/arabinofuranosidase			215	
bhn_I0895	*aroC*	chorismate synthase		3	211	
bhn_I1493	-	PspC domain-containing protein	1	3	210	
bhn_I1792	-	SAM-dependent methyltransferase			214	
bhn_I2735	-	flavin reductase	15	32	162	3
bhn_I2362	*csrA*	carbon storage regulator	8	16	185	1
bhn_I2367	*flgN*	flagellar biosynthesis protein	5	66	134	3
bhn_I0157	-	ATP synthase F0 I subunit	7	1	198	
bhn_I0971	-	α/β hydrolase family protein	6	78	106	14
bhn_I1233	-	xylose ABC transporter SBP			204	
bhn_I0272	-	NADPH-dependent FMN reductase			203	
bhn_I0065	-	prepilin peptidase A24 family			202	
bhn_I1863	*ispH*	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	1		200	
bhn_I0336	-	comEA protein		2	198	
bhn_I0594	-	glycosyl transferase			200	
bhn_I1675	-	GNAT family acetyltransferase			200	
bhn_I1797	-	PfkB family carbohydrate kinase	41	29	118	12
bhn_I2315	-	SAM-dependent methyltransferase			200	
bhn_I0253	-	SAM-dependent methyltransferase			199	
bhn_I2368	*flgM*	flagellar biosynthesis anti-sigma factor	3	110	83	2
bhn_I1369	*atpC*	ATP synthase F1 epsilon subunit		1	196	
bhn_I1591	*rpmA*	ribosomal protein L27	4		193	
bhn_I2129	-	two component system histidine kinase			197	
bhn_I2325	-	cupin domain-containing protein			195	
bhn_I2615	-	glycerol-3-phosphate cytidylyltransferase			195	
bhn_I0861	-	AraC family transcriptional regulator	34	15	142	3
bhn_I2153	-	CDP-diglyceride synthetase	2	1	191	
bhn_I2065	*rplB*	ribosomal protein L2	1		192	

Locus tag	Name	Annotation	XC1	XC2	PC1	PC2
bhn_I2099	*xsa43C*	xylosidase/arabinofuranosidase			193	
bhn_I2336	-	glycosyl transferase GT2 family	31	22	125	13
bhn_I0467	*acpP*	acyl carrier protein			190	
bhn_I1121	-	GGGtGRT family protein	8	73	67	41
bhn_I0322	-	histidine phosphatase			188	
bhn_I0582	-	HNH endonuclease family protein			188	
bhn_I1808	-	L-serine dehydratase		1	187	
bhn_I0159	*atpE*	ATP synthase F0 C subunit	2	5	180	
bhn_I0478	-	RNA binding protein	4	1	182	
bhn_I1280	-	SAM-dependent methyltransferase			187	
bhn_I2754	*trxA*	thioredoxin	5	4	178	
bhn_I0468	-	MaoC family protein			186	
bhn_I2258	*arf51A*	α-L-arabinofuranosidase			186	
bhn_I0271	-	α/β hydrolase family protein			185	
bhn_I1330	*gh31A*	glycoside hydrolase family 31			185	
bhn_I1604	-	metal-sensitive transcriptional repressor			185	
bhn_I2282	*coaX*	pantothenate kinase			185	
bhn_I1029	-	peptidase M29 family			184	
bhn_I0247	-	phospholipid methyltransferase			183	
bhn_I0249	-	AbrB family transcriptional regulator		7	175	
bhn_I0296	-	α/β hydrolase family protein			182	
bhn_I0560	-	ATPase YjeE family		1	181	
bhn_I0942	-	HTH domain-containing protein			182	
bhn_I2044	*rplX*	ribosomal protein L24			182	
bhn_I2091	-	WYL domain-containing protein			182	
bhn_I0482	-	ADP-ribosylglycohydrolase			181	
bhn_I0931	-	transposase			181	
bhn_I0380	-	GNAT family acetyltransferase			180	
bhn_I0593	-	QueT family transporter	61	35	79	5
bhn_I0851	*recR*	recombination protein			180	
bhn_I2276	*scr32A*	sucrose-6-phosphate hydrolase			180	
bhn_I2590	-	sugar-phosphate isomerase	5	19	153	3
bhn_I0610	*rplA*	ribosomal protein L1	1	1	177	
bhn_I0097	*iolB*	5-deoxy-glucuronate isomerase			178	
bhn_I0235	-	CAAX amino terminal protease	1	4	172	1
bhn_I0788	-	PAP2 family protein	6	2	170	
bhn_I0873	-	PilZ domain-containing protein	1	7	168	2
bhn_I1073	-	SAM-dependent methyltransferase			178	
bhn_I0184	-	AraC family transcriptional regulator	33	32	109	3
bhn_I1243	-	ATP-dependent DNA helicase	20	39	111	7
bhn_I1334	-	CobW/HypB/UreG nucleotide-binding domain-containing protein	38	3	133	3
bhn_I1505	-	pyridine nucleotide-disulphide oxidoreductase			176	
bhn_I0084	*gh27A*	glycoside hydrolase family 27			175	
bhn_I0242	-	ACT domain-containing protein	1	2	170	
bhn_I0473	-	glycosyl transferase GT4 family			173	
bhn_I1325	-	GNAT family acetyltransferase			173	

Transcript abundance is reported as RPKM values. Genes are ranked based on the total number of transcripts per gene. The top 200 genes with the highest overall transcript abundance are included. Abbreviations: XC1, MB2003 grown in mono-culture on xylan; XC2, MB2003 grown in co-culture on xylan; PC1, MB2003 grown in mono-culture on pectin; PC2, MB2003 grown in co-culture on pectin. Genes that did not possess a single transcript or annotated as "hypothetical proteins" in any of the treatment groups have been excluded from the analysis. The degree of redness indicates relative transcript abundance across all treatments and black indicates an absence of transcripts for that gene.

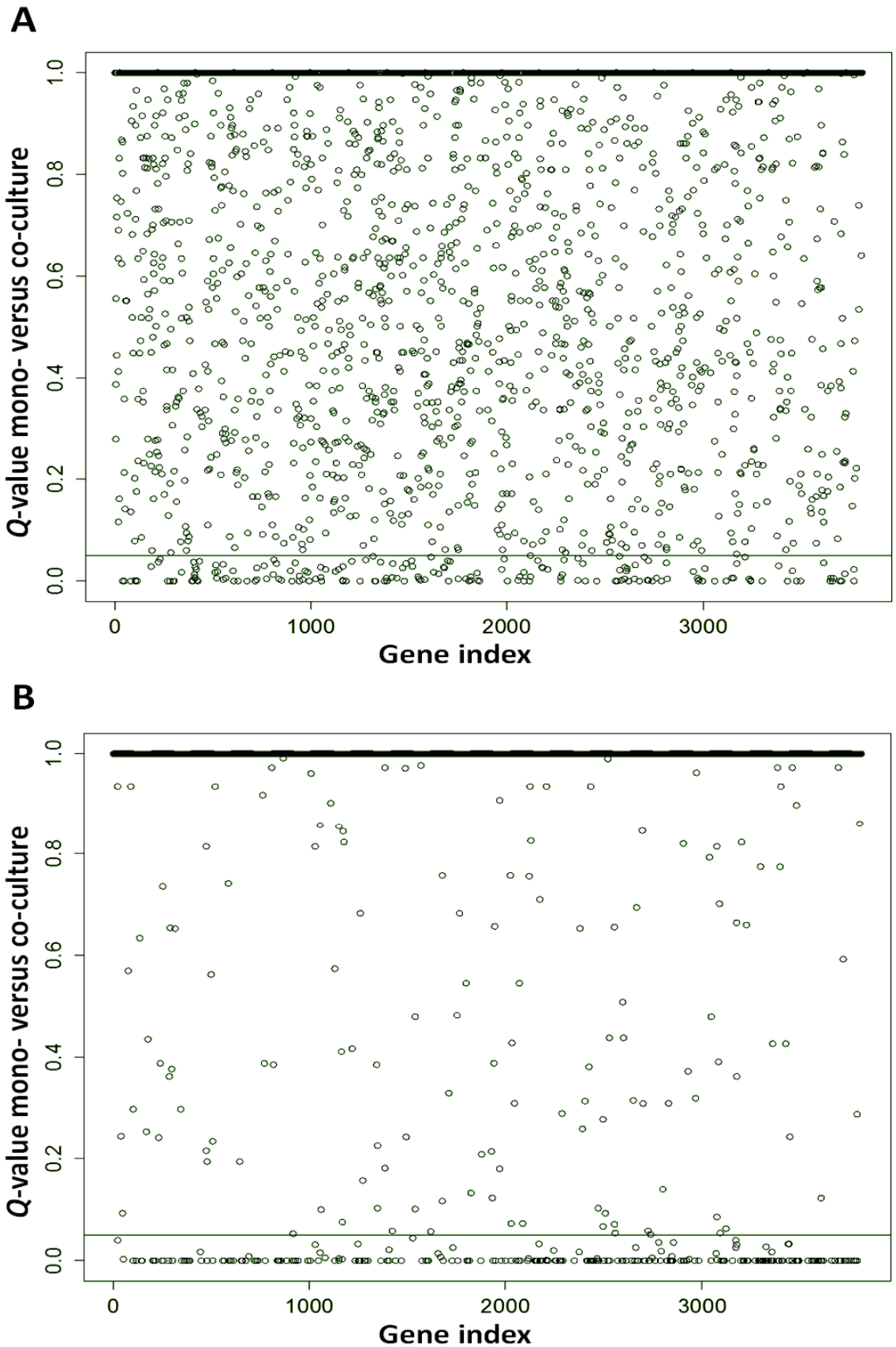


Figure A3.3. Distribution of all genes in *B. proteoclasticus* B316 based on *Q-value* from Rockhopper transcriptome assembly. A, transcripts of B316 grown on xylan; B, transcripts of B316 grown on pectin. The *Q-value* cut-off of < 0.05 is marked with line spanning the x axis.

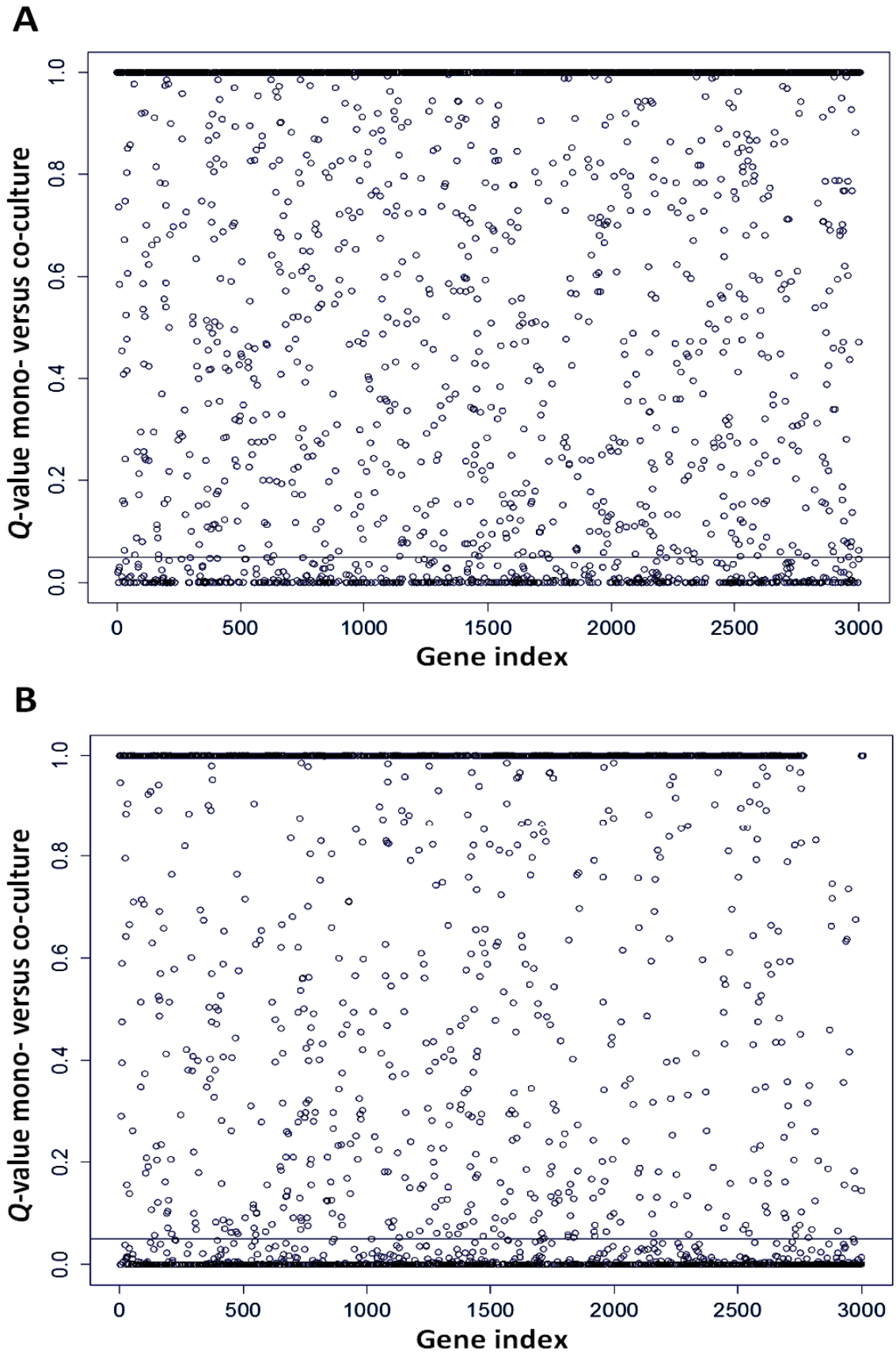


Figure A3.4. Distribution of all genes in *B. hungatei* MB2003 based on *Q-value* from Rockhopper transcriptome assembly. A, transcripts of MB2003 grown on xylan; B, transcripts of MB2003 grown on pectin. The *Q-value* cut-off of < 0.05 is marked with line spanning the x axis.

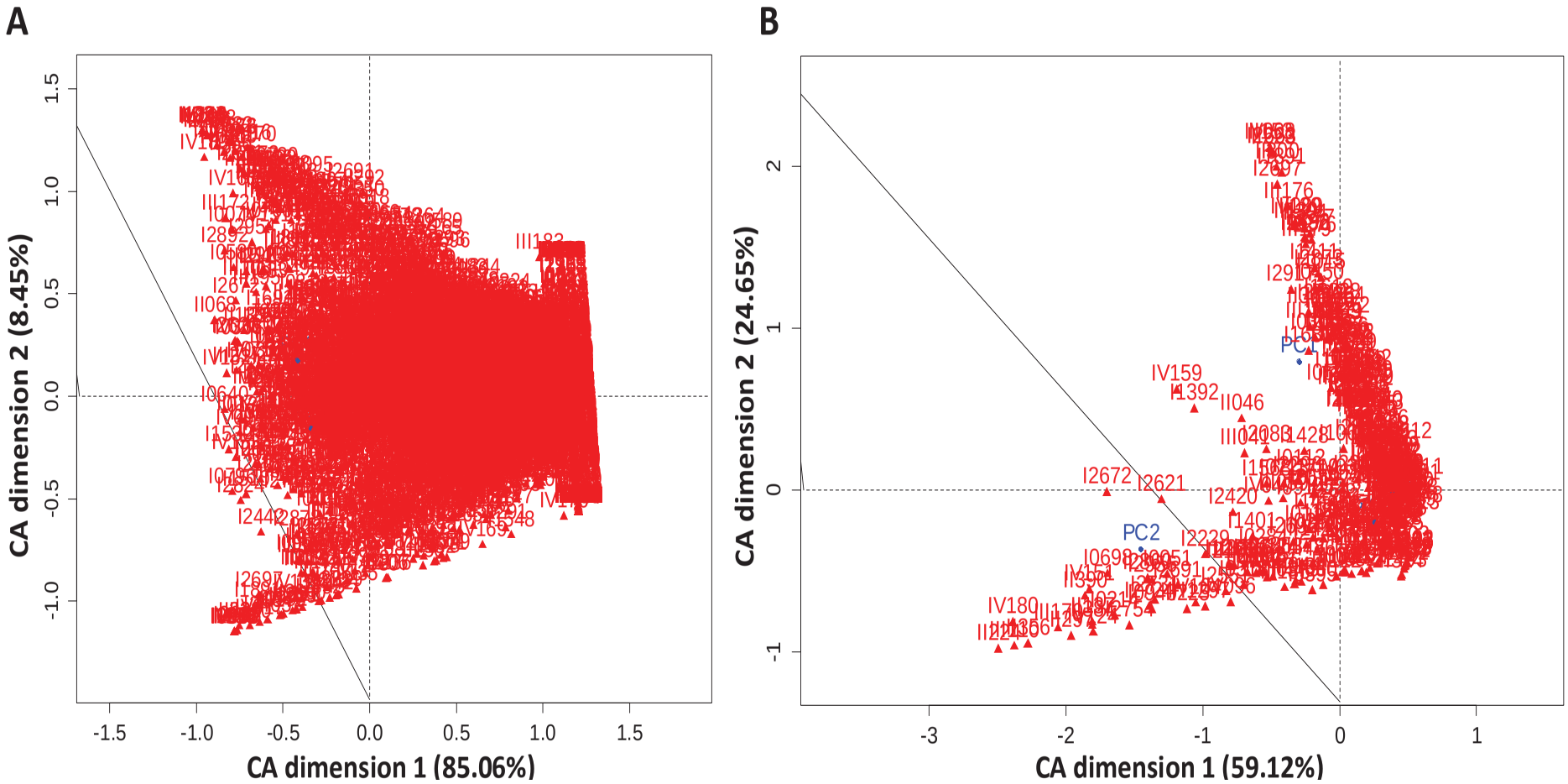


Figure A3.5. CA plots showing transcriptional homogeneity in *B. proteoclasticus* B316 at different treatment groups. A, CA analysis of entire B316 dataset; B, CA analysis of B316 subset dataset consisting of transcripts post *Q-value* cut-off of < 0.05. Two-dimensional CA plots showing genome-wide transcriptional homogeneity between mono- and co-culture growth on xylan and pectin conditions. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. Dashed lines represent the intersection between the first dimension (Dim 1) on the x axis and the second dimension (Dim 2) on the y axis.

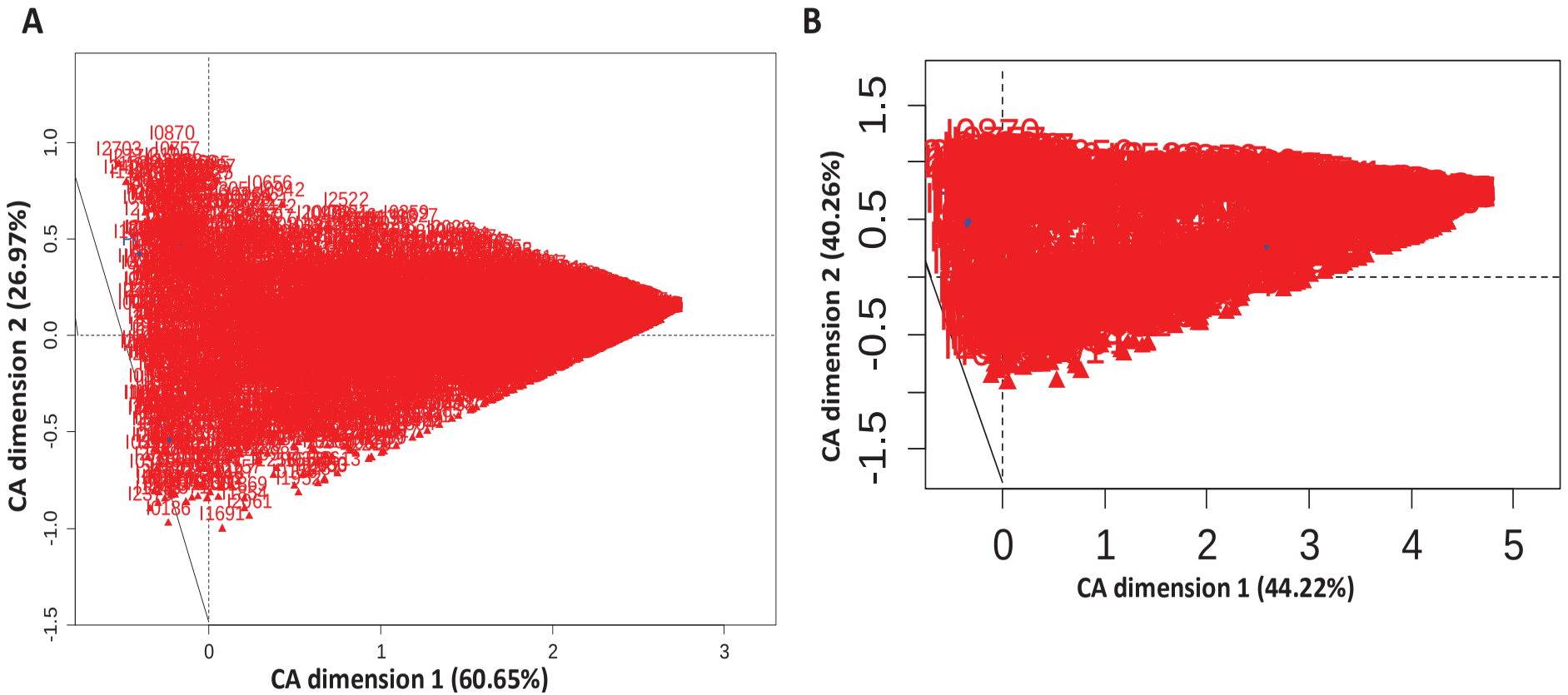


Figure A3.6. CA plots showing transcriptional homogeneity in *B. hungatei* MB2003 at different treatment groups. A, CA analysis of entire MB2003 dataset; B, CA analysis of MB2003 subset dataset consisting of transcripts post *Q-value* cut-off of < 0.05. Two-dimensional CA plots showing genome-wide transcriptional homogeneity between mono- and co-culture growth on xylan and pectin conditions. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. Dashed lines represent the intersection between the first dimension (Dim 1) on the x axis and the second dimension (Dim 2) on the y axis.

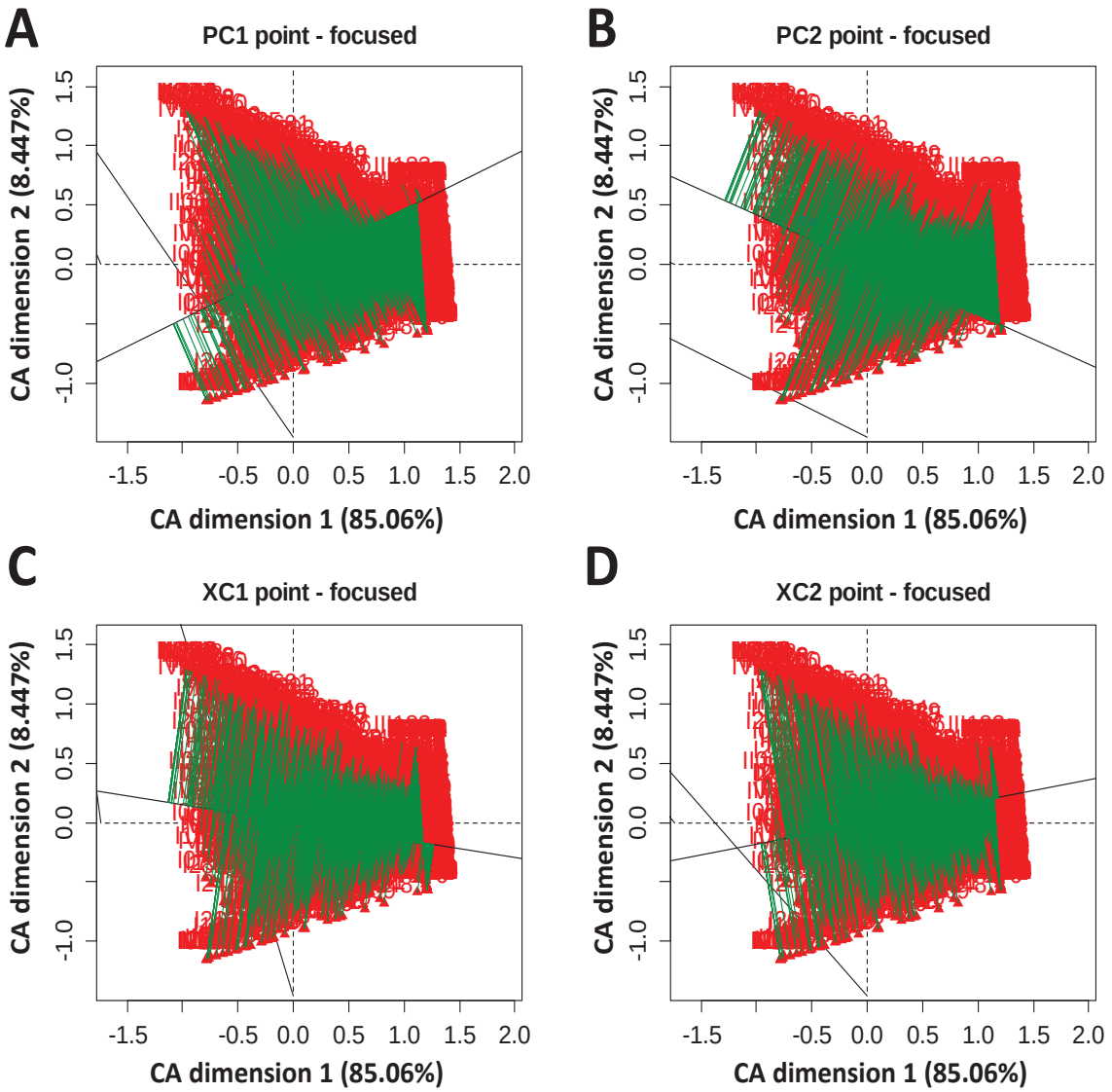


Figure A3.7. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. proteoclasticus* B316. Two-dimensional CA plots showing genome-wide transcriptional homogeneity between mono- and co-culture growth on xylan and pectin conditions using the full B316 transcript dataset. A, CA analysis based on transcripts of B316 grown on pectin in mono-culture. B, CA analysis based on transcripts of B316 grown on pectin in co-culture. C, CA analysis based on transcripts of B316 grown on xylan in mono-culture. D, CA analysis based on transcripts of B316 grown on xylan in co-culture. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. Dashed lines represent the intersection between the first dimension (Dim 1) on the x axis and the second dimension (Dim 2) on the y axis. The bisecting black line in each plot is the reference/centroid line of the dataset.

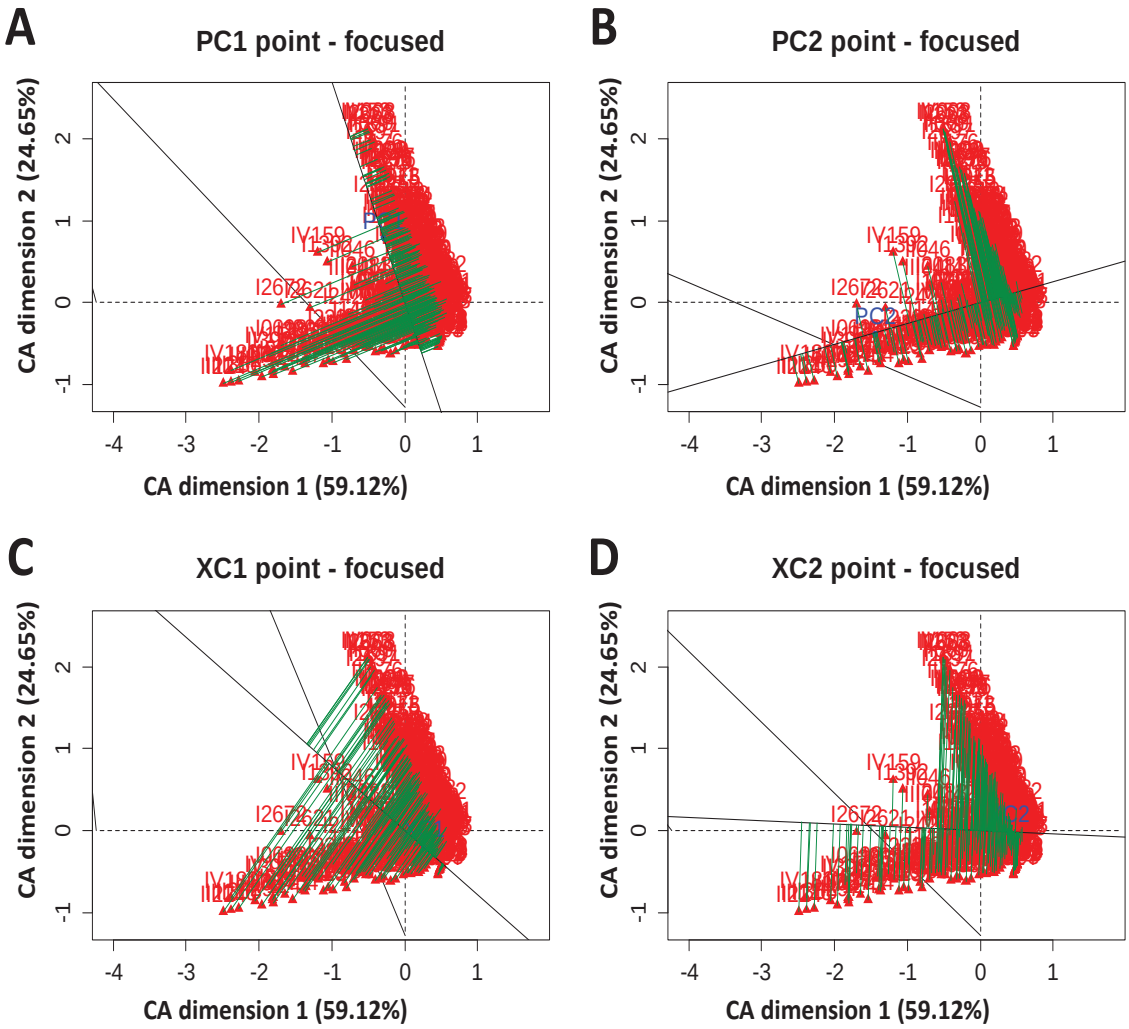


Figure A3.8. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. proteoclasticus* B316 at $Q < 0.05$. Two-dimensional CA plots showing genome-wide transcriptional homogeneity between mono- and co-culture growth on xylan and pectin conditions using the B316 subset dataset consisting of transcripts post *Q-value* cut-off of < 0.05. A, CA analysis based on transcripts of B316 grown on pectin in mono-culture. B, CA analysis based on transcripts of B316 grown on pectin in co-culture. C, CA analysis based on transcripts of B316 grown on xylan in mono-culture. D, CA analysis based on transcripts of B316 grown on xylan in co-culture. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. Dashed lines represent the intersection between the first dimension (Dim 1) on the x axis and the second dimension (Dim 2) on the y axis. The bisecting black line in each plot is the reference/centroid line of the dataset.

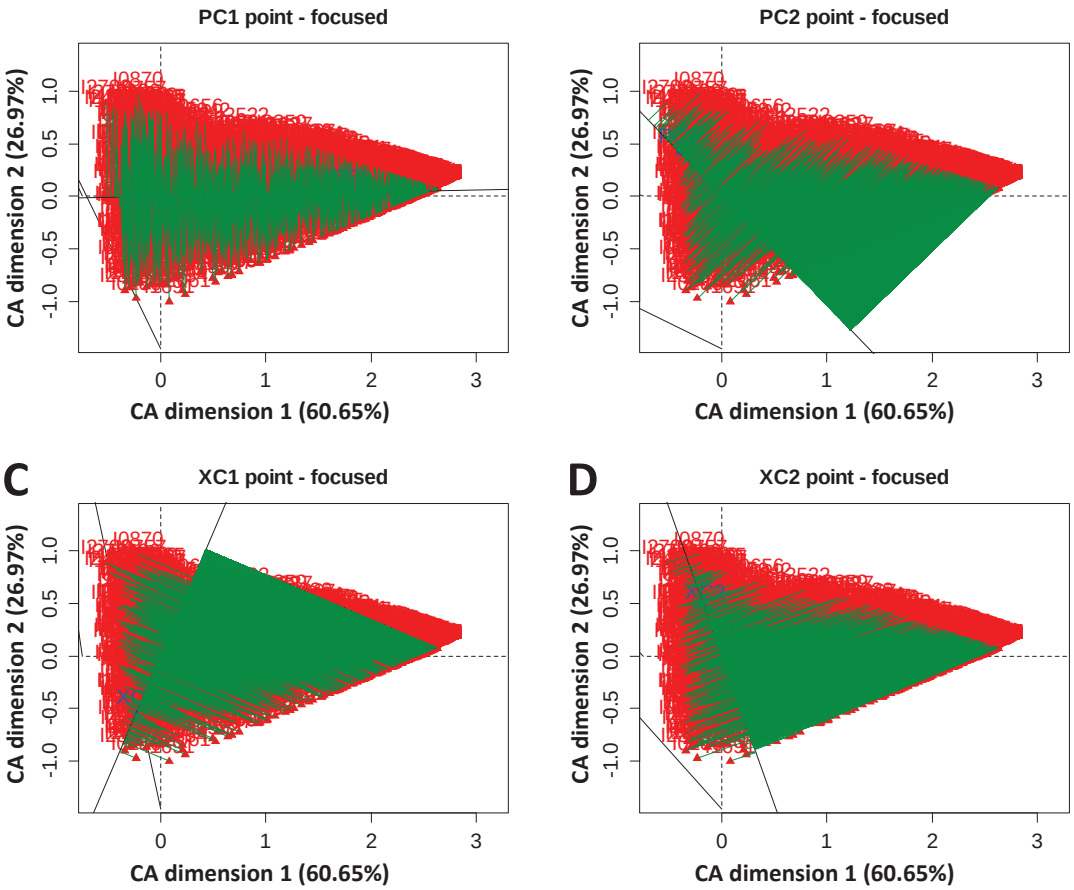


Figure A3.9. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. hungatei* MB2003. Two-dimensional CA plots showing genome-wide transcriptional homogeneity between mono- and co-culture growth on xylan and pectin conditions using the full MB2003 dataset. A, CA analysis based on transcripts of MB2003 grown on pectin in mono-culture. B, CA analysis based on transcripts of MB2003 grown on pectin in co-culture. C, CA analysis based on transcripts of MB2003 grown on xylan in mono-culture. D, CA analysis based on transcripts of MB2003 grown on xylan in co-culture. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. Dashed lines represent the intersection between the first dimension (Dim 1) on the x axis and the second dimension (Dim 2) on the y axis. The bisecting black line in each plot is the reference/centroid line of the dataset.

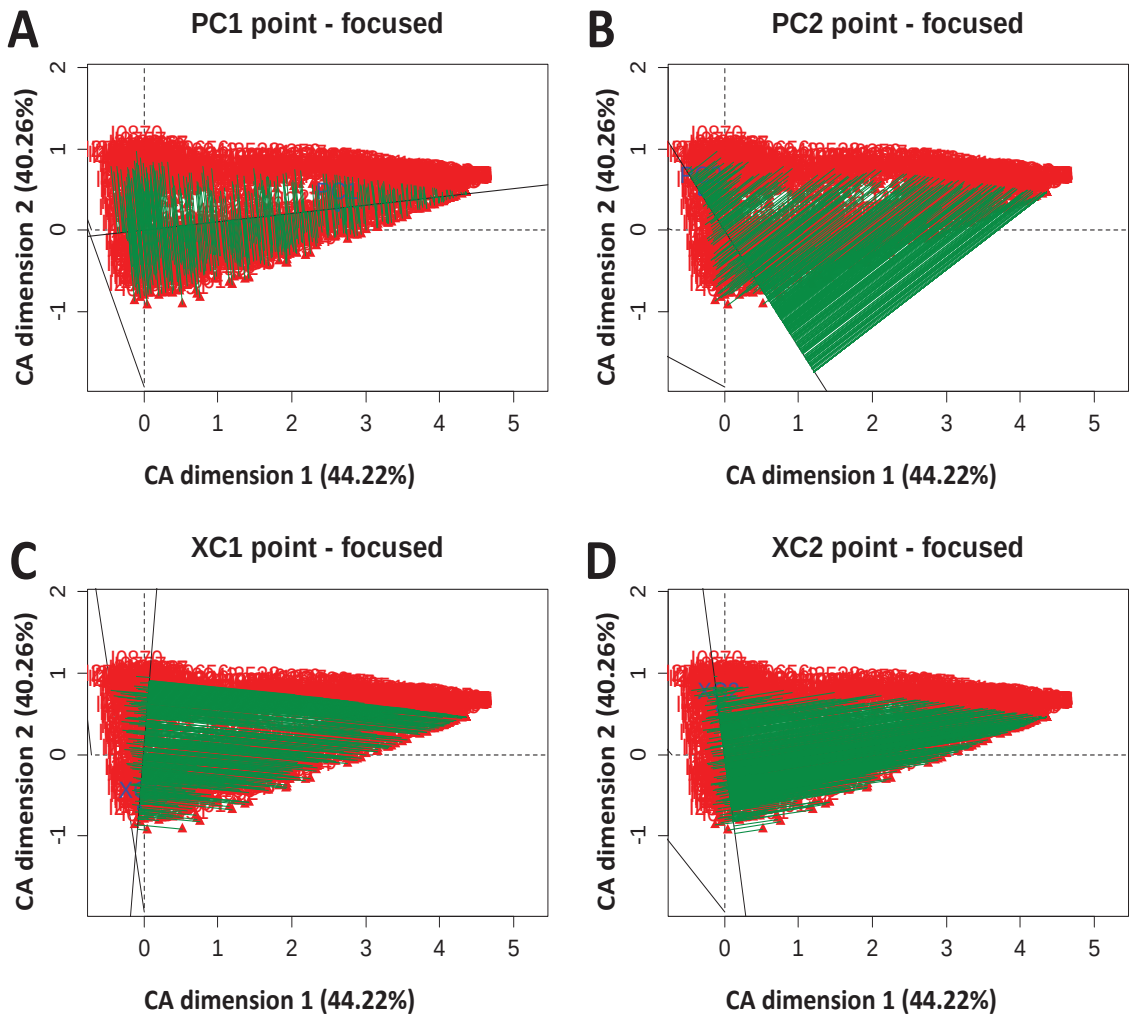


Figure A3.10. CA bi-plots showing transcriptional homogeneity at different treatment groups for *B. hungatei* MB2003 at $Q < 0.05$. Two-dimensional CA plots showing genome-wide transcriptional homogeneity between mono- and co-culture growth on xylan and pectin conditions using the MB2003 subset dataset consisting of transcripts post *Q-value* cut-off of < 0.05. A, CA analysis based on transcripts of MB2003 grown on pectin in mono-culture. B, CA analysis based on transcripts of MB2003 grown on pectin in co-culture. C, CA analysis based on transcripts of MB2003 grown on xylan in mono-culture. D, CA analysis based on transcripts of MB2003 grown on xylan in co-culture. Abbreviations: X, xylan; P, pectin; C1, mono-culture conditions; C2, co-culture condition. Dashed lines represent the intersection between the first dimension (Dim 1) on the x axis and the second dimension (Dim 2) on the y axis. The bisecting black line in each plot is the reference/centroid line of the dataset.

Table A3.14. List of *B. proteoclasticus* B316 subset genes from the network analysis.

Locus tag	Gene annotation	Name	Treatment correlation	
bpr_I1272	acetyl-CoA carboxylase	*accD*	↑PC1	
bpr_I2917	anaerobic ribonucleoside triphosphate reductase	*nrdG1*	↑PC1	
bpr_I1690	AraC family TR	-	↑PC1	
bpr_III162	exodeoxyribonuclease III	*xth*	↑PC1	
bpr_I1377	flagellar biosynthetic protein	*fliQ*	↑PC1	
bpr_I0573	glycosyl transferase 4	-	↑PC1	
bpr_I2228	GNAT acetyltransferase	-	↑PC1	
bpr_I2072	GNAT acetyltransferase	-	↑PC1	
bpr_I0282	HAD superfamily hydrolase	-	↑PC1	
bpr_III179	HxlR family TR	-	↑PC1	
bpr_I1411	hydrogenase chaperone	*hypC*	↑PC1	
bpr_I0206	hypothetical protein	-	↑PC1	
bpr_I1696	hypothetical protein	-	↑PC1	
bpr_I1064	hypothetical protein	-	↑PC1	
bpr_I0561	hypothetical protein	-	↑PC1	
bpr_I0659	hypothetical protein	-	↑PC1	
bpr_I2275	hypothetical protein	-	↑PC1	
bpr_I1178	hypothetical protein	-	↑PC1	
bpr_I2929	hypothetical protein	-	↑PC1	
bpr_I2236	hypothetical protein	-	↑PC1	
bpr_I1966	hypothetical protein	-	↑PC1	
bpr_I2132	hypothetical protein	-	↑PC1	
bpr_I2176	hypothetical protein	-	↑PC1	
bpr_I0450	hypothetical protein	-	↑PC1	
bpr_I2276	hypothetical protein	-	↑PC1	
bpr_I1619	hypothetical protein	-	↑PC1	
bpr_II145	hypothetical protein	-	↑PC1	
bpr_II309	hypothetical protein	-	↑PC1	
bpr_II304	hypothetical protein	-	↑PC1	
bpr_II106	hypothetical protein	-	↑PC1	
bpr_II180	hypothetical protein	-	↑PC1	
bpr_II166	hypothetical protein	-	↑PC1	
bpr_II028	hypothetical protein	-	↑PC1	
bpr_IV064	hypothetical protein	-	↑PC1	
bpr_I0746	iron ABC transporter SBP	-	↑PC1	
bpr_I2471	N-acetylmuramoyl-L-alanine amidase	-	↑PC1	
bpr_I2873	NUDIX hydrolase	-	↑PC1	
bpr_III024	oligopeptide ABC TP	*oppB2*	↑PC1	
bpr_I0863	orotate phosphoribosyltransferase	*pyrE1*	↑PC1	
bpr_I2749	peptide ABC TP	-	↑PC1	
bpr_III049	peptide methionine sulfoxide reductase	*msrA*	↑PC1	
bpr_I1682	phosphotyrosine protein phosphatase	-	↑PC1	
bpr_I1015	radical SAM DCP	-	↑PC1	
bpr_II203	resolvase family protein	-	↑PC1	
bpr_I2677	rhodanese DCP	-	↑PC1	
bpr_I0306	sugar ABC TP	-	↑PC1	
bpr_I2867	TCShistidine kinase	-	↑PC1	
bpr_I1245	tyrosine recombinase	*xerD*	↑PC1	
bpr_II009	hypothetical protein	-	↑PC1	↓XC1
bpr_II153	hypothetical protein	-	↑↑PC1	↓XC1

Locus tag	Gene annotation	Name	Treatment correlation			
bpr_I1891	RNA polymerase alpha subunit C terminal DCP	-	↓XC1	↓XC2		
bpr_I2878	GNAT acetyltransferase	-	↑PC1	↓XC1	↓XC2	
bpr_I2697	HNH endonuclease DCP	-	↑PC1	↓XC1	↓XC2	
bpr_II139	hypothetical protein	-	↑PC1	↓XC1	↓XC2	
bpr_IV139	hypothetical protein	-	↑PC1	↓XC1	↓XC2	
bpr_III176	MarR family TR	-	↑PC1	↓XC1	↓XC2	
bpr_I2663	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2	
bpr_II300	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2	
bpr_II223	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2	
bpr_IV008	hypothetical protein	-	↑↑PC1	↓XC1	↓XC2	
bpr_I1695	hypothetical protein	-	↑PC1	↑PC2	↓XC1	↓XC2
bpr_I1392	hypothetical protein	-	↑PC1	↑PC2	↓XC1	↓XC2
bpr_IV159	YolD-like protein	-	↑PC1	↑PC2	↓XC1	↓XC2
bpr_II224	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_II356	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_II297	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_II390	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_II384	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_IV180	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_IV124	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_IV151	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_II137	hypothetical protein	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_III170	HTH DCP	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_III110	response regulator DCP	-	↑↑↑PC2	↓XC1	↓XC2	
bpr_I2621	cysteine desulfurase	-	↑↑PC2	↓XC1	↓XC2	
bpr_I2672	GNAT acetyltransferase	-	↑↑PC2	↓XC1	↓XC2	
bpr_I0698	sugar ABC TP	-	↑↑PC2	↓XC1	↓XC2	
bpr_I2817	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I2691	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I2860	glycoside hydrolase 31	-	↑↑PC2	↓XC2		
bpr_I2682	sugar ABC TP	-	↑↑PC2	↓XC2		
bpr_I1673	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I2956	tRNA modification GTPase	*trmE*	↑↑PC2	↓XC2		
bpr_I2731	GNAT acetyltransferase	-	↑↑PC2	↓XC2		
bpr_I2724	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I2754	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I0214	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I0940	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_I0051	FeS assembly ATPase	*sufC*	↑↑PC2	↓XC2		
bpr_II225	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_II141	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_II257	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_IV164	hypothetical protein	-	↑↑PC2	↓XC2		
bpr_III240	sugar ABC transporter SBP	-	↑↑PC2	↓XC2		
bpr_III242	sugar ABC TP	-	↑↑PC2	↓XC2		
bpr_I2229	HAD superfamily hydrolase	-	↑PC2	↓XC2		
bpr_III238	HTH DCP	-	↑PC2	↓XC2		
bpr_II046	hypothetical protein	-	↑PC2	↓XC2		
bpr_IV160	hypothetical protein	-	↑PC2	↓XC2		
bpr_I1297	SsrA-binding protein	*smpB*	↑↑PC2			
bpr_III041	AAA family ATPase	-	↑PC2			
bpr_I1264	acyl carrier protein	-	↑PC2			
bpr_I1568	AraC family TR	-	↑PC2			
bpr_I0184	β-xylosidase	-	↑PC2			

bpr_I1401	cation-transporting P-type ATPase	-	↑PC2
bpr_I1292	CDF family transporter	-	↑PC2
bpr_I1261	fibronectin-binding A DCP	-	↑PC2
bpr_I2551	GtrA-like protein	-	↑PC2
bpr_I1541	hypothetical protein	-	↑PC2
bpr_I2420	hypothetical protein	-	↑PC2
bpr_I0864	hypothetical protein	-	↑PC2
bpr_I0681	hypothetical protein	-	↑PC2
bpr_I2083	hypothetical protein	-	↑PC2
bpr_I2785	hypothetical protein	-	↑PC2
bpr_I2086	hypothetical protein	-	↑PC2
bpr_I0379	hypothetical protein	-	↑PC2
bpr_I0284	hypothetical protein	-	↑PC2
bpr_I2134	hypothetical protein	-	↑PC2
bpr_II124	hypothetical protein	-	↑PC2
bpr_II258	hypothetical protein	-	↑PC2
bpr_II008	hypothetical protein	-	↑PC2
bpr_II112	hypothetical protein	-	↑PC2
bpr_II409	hypothetical protein	-	↑PC2
bpr_III059	hypothetical protein	-	↑PC2
bpr_III185	hypothetical protein	-	↑PC2
bpr_II378	IS200/IS605 family transposase	-	↑PC2
bpr_I0232	lacto-N-biose phosphorylase	-	↑PC2
bpr_I1263	MerR family TR	-	↑PC2
bpr_I2696	polypeptide deformylase	-	↑PC2
bpr_I0513	radical SAM DCP	-	↑PC2
bpr_I2171	radical SAM DCP	-	↑PC2
bpr_III140	radical SAM DCP	-	↑PC2
bpr_IV096	restriction endonuclease family protein	-	↑PC2
bpr_II260	signal peptidase I	*traF*	↑PC2
bpr_IV047	single stranded DNA-binding protein	*ssb4*	↑PC2
bpr_I0022	UmuC-like DNA repair protein	-	↑PC2
bpr_I1562	16S rRNA-processing protein	*rimM*	↓PC2
bpr_I1294	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	-	↓PC2
bpr_I2019	30S ribosomal protein	*s20*	↓PC2
bpr_I1222	30S ribosomal protein	*s21*	↓PC2
bpr_I0611	30S ribosomal protein	*s5*	↓PC2
bpr_I2299	50S ribosomal protein	*l1*	↓PC2
bpr_I0610	50S ribosomal protein	*l18*	↓PC2
bpr_I0596	50S ribosomal protein	*l23*	↓PC2
bpr_I0605	50S ribosomal protein	*l24*	↓PC2
bpr_I0612	50S ribosomal protein	*l30*	↓PC2
bpr_I1823	50S ribosomal protein	*l31*	↓PC2
bpr_I2303	50S ribosomal protein	*l33*	↓PC2
bpr_I0784	50S ribosomal protein	*l35*	↓PC2
bpr_I0595	50S ribosomal protein	*l4*	↓PC2
bpr_I2370	50S ribosomal protein	*l7/l12*	↓PC2
bpr_I0852	ABC transporter ABP	-	↓PC2
bpr_I2192	acyl carrier protein	-	↓PC2
bpr_III223	amino acid ABC transporter ABP	-	↓PC2

Locus tag	Gene annotation	Name	Treatment correlation
bpr_III224	amino acid ABC TP	-	↓PC2
bpr_I2944	anti-sigma factor antagonist	-	↓PC2
bpr_I2838	AsnC family TR	-	↓PC2
bpr_I0407	asparagine synthase	-	↓PC2
bpr_I2542	bacterial sugar transferase	-	↓PC2
bpr_I1270	β-hydroxyacyl-ACP dehydratase	-	↓PC2
bpr_III237	β-mannosidase Man2A	-	↓PC2
bpr_I0144	β-phosphoglucomutase	-	↓PC2
bpr_I1586	carbohydrate esterase 12	*est12C*	↓PC2
bpr_I2451	cold-shock protein	*d1*	↓PC2
bpr_I2483	electron transfer flavoprotein	*etfA*	↓PC2
bpr_I0363	flavin reductase DCP	-	↓PC2
bpr_I0421	GFO/IDH/MOCA family oxidoreductase	-	↓PC2
bpr_I0100	glucokinase	*glk*	↓PC2
bpr_I0556	glycosyl transferase 2	-	↓PC2
bpr_I0409	glycosyl transferase 2	-	↓PC2
bpr_I2183	glycosyl transferase 2	-	↓PC2
bpr_I0390	glycosyl transferase 2	-	↓PC2
bpr_I1142	HIT DCP	-	↓PC2
bpr_I1819	Hpt DCP	-	↓PC2
bpr_I0436	hypothetical protein	-	↓PC2
bpr_I2293	hypothetical protein	-	↓PC2
bpr_I2593	hypothetical protein	-	↓PC2
bpr_I0922	hypothetical protein	-	↓PC2
bpr_I2309	hypothetical protein	-	↓PC2
bpr_I0635	hypothetical protein	-	↓PC2
bpr_I0730	hypothetical protein	-	↓PC2
bpr_I0103	hypothetical protein	-	↓PC2
bpr_I1998	hypothetical protein	-	↓PC2
bpr_I1419	hypothetical protein	-	↓PC2
bpr_I2581	hypothetical protein	-	↓PC2
bpr_I2613	hypothetical protein	-	↓PC2
bpr_I2023	hypothetical protein	-	↓PC2
bpr_I1057	hypothetical protein	-	↓PC2
bpr_I1001	hypothetical protein	-	↓PC2
bpr_I2005	hypothetical protein	-	↓PC2
bpr_I0781	hypothetical protein	-	↓PC2
bpr_I0723	hypothetical protein	-	↓PC2
bpr_I0203	hypothetical protein	-	↓PC2
bpr_I2438	hypothetical protein	-	↓PC2
bpr_I1741	hypothetical protein	-	↓PC2
bpr_I2589	hypothetical protein	-	↓PC2
bpr_I0343	hypothetical protein	-	↓PC2
bpr_I0372	hypothetical protein	-	↓PC2
bpr_I1310	hypothetical protein	-	↓PC2
bpr_I2201	hypothetical protein	-	↓PC2
bpr_I2178	hypothetical protein	-	↓PC2
bpr_I2432	hypothetical protein	-	↓PC2
bpr_III044	hypothetical protein	-	↓PC2
bpr_III213	hypothetical protein	-	↓PC2
bpr_II126	hypothetical protein	-	↓PC2
bpr_II026	hypothetical protein	-	↓PC2
bpr_II172	hypothetical protein	-	↓PC2
bpr_II054	hypothetical protein	-	↓PC2

Locus tag	Gene annotation	Name	Treatment correlation
bpr_II274	hypothetical protein	-	↓PC2
bpr_II088	hypothetical protein	-	↓PC2
bpr_II043	hypothetical protein	-	↓PC2
bpr_II115	hypothetical protein	-	↓PC2
bpr_II045	hypothetical protein	-	↓PC2
bpr_II057	hypothetical protein	-	↓PC2
bpr_IV088	hypothetical protein	-	↓PC2
bpr_IV038	hypothetical protein	-	↓PC2
bpr_IV049	hypothetical protein	-	↓PC2
bpr_I2437	IMP dehydrogenase	-	↓PC2
bpr_I1714	L-serine dehydratase	-	↓PC2
bpr_I2448	metallopeptidase M23	-	↓PC2
bpr_I2603	metallopeptidase M23	-	↓PC2
bpr_I1481	methylglyoxal synthase	*mgsA1*	↓PC2
bpr_I1192	molecular chaperone	*groES*	↓PC2
bpr_III215	Na+/H+ antiporter	-	↓PC2
bpr_I2490	NLPC/P60 DCP	-	↓PC2
bpr_I0739	peptidyl-prolyl cis-trans isomerase cyclophilin-type	-	↓PC2
bpr_I2783	peptidyl-prolyl cis-trans isomerase FKBP-type	-	↓PC2
bpr_I0419	phosphopantetheine-binding protein	-	↓PC2
bpr_I0250	phosphopantetheine-binding protein	-	↓PC2
bpr_I1341	phosphoserine aminotransferase	*serC*	↓PC2
bpr_I1992	polypeptide deformylase	-	↓PC2
bpr_I0951	polysaccharide deacetylase	-	↓PC2
bpr_I2106	PTS system I PEP-phosphotransferase	-	↓PC2
bpr_I0273	ribose 5-phosphate isomerase	*rpiB1*	↓PC2
bpr_III014	SPFH domain/band 7 family protein	-	↓PC2
bpr_I2743	succinylglutamate desuccinylase/aspartoacylase	-	↓PC2
bpr_I0117	sugar ABC transporter SBP	-	↓PC2
bpr_I1616	sugar transporter	-	↓PC2
bpr_I2488	thiolase	*thlA1*	↓PC2
bpr_I2196	transketolase subunit A	*tktA2*	↓PC2
bpr_I0617	translation initiation factor IF1	*infA*	↓PC2
bpr_I0783	translation initiation factor IF3	*infC*	↓PC2
bpr_I2048	triosephosphate isomerase	*tpiA*	↓PC2
bpr_I1806	TCSresponse regulator	-	↓PC2
bpr_I0516	UDP-phosphate glucose phosphotransferase	-	↓PC2
bpr_I0953	uridylate kinase	-	↓PC2
bpr_I1955	xanthine phosphoribosyltransferase	-	↓PC2

The network analysis was based on CA of the *B. proteoclasticus* B316 dataset with *Q-value* cut-off of < 0.05. Increase or decrease in gene expression is represented by ↑ or ↓, respectively. The colour key at the top left in Figure 5.14 indiactes expression levels and the threshold for expression was set at 0.15. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture. Abbreviations: ABP, ATP-binding protein; DCP, domain-containing protein; TR, transcriptional regulator; TCS, two component system; TP, transporter permease.

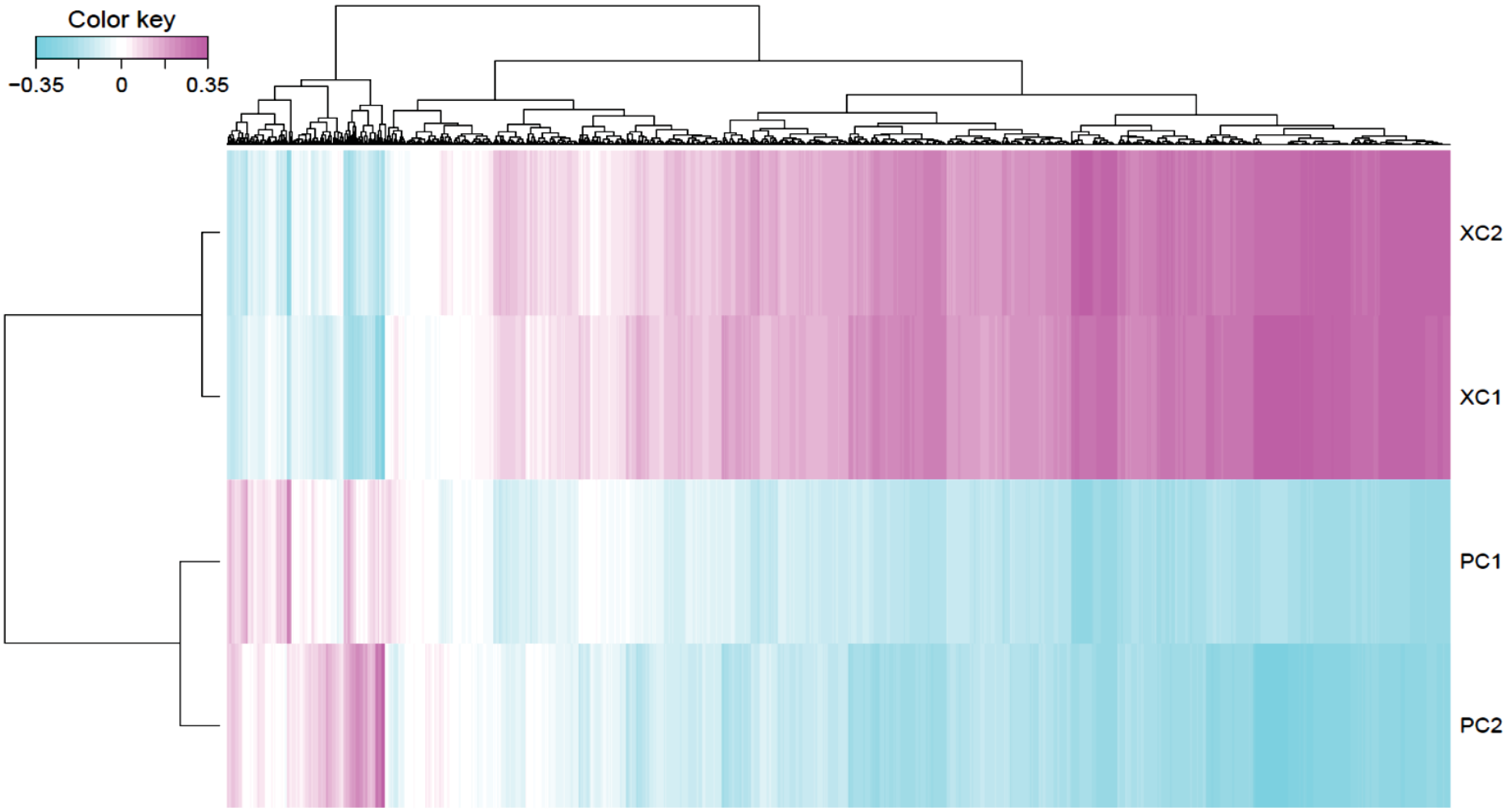


Figure A3.11. CIM analysis of the entire *B. proteoclasticus* B316 dataset. CIM was constructed using the similarity matrix from hierarchical clustering based on CA of the *B. proteoclasticus* B316 dataset. The colour key at the top left indicates expression levels. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture. The gene names are not included due to the large number of transcripts.

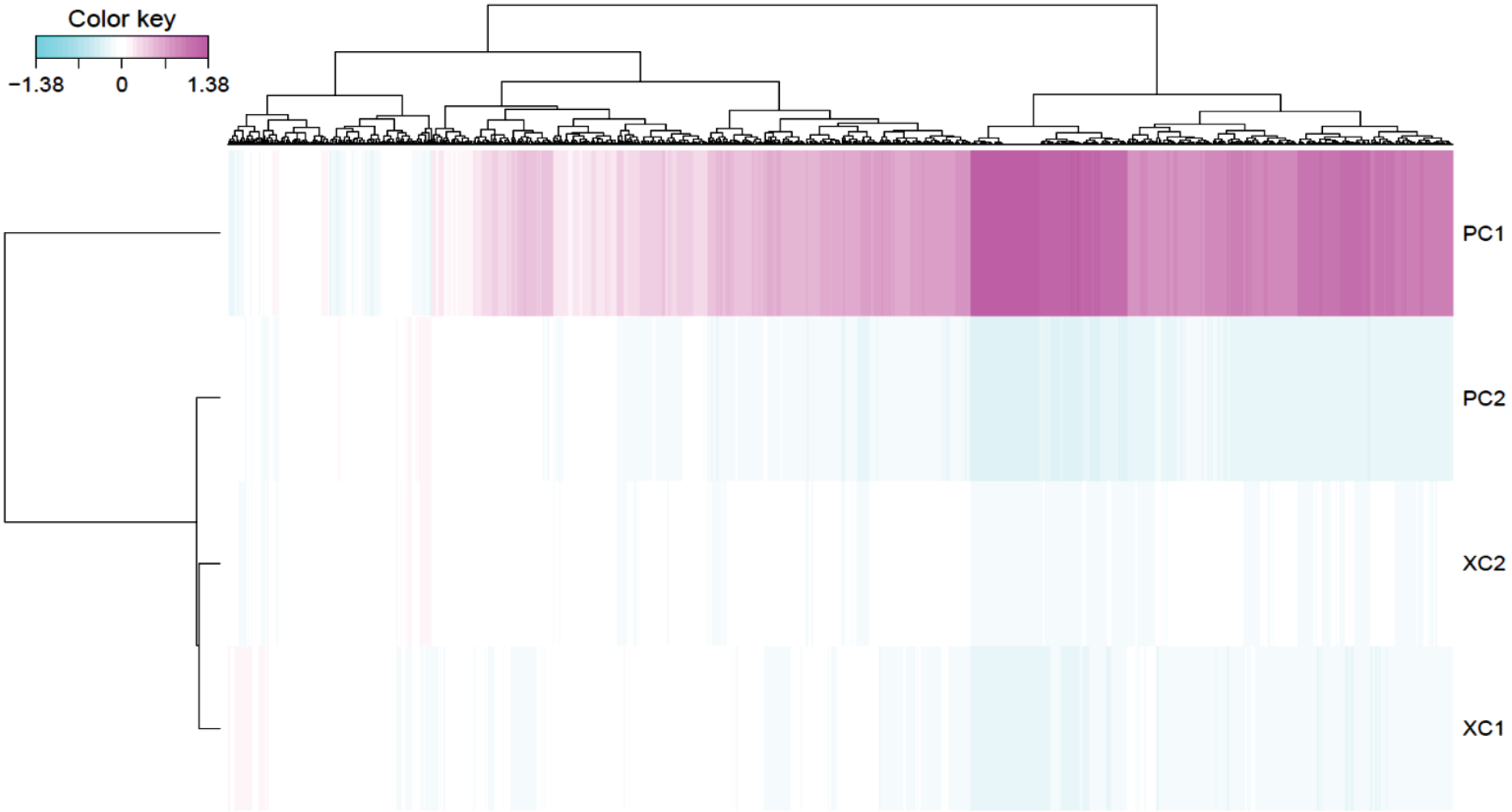


Figure A3.12. CIM analysis of the entire *B. hungatei* MB2003 dataset. CIM was constructed using the similarity matrix from hierarchical clustering based on CA of the *B. hungatei* MB2003 dataset. The colour key at the top left indicates expression levels. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture. The gene names are not included due to the large number of transcripts.

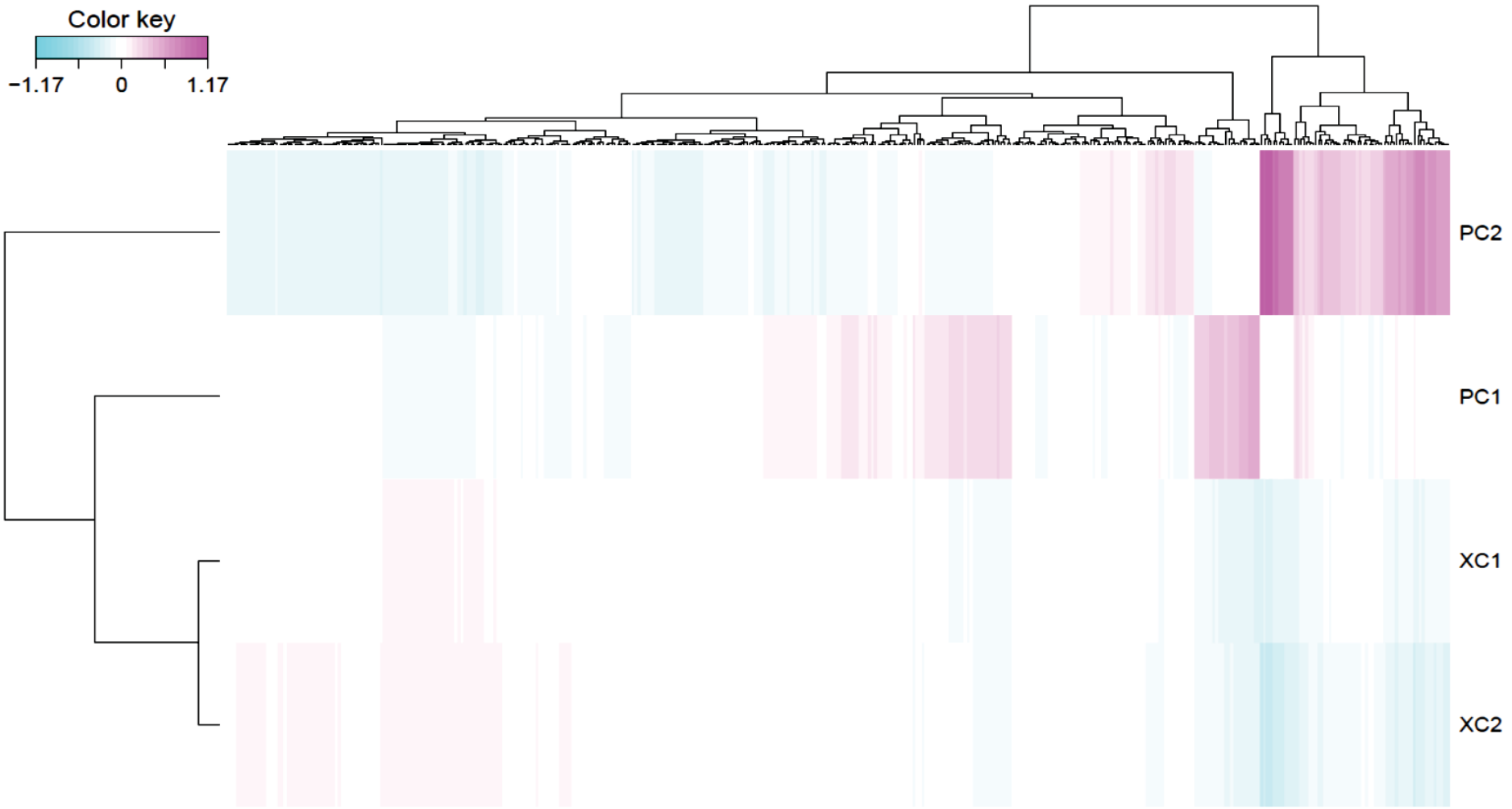


Figure A3.13. CIM analysis of the subset *B. proteoclasticus* B316 dataset. CIM was constructed using the similarity matrix from hierarchical clustering based on CA of the *B. proteoclasticus* B316 dataset post *Q-value* cut-off of < 0.05. The colour key at the top left indicates expression levels. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture. The gene names are not included due to the large number of transcripts.

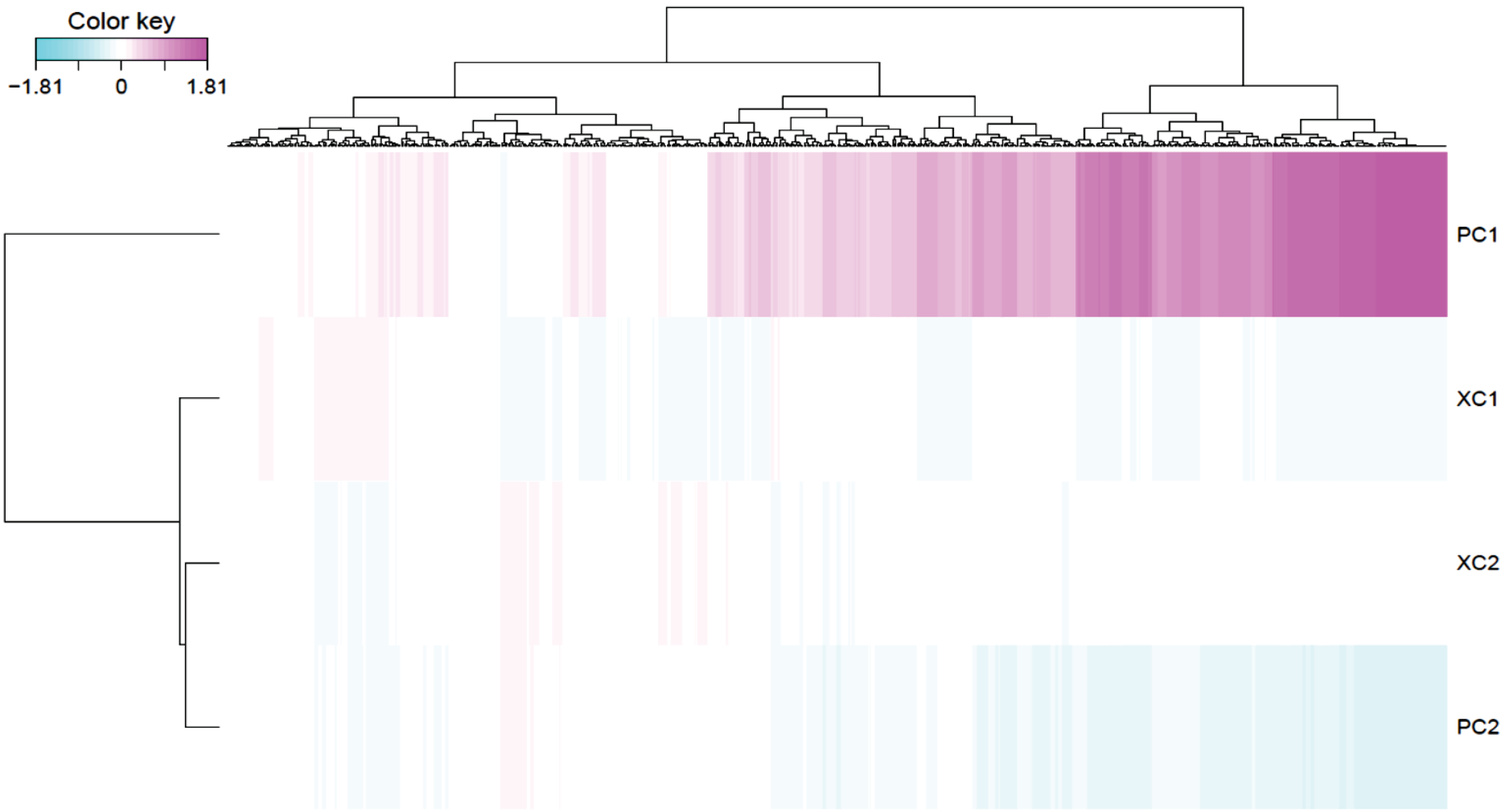


Figure A3.14. CIM analysis of the subset *B. hungatei* MB2003 dataset. CIM was constructed using the similarity matrix from hierarchical clustering based on CA of the *B. hungatei* MB2003 dataset post *Q-value* cut-off of < 0.05. The colour key at the top left indicates expression levels. Abbreviations: X, xylan; P, pectin; C1, mono-culture; C2, co-culture. The gene names are not included due to the large number of transcripts.

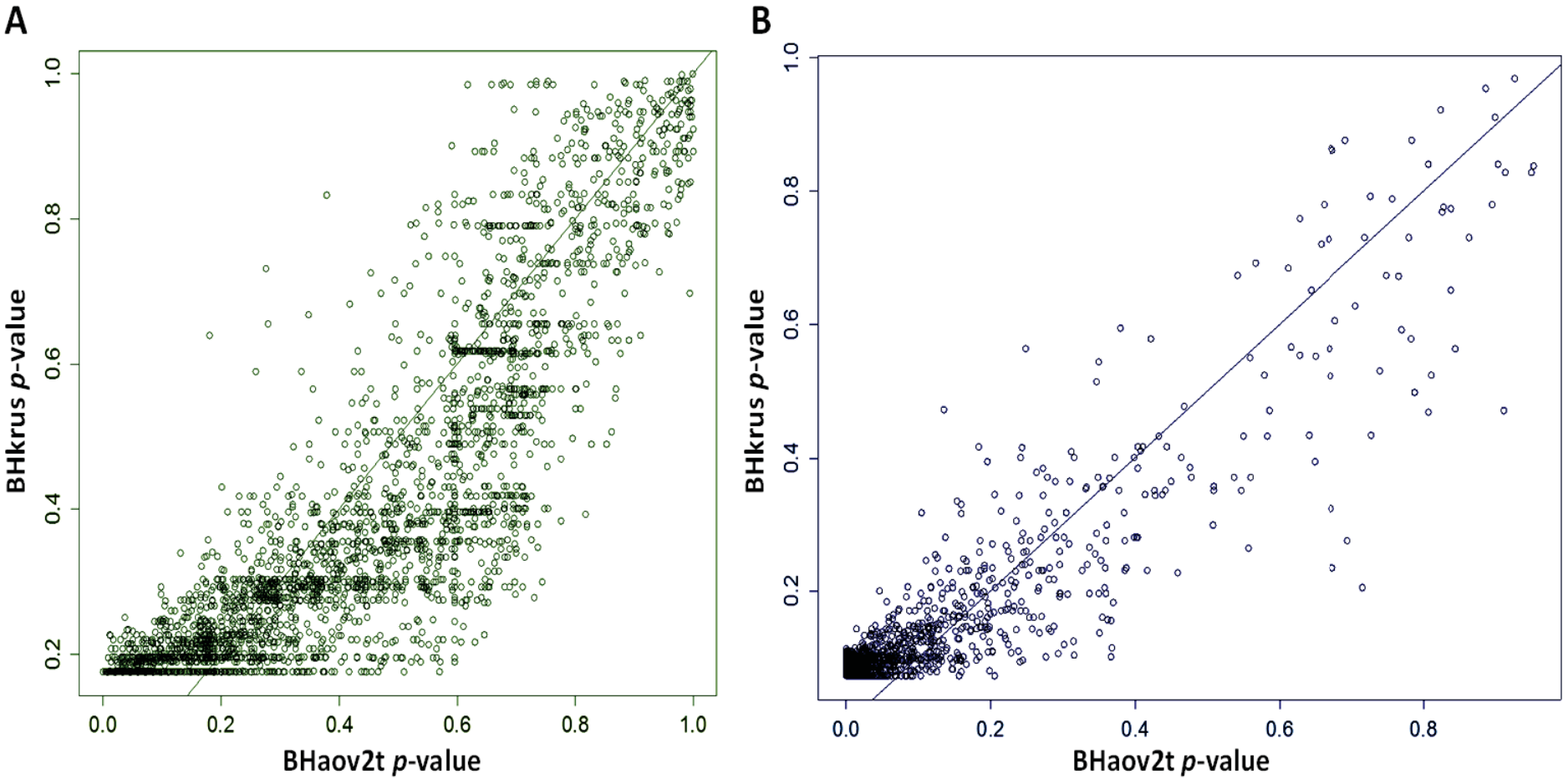


Figure A3.15. Ordinate plots comparing BH-adjusted ANOVA and BH-adjusted KW analyses. A, *B. proteoclasticus* B316 and B, *B. hungatei* MB2003 xylan and pectin complete transcriptome datasets. Abbreviations: BHaov2t *p*-value, Benjamin Hochberg adjusted ANOVA *t*-test *p*-values; BHkrus *p*-value, Benjamin Hochberg adjusted KW analysis of variance by ranks *p*-values.

Table A3.15. List of genes significantly up-regulated during mono- and co-culture growth of *B. proteoclasticus* B316 on xylan.

Condition	Locus tag	Name	Annotation	CAZy	COG	FDR	log_2 FC
B316 mono-culture growth on xylan	bpr_I1403	-	hypothetical protein	-	-	0.00E+00	6.38
	bpr_I0302	*xsa43A*	xylosidase/arabinofuranosidase	GH43/CBM6	[G]	8.57E-29	3.36
	bpr_I2581	-	hypothetical protein	-	-	2.00E-05	2.66
	bpr_I1493	*pepT*	tripeptide aminopeptidase	-	[E]	2.07E-07	2.58
	bpr_I2420	-	hypothetical protein	-	-	6.89E-10	2.32
	bpr_I0082	-	Orn/Lys/Arg decarboxylase	-	[E]	1.91E-08	2.10
	bpr_III122	-	GNAT family acetyltransferase	-	[R]	1.86E-07	2.07
	bpr_I0273	*rpiB1*	ribose 5-phosphate isomerase	-	[G]	2.58E-06	2.06
	bpr_I2005	-	conserved hypothetical protein	-	-	4.31E-02	2.03
	bpr_IV087	-	hypothetical protein	-	-	6.96E-04	2.02
B316 co-culture growth on xylan	bpr_I0612	*rpmD*	ribosomal protein L30	-	[J]	9.44E-109	4.75
	bpr_I2105	-	PTS system HPr phosphocarrier	-	[G]	1.22E-06	3.29
	bpr_I1264	-	acyl carrier protein	-	[R]	8.10E-05	3.17
	bpr_II046	-	hypothetical protein	-	-	4.24E-03	3.12
	bpr_I2083	-	hypothetical protein	-	-	8.42E-03	2.68
	bpr_I2106	*fruB*	PTS system I PEP-phosphotransferase	-	[G]	1.75E-02	2.35
	bpr_I0611	*rpsE*	ribosomal protein S5	-	[J]	2.07E-07	2.25

Genes were deemed "significantly expressed" if all three thresholds were met: FDR calculated using the Benjamini-Hochberg procedure *Q*-values < 0.05, KW analysis of variance with $p < 0.05$, and ≥ 2-fold log_2 difference in expression. CAZy designations have been assigned to genes containing the relevant domains. COG classifications: [C], Energy production and conversion; [D], Cell cycle control, cell division, chromosome partitioning; [E], Amino acid transport and metabolism; [F], Nucleotide transport and metabolism; [G], Carbohydrate transport and metabolism; [H], Coenzyme transport and metabolism; [I], Lipid transport and metabolism; [J], Translation, ribosomal structure and biogenesis; [K], Transcription; [L], Replication, recombination and repair; [M], Cell wall-membrane-envelope biogenesis; [N], Cell motility; [O], Posttranslational modification, protein turnover, chaperones; [P], Inorganic ion transport and metabolism; [Q], Secondary metabolites biosynthesis, transport and catabolism; [R], General function prediction only; [S], Function unknown; [T], Signal transduction mechanisms; [U], Intracellular trafficking, secretion, and vesicular transport; [V], Defence mechanisms.

Table A3.16. List of genes significantly up-regulated during mono- and co-culture growth of *B. hungatei* MB2003 on xylan.

Condition	Locus tag	Name	Annotation	CAZy	COG	FDR	log_2 FC
MB2003 mono-culture growth on xylan	bhn_I2259	-	sugar ABC transporter permease	-	[G]	0.00E+00	6.69
	bhn_III50	*cheY*	chemotaxis protein	-	[T]	3.18E-172	5.75
	bhn_I2653	*cel9A*	endo-1,4-β-glucanase	GH9/CelD	[G]	1.42E-77	5.36
	bhn_I1279	*rpsP*	30S ribosomal protein S16	-	[J]	2.30E-107	5.34
	bhn_I0156	-	F0F1-ATPase subunit	-	-	6.31E-127	5.12
	bhn_I1854	*rpsT*	30S ribosomal protein S20	-	[J]	1.31E-115	5.09
	bhn_I0186	*hpaA*	AraC family transcriptional regulator	-	[K]	2.89E-299	5.04
	bhn_I1243	*cobW1*	cobalamin biosynthesis protein	-	[R]	0.00E+00	4.74
	bhn_I1371	*rplU*	50S ribosomal protein L21	-	[J]	8.79E-56	4.71
	bhn_I0069	-	rhomboid family peptidase S54 family	-	[R]	2.70E-44	4.70
	bhn_I1369	*rpmA*	50S ribosomal protein L27	-	[J]	8.62E-56	4.65
	bhn_I1100	-	hypothetical protein	-	-	1.54E-50	4.61
	bhn_I2102	*rplC*	50S ribosomal protein L3	-	[J]	9.62E-40	4.55
	bhn_I0363	-	hypothetical protein	-	-	7.72E-39	4.50
	bhn_I2099	*rplB*	50S ribosomal protein L2	-	[J]	1.60E-35	4.45
	bhn_II041	-	hypothetical protein	-	-	5.09E-45	4.45
	bhn_I2022	*uxuA*	mannonate dehydratase	-	[G]	4.94E-33	4.44
	bhn_I0140	*asnB*	asparagine synthase	-	[E]	4.78E-104	4.36
	bhn_II078	-	hypothetical protein	-	-	5.83E-34	4.35
	bhn_I2498	*smc*	peptidase M23	-	[M]	5.81E-36	4.33
	bhn_I1658	*deoC*	deoxyribose-phosphate aldolase	-	[F]	2.19E-23	4.23
	bhn_I1780	*xyn8A*	reducing end xylose-releasing exo-oligoxylanase	GH8	[G]	7.38E-98	4.20
	bhn_I2748	*proS*	prolyl-tRNA synthetase	-	[J]	6.21E-50	4.16
	bhn_I2692	*rumA*	RNA methyltransferase	-	[J]	7.33E-13	4.13
	bhn_I2031	-	ABC transporter permease	-	[V]	5.68E-24	4.13
	bhn_I0751	*purM*	phosphoribosylaminoimidazole synthetase	-	[F]	1.05E-17	4.09
	bhn_I2379	-	chloride channel protein	-	[P]	7.05E-83	4.09
	bhn_I1633	-	heme ABC transporter ATP-binding protein	-	[R]	4.02E-23	4.05
	bhn_I0834	-	HD-GYP domain-containing protein	-	[K]	7.27E-18	4.00
	bhn_I0157	-	hypothetical protein	-	-	2.05E-32	3.94
	bhn_II111	-	hypothetical protein	-	-	3.06E-17	3.93
	bhn_I1091	*nadp*	isocitrate dehydrogenase	-	[C]	2.06E-09	3.91

	bhn_I2239	*matE*	multidrug transporter	-	[V]	4.76E-17	3.91
	bhn_I1541	-	hypothetical protein	-	[S]	5.95E-14	3.88
	bhn_I0945	-	TetR family transcriptional regulator	-	[K]	1.81E-12	3.87
	bhn_I0719	*gtrA*	polysaccharide synthesis protein	-	-	3.32E-17	3.83
	bhn_I1370	-	hypothetical protein	-	[J]	6.23E-14	3.83
	bhn_I0223	*lysS*	lysyl-tRNA synthetase	-	[J]	3.33E-13	3.81
	bhn_III79	-	zinc/iron permease/ATP-binding protein	-	[R]	4.97E-04	3.77
	bhn_I1497	*modB*	putrescine aminotransferase	-	[E]	9.63E-12	3.77
	bhn_I0565	*gcp*	O-sialoglycoprotein endopeptidase	-	[O]	2.73E-05	3.73
	bhn_I2663	*metG*	methionyl-tRNA synthetase	-	[J]	1.21E-09	3.70
	bhn_I0217	*argF*	ornithine carbamoyltransferase	-	[E]	1.40E-11	3.68
	bhn_I0874	*matE*	MatE efflux family protein	-	[V]	2.59E-77	3.65
	bhn_I0612	*rplJ*	50S ribosomal protein L10	-	[J]	1.80E-10	3.64
	bhn_I1153	*gldE*	CBS domain-containing protein	-	[R]	4.08E-09	3.61
	bhn_II144	-	hypothetical protein	-	-	2.38E-16	3.61
	bhn_I1262	*cyt*	nucleotidyltransferase	-	[R]	5.56E-25	3.56
	bhn_I1863	*ispH*	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	-	[I]	2.65E-12	3.53
	bhn_I0082	*prmC*	methyltransferase domain-containing protein	-	[R]	1.59E-12	3.51
	bhn_I1649	*pyrB*	aspartate carbamoyltransferase	-	[F]	7.90E-09	3.51
	bhn_I1820	*rlmN*	ribosomal RNA large subunit methyltransferase N	-	[R]	3.03E-05	3.50
	bhn_I2172	*ssuA*	ABC transporter substrate-binding protein	-	[P]	2.15E-16	3.45
	bhn_I1225	-	hypothetical protein	-	-	2.23E-07	3.44
	bhn_I1147	-	CoA-binding domain-containing protein	-	[R]	2.18E-13	3.41
	bhn_I0221	*dusA2*	tRNA dihydrouridine synthase A	-	[J]	1.04E-09	3.38
	bhn_I0768	-	hypothetical protein	-	-	3.50E-56	3.36
	bhn_I1698	*azlC*	branched-chain amino acid ABC transporter permease	-	[E]	1.38E-05	3.35
	bhn_I1289	*rplS*	50S ribosomal protein L19	-	[J]	4.61E-10	3.34
	bhn_I2593	*pfk*	6-phosphofructokinase	-	[G]	7.43E-06	3.32
	bhn_I1256	*ccpA*	LacI family transcriptional regulator	-	[K]	5.04E-07	3.30
	bhn_I0603	-	transcriptional regulator	-	[K]	3.79E-10	3.28
	bhn_I1642	-	HAD superfamily hydrolase	-	[R]	1.82E-39	3.28
	bhn_I1948	-	hypothetical protein	-	-	5.74E-08	3.24
	bhn_I0340	*fabG*	short chain dehydrogenase/reductase	-	[I]	5.59E-56	3.23
	bhn_I0791	-	hypothetical protein	-	[S]	1.18E-04	3.19
	bhn_I0148	*cstA*	carbon starvation protein	-	[T]	5.47E-08	3.19

	bhn_I0962	-	hypothetical protein	-	-	5.19E-05	3.17
	bhn_I2412	-	Zn-dependent protease	-	[O]	9.20E-04	3.17
	bhn_I1734	*ftsZ*	cell division protein	-	[D]	5.23E-31	3.17
	bhn_I1674	*phoB*	PhoP family transcriptional regulator	-	[T]	1.14E-06	3.13
	bhn_I0856	*est4C*	polysaccharide deacetylase	CE4	G]	1.57E-05	3.12
	bhn_I1766	-	AraC family transcriptional regulator	-	[K]	2.37E-14	3.10
	bhn_I0982	-	hypothetical protein	-	-	2.01E-04	3.09
	bhn_I1729	*nikD*	nickel/peptide ABC transporter ATP-binding protein	-	[E]	3.15E-04	3.07
	bhn_I1581	*ilvC*	ketol-acid reductoisomerase	-	[E]	3.26E-09	3.05
	bhn_I2788	-	SpoIIE domain-containing protein	-	[T]	5.12E-24	3.05
	bhn_I1834	-	hypothetical protein	-	-	7.44E-05	3.04
	bhn_I1713	-	hypothetical protein	-	-	1.57E-05	3.04
	bhn_I0353	-	hypothetical protein	-	-	3.97E-30	3.03
	bhn_I0996	-	hypothetical protein	-	-	1.77E-05	3.02
	bhn_I2228	-	hypothetical protein	-	[M]	8.73E-06	3.01
	bhn_I1045	*nikE*	nickel/peptide ABC transporter ATP-binding protein	-	[E]	1.29E-08	3.01
	bhn_II042	-	hypothetical protein	-	-	5.13E-11	2.98
	bhn_I0022	-	hypothetical protein	-	-	3.00E-03	2.96
	bhn_II075	-	hypothetical protein	-	-	3.61E-18	2.92
	bhn_I1878	*gap*	glyceraldehyde-3-phosphate dehydrogenase	-	[G]	2.06E-23	2.92
	bhn_I1496	*modB*	ABC transporter permease	-	[E]	6.83E-05	2.91
	bhn_I2262	*ccpA*	LacI family transcription regulator	-	[K]	9.63E-03	2.91
	bhn_I0470	-	MBOAT family acyltransferase	-	[M]	3.51E-05	2.91
	bhn_I1540	-	hypothetical protein	-	[S]	8.32E-04	2.87
	bhn_I0118	-	hypothetical protein	-	[S]	2.53E-07	2.87
	bhn_I1694	-	ribonuclease III	-	[S]	2.11E-05	2.87
	bhn_I1594	*atpA*	ATP synthase F1 subunit alpha	-	[C]	1.59E-07	2.85
	bhn_II005	-	hypothetical protein	-	-	1.00E-02	2.85
	bhn_I0451	-	LicD family protein	-	[M]	9.22E-04	2.85
	bhn_I0576	-	hypothetical protein	-	[N]	2.70E-02	2.84
	bhn_I0520	*pds*	flavin containing amine oxidoreductase	-	[H]	9.58E-05	2.83
	bhn_I1592	*atpD*	ATP F0F1 synthase subunit beta	-	[C]	2.08E-05	2.82
	bhn_I2140	-	hypothetical protein	-	-	7.04E-03	2.81
	bhn_I0364	-	membrane protein	-	[S]	9.97E-04	2.79
	bhn_I0873	-	hypothetical protein	-	[I]	6.08E-12	2.79

	bhn_II143	-	HTH domain-containing protein	-	[K]	1.86E-21	2.78
	bhn_I1317	-	ribonuclease H	-	[L]	5.74E-03	2.78
	bhn_I0686	-	GntR family transcriptional regulator	-	[K]	6.28E-06	2.78
	bhn_I2092	*rplN*	ribosomal protein L14	-	[J]	4.05E-03	2.78
	bhn_I1348	*mrdB*	rod shape-determining protein RodA	-	[D]	5.09E-05	2.77
	bhn_II115	-	hypothetical protein	-	-	3.92E-03	2.77
	bhn_I1790	-	AAA domain-containing protein	-	[R]	4.94E-03	2.75
	bhn_I0388	-	EAL domain-containing protein	-	[T]	3.62E-03	2.75
	bhn_I1485	-	hypothetical protein	-	[S]	1.53E-03	2.71
	bhn_I1590	-	hypothetical protein	-	-	9.15E-03	2.71
	bhn_I0203	*gh2B*	glycosyl hydrolase family 2	GH2	[G]	4.01E-16	2.70
	bhn_III78	*est1A*	feruloyl esterase	CE1	[G]	1.17E-17	2.69
	bhn_I2515	-	PTS sugar transporter	-	-	2.37E-05	2.69
	bhn_II114	-	hypothetical protein	-	-	4.95E-03	2.69
	bhn_I0528	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	-	[I]	4.59E-02	2.68
	bhn_I2413	*ugl88B*	unsaturated glucuronyl hydrolase	-	[R]	9.30E-03	2.68
	bhn_I1165	*scpA*	chromosome segregation and condensation protein	-	[K]	7.27E-03	2.66
	bhn_I0850	-	zinc/iron permease	-	[P]	1.61E-10	2.66
	bhn_I0298	-	hypothetical protein	-	-	2.10E-02	2.65
	bhn_I1551	-	hypothetical protein	-	[T]	1.78E-05	2.64
	bhn_I0833	-	GGDEF domain-containing protein	-	[T]	1.39E-03	2.64
	bhn_I2631	-	membrane protein	-	[R]	3.14E-08	2.64
	bhn_I1280	-	RNA binding protein	-	[J]	4.56E-04	2.63
	bhn_I1686	-	hypothetical protein	-	-	4.24E-08	2.63
	bhn_I2737	*fba*	fructose-1,6-bisphosphate aldolase	-	[G]	2.20E-18	2.62
	bhn_I1509	-	hypothetical protein	-	[R]	8.61E-03	2.62
	bhn_I0427	-	hypothetical protein	-	[S]	8.49E-04	2.62
	bhn_I2059	*rpsD*	30S ribosomal protein S4	-	[J]	8.74E-03	2.58
	bhn_I2481	*birA*	YbbR-like protein	-	[S]	3.72E-02	2.58
	bhn_I1624	*ihfB*	transcriptional regulator/DNA-binding protein	-	[L]	7.93E-18	2.58
	bhn_I2238	*pbp1A*	penicillin-binding protein 1A family	-	[M]	3.11E-03	2.55
	bhn_I1333	*gltX*	glutamyl-tRNA synthetase	-	[J]	2.26E-03	2.55
	bhn_I2715	*pgl28B*	polygalacturonase	PL3, GH28	[G]	1.03E-02	2.54
	bhn_I1219	-	hypothetical protein	-	-	1.21E-02	2.54
	bhn_I2054	*cbiO*	cobalt ABC transporter ATP-binding protein	-	[P]	3.97E-02	2.53

	bhn_I1018	-	transposase	-	[L]	3.36E-06	2.53
	bhn_I0913	*pcrA*	ATP-dependent DNA helicase	-	[L]	3.46E-03	2.51
	bhn_I0405	-	hypothetical protein	-	-	6.20E-05	2.51
	bhn_III13	-	hypothetical protein	-	-	3.47E-03	2.50
	bhn_I1283	*udp*	uridine phosphorylase	-	[F]	2.08E-03	2.50
	bhn_I2139	*xsa43D*	xylosidase/arabinofuranosidase	GH43	[G]	9.88E-03	2.49
	bhn_I2501	-	CdaR family transcriptional regulator	-	[K]	2.82E-02	2.47
	bhn_I1862	*rpsA*	30S ribosomal protein S1	-	[J]	9.11E-13	2.47
	bhn_I1833	-	radical SAM domain-containing protein	-	[R]	1.26E-02	2.46
	bhn_I1560	*rny*	ribonuclease	-	[L]	3.07E-10	2.45
	bhn_I2765	-	LytTr domain-containing protein	-	[K]	4.02E-02	2.44
	bhn_I0858	*pyrH*	UMP kinase	-	[F]	4.33E-02	2.44
	bhn_I0049	*sufE*	FeS assembly protein	-	[C]	9.80E-03	2.43
	bhn_I2153	*hpaA*	AraC family transcriptional regulator	-	[K]	9.72E-16	2.43
	bhn_I0700	*ygbF*	TPR domain-containing protein	-	[R]	3.66E-04	2.42
	bhn_II093	-	hypothetical protein	-	-	2.21E-03	2.39
	bhn_I1107	-	hypothetical protein	-	-	8.04E-13	2.39
	bhn_I1634	*purC*	phosphoribosylaminoimidazole-succinocarboxamide synthase	-	[F]	7.58E-03	2.37
	bhn_I0453	*galU*	choline kinase	-	[M]	9.47E-03	2.37
	bhn_I2511	-	hypothetical protein	-	[S]	7.76E-04	2.37
	bhn_I2789	-	hypothetical protein	-	[V]	3.17E-12	2.36
	bhn_I1007	*xylF*	sugar ABC transporter substrate-binding protein	-	[G]	1.32E-12	2.35
	bhn_I0634	-	RelA/SpoT domain-containing protein	-	[S]	3.97E-02	2.35
	bhn_I1992	*dnaQ*	exonuclease	-	[L]	3.01E-02	2.34
	bhn_I0300	-	hypothetical protein	-	-	4.90E-02	2.33
	bhn_I0529	-	MBOAT family acyltransferase	-	[M]	4.60E-02	2.32
	bhn_I0583	-	glycosyl transferase GT2 family	GT2	[M]	3.49E-04	2.32
	bhn_I0119	-	phosphocarrier HPr family	-	[G]	1.24E-04	2.31
	bhn_I0683	*purH*	5-aminoimidazole-4-carboxamide ribonucleotide transformylase	-	[F]	2.16E-02	2.30
	bhn_I0809	-	iron ABC transporter permease protein	-	[P]	1.67E-04	2.29
	bhn_I2553	-	hypothetical protein	-	[S]	5.11E-10	2.27
	bhn_I0776	-	hypothetical protein	-	-	1.79E-07	2.25
	bhn_II021	-	type II/IV secretion system protein	-	[U]	1.18E-03	2.24
	bhn_I2759	-	glutathione peroxidase	-	[O]	8.92E-10	2.23
	bhn_III23	*pta*	phosphate acetyltransferase	-	[C]	3.42E-25	2.22

	bhn_I1273	*ilvA*	threonine dehydratase	-	[E]	1.22E-02	2.22
	bhn_I1167	*mmdA*	carboxyl transferase	-	[I]	3.20E-05	2.22
	bhn_I1568	*mrdA*	peptidoglycan glycosyltransferase	-	[M]	2.69E-08	2.21
	bhn_I1687	-	Hpt domain-containing protein	-	[T]	5.46E-05	2.21
	bhn_I1779	-	methyl-accepting chemotaxis protein	-	[N]	1.61E-09	2.20
	bhn_I1293	*rnhB*	ribonuclease HII	-	[L]	3.95E-02	2.20
	bhn_I1151	-	hypothetical protein	-	-	2.05E-05	2.19
	bhn_I0005	*gyrB*	DNA gyrase subunit B	-	[L]	4.61E-04	2.18
	bhn_III76	*lyc25B*	lysozyme	GH25, 2xSH3	[G]	1.73E-18	2.17
	bhn_I2149	-	cNMP binding domain-containing protein	-	[T]	2.88E-09	2.17
	bhn_I0783	-	hypothetical protein	-	-	1.44E-02	2.16
	bhn_I0358	*cotH*	spore coat protein	-	[M]	1.49E-03	2.14
	bhn_I1353	*sepF*	cell division protein	-	[S]	2.71E-02	2.14
	bhn_I1650	*pyrI*	aspartate carbamoyltransferase	-	[F]	6.86E-03	2.12
	bhn_I1667	-	PAP2 family protein	-	[I]	3.14E-02	2.12
	bhn_I0879	*matE*	MatE efflux family protein	-	[V]	2.99E-02	2.11
	bhn_I2256	*thlA1*	acetyl-CoA acetyltransferase	-	[I]	1.42E-03	2.10
	bhn_I1819	-	serine/threonine protein phosphatase	-	[T]	2.16E-02	2.09
	bhn_I0412	*pseB*	UDP-N-acetylglucosamine 4,6-dehydratase	-	[M]	4.73E-02	2.08
	bhn_I0851	-	QueT family transporter	-	[S]	3.17E-09	2.08
	bhn_I0420	-	hypothetical protein	-	-	8.18E-04	2.07
	bhn_III14	-	DNA-binding protein	-	[L]	1.84E-16	2.07
	bhn_I0741	-	ABC transporter ATP-binding protein	-	[P]	4.39E-03	2.06
	bhn_II015	-	hypothetical protein	-	-	9.18E-03	2.04
	bhn_I1662	-	hypothetical protein	-	-	9.26E-07	2.02
	bhn_I0289	-	hypothetical protein	-	-	2.82E-02	2.02
	bhn_I0421	-	hypothetical protein	-	-	9.87E-04	2.01
	bhn_I0861	*pgsA*	CDP-diglyceride synthetase	-	[I]	7.91E-03	2.01
	bhn_I0780	-	CAAX amino terminal protease family protein	-	[R]	1.28E-02	2.00
	bhn_I2550	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II105	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I1168	-	sodium ion-translocating decarboxylase	-	-	0.00E+00	N/A
	bhn_III84	-	AAA domain-containing protein	-	-	1.17E-08	N/A
	bhn_III85	-	hypothetical protein	-	-	1.47E-04	N/A
	bhn_I2160	*thyA*	thymidylate synthase	-	[F]	3.03E-85	N/A

	bhn_I0650	*rpmI*	50S ribosomal protein L35	-	[J]	0.00E+00	N/A
	bhn_II065	-	single-stranded DNA-binding protein	-	[L]	0.00E+00	N/A
	bhn_I1311	-	heavy metal transporter	-	[P]	0.00E+00	N/A
	bhn_III61	-	CorA-like Mg2+ transporter protein	-	[P]	3.87E-02	N/A
	bhn_I0502	-	radical SAM protein	-	[R]	4.37E-272	N/A
	bhn_I2629	-	prolyl-tRNA synthetase	-	[S]	0.00E+00	N/A
	bhn_III77	-	GNAT family acetyltransferase	-	[S]	3.18E-135	N/A
	bhn_I1292	-	signal peptidase I	-	[U]	0.00E+00	N/A
MB2003 co-culture growth on xylan	bhn_I0870	-	two component system histidine kinase	-	[T]	0.00E+00	6.33
	bhn_I2522	*grpE*	co-chaperone	-	[O]	4.06E+106	5.01
	bhn_I0757	*phnF*	GntR family transcriptional regulator	-	[K]	1.55E+147	4.35
	bhn_I0208	-	hypothetical protein	-	-	1.04E+53	4.32
	bhn_I1136	-	PTS ascorbate transporter subunit IIC	-	[S]	1.01E+34	4.29
	bhn_I1765	*ngcE*	sugar ABC transporter substrate-binding protein	-	[G]	7.22E+106	4.09
	bhn_III58	-	hypothetical protein	-	-	6.88E+08	4.03
	bhn_I0656	*ric*	disulfide oxidoreductase	-	[C]	4.57E+39	3.98
	bhn_I2399	*pncB*	nicotinate phosphoribosyltransferase	-	[H]	2.43E+15	3.91
	bhn_I0869	*hsp20*	heat-shock protein	-	[O]	2.25E+112	3.82
	bhn_I0942	*abrB*	AbrB family transcriptional regulator	-	-	7.42E+34	3.81
	bhn_I2402	-	NUDIX family hydrolase	-	[F]	8.38E+15	3.81
	bhn_I2190	-	hypothetical protein	-	-	9.76E+97	3.77
	bhn_I1599	-	hypothetical protein	-	-	7.19E+49	3.76
	bhn_I0681	*agn53A*	arabinogalactan endo-1,4-β-galactosidase	GH53	[G]	1.78E+10	3.70
	bhn_I2358	*fliS*	flagellar biosynthesis protein	-	[N]	1.01E+67	3.69
	bhn_I2555	-	CMP deaminase	-	[F]	3.25E+09	3.64
	bhn_I2646	*pyrD*	dihydroorotate dehydrogenase	-	[F]	1.63E+62	3.55
	bhn_I2046	-	hypothetical protein	-	-	1.11E+17	3.54
	bhn_I0825	-	von Willebrand factor type A domain-containing protein	-	[R]	1.97E+94	3.52
	bhn_I0881	-	hypothetical protein	-	-	6.06E+22	3.51
	bhn_I2368	*flgM*	flagellar biosynthesis anti-sigma factor	-	[N]	8.61E+125	3.51
	bhn_I1135	-	SPFH domain/Band 7 family protein	-	[O]	1.20E+15	3.49
	bhn_I2601	-	RNA polymerase sigma factor sigma-70 family	-	[K]	1.33E+93	3.44
	bhn_I1637	*thrH*	phosphoserine phosphatase	-	[E]	7.10E+52	3.44
	bhn_I0972	*xylF*	xylose ABC transporter substrate-binding protein	-	[G]	2.72E+149	3.32
	bhn_I2047	-	hypothetical protein	-	-	1.12E+06	3.25

	bhn_I2257	*raiA*	ribosomal subunit interface protein	-	[J]	4.32E+62	3.19
	bhn_I2331	-	hypothetical protein	-	-	3.60E+07	3.18
	bhn_I0973	*xylG*	xylose ABC transporter ATP-binding protein	-	[G]	2.74E+11	3.17
	bhn_I1324	*rimI*	GNAT family acetyltransferase	-	[K]	1.10E+19	3.10
	bhn_I2333	-	HAD family hydrolase	-	[R]	1.36E-05	3.09
	bhn_I2048	-	hypothetical protein	-	-	2.36E+09	2.99
	bhn_I2221	*gh31C*	glycoside hydrolase family 31	GH31	[G]	1.73E+40	2.98
	bhn_I0964	-	hypothetical protein	-	-	1.08E+12	2.97
	bhn_I1860	-	hypothetical protein	-	-	1.15E-04	2.95
	bhn_I0237	*eutE*	succinylglutamate desuccinylase	-	[R]	1.61E-04	2.95
	bhn_I2642	-	hypothetical protein	-	-	1.77E+38	2.92
	bhn_I2473	*galE*	GDP-L-fucose synthase	-	[M]	2.57E+69	2.92
	bhn_I0758	*clpB*	ATP-dependent chaperone protein	-	[O]	2.37E+32	2.91
	bhn_I1870	-	metal-dependent hydrolase	-	[R]	7.27E+36	2.89
	bhn_I1216	*hsdS*	type I restriction-modification system S subunit	-	[V]	1.43E-04	2.86
	bhn_I0963	-	LytTR domain-containing protein	-	[K]	6.88E+11	2.84
	bhn_I2225	*ccpA*	LacI family transcriptional regulator	-	[K]	2.84E+28	2.83
	bhn_I0209	-	hypothetical protein	-	-	1.81E-04	2.83
	bhn_I2774	-	GGDEF/PAS domain-containing protein	-	[T]	6.26E+73	2.81
	bhn_I2662	-	α/β hydrolase	-	[I]	1.37E+28	2.78
	bhn_I2354	*fliC*	flagellin	-	[N]	1.66E+68	2.76
	bhn_I2472	*gmd*	GDP-mannose 4,6-dehydratase	-	[M]	7.11E+11	2.74
	bhn_I1351	*pilT*	twitching motility protein	-	-	1.99E+40	2.71
	bhn_I2523	*hrcA*	heat-inducible transcription repressor	-	[K]	1.68E-05	2.70
	bhn_I2776	*modB*	ABC transporter permease	-	[E]	1.73E+60	2.69
	bhn_I1841	*bga35B*	β-galactosidase	GH35	[G]	9.70E+11	2.66
	bhn_I0396	-	molecular chaperone Hsp20	-	[O]	9.01E+10	2.66
	bhn_I2488	-	hypothetical protein	-	-	9.45E-04	2.64
	bhn_I0109	-	ROK family protein	-	[K]	3.22E-05	2.64
	bhn_I2360	*flaG*	flagellar protein	-	[N]	4.53E+14	2.60
	bhn_I1845	-	sugar ABC transporter substrate-binding protein	-	[G]	1.48E+30	2.59
	bhn_I0584	-	thioesterase	-	[I]	7.95E+12	2.56
	bhn_I2224	-	sugar ABC transporter permease protein	-	[G]	2.41E+21	2.55
	bhn_I2474	-	hypothetical protein	-	-	4.80E+08	2.55
	bhn_I0805	-	MecA family protein	-	-	1.70E+24	2.50

	bhn_I2521	*dnaK*	molecular chaperone	-	[O]	4.33E+18	2.49
	bhn_I1598	-	hypothetical protein	-	-	5.34E+25	2.48
	bhn_I2516	*thiI*	thiamine biosynthesis/tRNA modification protein	-	[H]	7.73E+33	2.41
	bhn_I2230	*cspD*	cold-shock protein	-	[K]	9.04E+08	2.41
	bhn_I2305	-	hypothetical protein	-	-	7.91E-03	2.40
	bhn_I1528	-	cation diffusion facilitator family transporter	-	[P]	1.03E-03	2.39
	bhn_I0334	-	hypothetical protein	-	-	2.97E-04	2.38
	bhn_I1097	*rho*	transcription termination factor	-	[K]	1.09E+23	2.36
	bhn_III86	*hicA*	HicA family protein	-	-	5.14E-04	2.28
	bhn_I0153	-	hypothetical protein	-	-	3.88E+18	2.28
	bhn_I1233	-	α/β hydrolase family protein	-	[R]	3.75E+31	2.27
	bhn_I2367	*flgN*	flagellar biosynthesis protein	-	[N]	8.97E+26	2.26
	bhn_I1532	*arf51C*	α-L-arabinofuranosidase	GH51/CBM4	[G]	4.33E+22	2.25
	bhn_I1407	*est4D*	polysaccharide deacetylase	CE4	[G]	5.31E-03	2.23
	bhn_I1241	*cbiO*	cobalt ABC transporter ATP-binding protein	-	[R]	5.15E-05	2.21
	bhn_I1461	*fliK*	flagellar hook length control protein	-	[N]	2.09E+23	2.20
	bhn_I1224	*recF*	ATPase AAA family	-	[R]	2.32E-03	2.17
	bhn_I2560	*mecA*	competence protein	-	[O]	1.96E+15	2.14
	bhn_I0659	-	sugar ABC transporter substrate-binding protein	-	[G]	1.56E+14	2.13
	bhn_I2132	*ngcE*	sugar ABC transporter substrate-binding protein	-	[G]	5.83E+14	2.12
	bhn_I2597	-	hypothetical protein	-	-	2.18E-05	2.12
	bhn_I2777	-	ABC transporter substrate-binding protein	-	[E]	8.13E+10	2.01
	bhn_I1357	*cmk*	cytidylate kinase	-	[F]	5.82E-04	2.01
	bhn_I2503	-	CAAX amino terminal protease family protein	-	[R]	6.00E+18	2.00
	bhn_I0167	*cel5A*	endo-1,4-β-glucanase/xylanase	GH5	[G]	4.19E-03	2.00
	bhn_I0955	*matE*	MatE efflux family protein	-	[V]	1.44E-02	2.00
	bhn_I0629	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I2636	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I1484	-	MFS transporter	-	[H]	0.00E+00	N/A
	bhn_I2141	-	phosphomethylpyrimidine synthase	-	[H]	0.00E+00	N/A

Genes were deemed "significantly expressed" if all three thresholds were met: FDR calculated using the Benjamini-Hochberg procedure *Q*-values < 0.05, KW analysis of variance with $p < 0.05$, and ≥ 2-fold $\log_2$ difference in expression. CAZy designations have been assigned to genes containing the relevant domains. COG classifications: [C], Energy production and conversion; [D], Cell cycle control, cell division, chromosome partitioning; [E], Amino acid transport and metabolism; [F], Nucleotide transport and metabolism; [G], Carbohydrate transport and metabolism; [H], Coenzyme transport and metabolism; [I], Lipid transport and metabolism; [J], Translation, ribosomal structure and biogenesis; [K], Transcription; [L], Replication, recombination and repair; [M], Cell wall-membrane-envelope biogenesis; [N], Cell motility; [O], Posttranslational modification, protein turnover, chaperones; [P], Inorganic ion transport and metabolism; [Q], Secondary metabolites biosynthesis, transport and catabolism; [R], General function prediction only; [S], Function unknown; [T], Signal transduction mechanisms; [U], Intracellular trafficking, secretion, and vesicular transport; [V], Defence mechanisms.

Table A3.17. List of genes significantly up-regulated during mono- and co-culture growth of *B. proteoclasticus* B316 on pectin.

Condition	Locus tag	Name	Annotation	CAZy	COG	FDR	log_2 FC
B316 mono-culture growth on pectin	bpr_I0682	-	hypothetical protein	-	-	3.52E-55	4.60
	bpr_I0848	-	hypothetical protein	-	-	7.47E-48	4.54
	bpr_I1955	*xptA*	xanthine phosphoribosyltransferase	-	[F]	1.02E-40	4.52
	bpr_I0781	-	hypothetical protein	-	-	4.22E-13	4.03
	bpr_II309	-	hypothetical protein	-	-	8.65E-11	3.81
	bpr_II209	-	hypothetical protein	-	-	3.28E-13	3.79
	bpr_II021	-	hypothetical protein	-	-	1.57E-11	3.75
	bpr_I0931	*eda*	2-keto-3-deoxygluconate 6-phosphate aldolase	-	[G]	2.85E-06	3.47
	bpr_III162	*xth*	exodeoxyribonuclease III	-	[L]	1.87E-05	3.47
	bpr_I0306	-	sugar ABC transporter permease	-	[G]	4.98E-06	3.43
	bpr_II396	-	helicase	-	[L]	2.08E-04	3.39
	bpr_II086	-	phage integrase family protein	-	[L]	1.12E-04	3.24
	bpr_I1178	-	hypothetical protein	-	-	2.74E-03	3.12
	bpr_I2844	-	toxic anion resistance family protein	-	[P]	1.26E-04	3.06
	bpr_I1091	-	PilZ domain-containing protein	-	-	5.31E-03	3.04
	bpr_I2136	-	GGDEF/EAL domain-containing protein	-	[T]	5.07E-03	2.94
	bpr_I2190	*dltA*	AMP-binding enzyme	-	[Q]	3.38E-02	2.94
	bpr_I1420	-	PAP2 family protein	-	[I]	2.20E-02	2.86
	bpr_I0446	-	conserved hypothetical transmembrane protein	-	-	1.78E-02	2.81
	bpr_I1682	-	phosphotyrosine protein phosphatase	-	[T]	7.67E-03	2.77
	bpr_I2395	-	hypothetical secreted protein	-	-	1.68E-02	2.68
	bpr_I1748	-	hypothetical protein	-	-	2.62E-02	2.64
	bpr_I2804	-	conserved hypothetical transmembrane protein	-	-	3.59E-02	2.59
	bpr_I0102	-	P-loop-containing ATPase	-	[R]	0.00E+00	N/A
	bpr_I0117	-	sugar ABC transporter substrate-binding protein	-	[G]	0.00E+00	N/A
	bpr_I0363	-	flavin reductase domain-containing protein	-	[R]	0.00E+00	N/A
	bpr_I0390	-	glycosyl transferase 2	GT2	[M]	0.00E+00	N/A
	bpr_I0730	-	hypothetical protein	-	-	0.00E+00	N/A
	bpr_I0953	*pyrH*	uridylate kinase	-	[F]	0.00E+00	N/A
	bpr_I1142	-	HIT domain-containing protein	-	[R]	0.00E+00	N/A
	bpr_I1245	*xerD*	tyrosine recombinase	-	[L]	0.00E+00	N/A
	bpr_I1272	*accD*	acetyl-CoA carboxylase carboxyl transferase beta subunit	-	[I]	0.00E+00	N/A
	bpr_I1294	*gpmA*	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	-	[G]	0.00E+00	N/A
	bpr_I1616	*gph*	sugar transporter GPH family	-	[G]	0.00E+00	N/A

	bpr_I1831	*ispF*	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	-	[I]	0.00E+00	N/A
	bpr_I2048	*tpiA*	triosephosphate isomerase	-	[G]	0.00E+00	N/A
	bpr_I2176	-	hypothetical protein	-	-	0.00E+00	N/A
	bpr_I2183	-	glycosyl transferase 2	GT2	[M]	0.00E+00	N/A
	bpr_I2275	-	hypothetical protein	-	-	0.00E+00	N/A
	bpr_I2293	-	hypothetical protein	-	-	0.00E+00	N/A
	bpr_I2437	*guaB*	IMP dehydrogenase	-	[F]	0.00E+00	N/A
	bpr_I2471	-	N-acetylmuramoyl-L-alanine amidase	-	[M]	0.00E+00	N/A
	bpr_I2749	-	peptide ABC transporter permease	-	[P]	0.00E+00	N/A
	bpr_I2933	-	RNA-binding S4 domain-containing protein	-	[K]	0.00E+00	N/A
	bpr_II009	-	hypothetical protein	-	-	0.00E+00	N/A
	bpr_II088	-	hypothetical protein	-	-	0.00E+00	N/A
	bpr_II203	-	resolvase family protein	-	[L]	0.00E+00	N/A
	bpr_III224	-	amino acid ABC transporter permease	-	[E]	0.00E+00	N/A
	bpr_IV097	-	restriction endonuclease family protein	-	[S]	0.00E+00	N/A
B316 co-culture growth on pectin	bpr_I2860	*gh31D*	glycoside hydrolase 31	GH31	[G]	1.55E-05	3.64
	bpr_I1161	*atpG2*	ATP synthase F1 subunit gamma	-	[C]	1.07E-08	3.52
	bpr_III140	-	GGDEF/EAL domain-containing protein	-	[T]	8.03E-04	3.06
	bpr_I1037	*pgm*	phosphoglucomutase	-	[G]	1.83E-03	2.94
	bpr_I2264	-	sugar ABC transporter substrate-binding protein	-	[G]	1.17E-03	2.68
	bpr_II257	-	hypothetical protein	-	-	1.96E-03	2.65
	bpr_III242	*urtB*	sugar ABC transporter permease protein	-	[G]	1.83E-03	2.18

Genes were deemed "significantly expressed" if all three thresholds were met: FDR calculated using the Benjamini-Hochberg procedure *Q*-values < 0.05, KW analysis of variance with $p < 0.05$, and ≥ 2-fold $\log_2$ difference in expression. CAZy designations have been assigned to genes containing the relevant domains. COG classifications: [C], Energy production and conversion; [D], Cell cycle control, cell division, chromosome partitioning; [E], Amino acid transport and metabolism; [F], Nucleotide transport and metabolism; [G], Carbohydrate transport and metabolism; [H], Coenzyme transport and metabolism; [I], Lipid transport and metabolism; [J], Translation, ribosomal structure and biogenesis; [K], Transcription; [L], Replication, recombination and repair; [M], Cell wall-membrane-envelope biogenesis; [N], Cell motility; [O], Posttranslational modification, protein turnover, chaperones; [P], Inorganic ion transport and metabolism; [Q], Secondary metabolites biosynthesis, transport and catabolism; [R], General function prediction only; [S], Function unknown; [T], Signal transduction mechanisms; [U], Intracellular trafficking, secretion, and vesicular transport; [V], Defence mechanisms.

Table A3.18. List of genes significantly up-regulated during mono- and co-culture growth of *B. hungatei* MB2003 on pectin.

Condition	Locus tag	Name	Annotation	CAZy	COG	FDR	log_2 FC
MB2003 mono-culture growth on pectin	bhn_I1925		hypothetical protein	-	-	0.00E+00	6.99
	bhn_I0806	-	arginase family protein	-	-	0.00E+00	6.98
	bhn_II008	-	TraG/TraD family protein	-	-	0.00E+00	6.92
	bhn_III55	*rsmH*	16S rRNA methyltransferase	-	[M]	0.00E+00	6.91
	bhn_III007	-	HTH domain-containing protein	-	[K]	0.00E+00	6.61
	bhn_I2638	*guaA*	GMP synthase	-	[F]	0.00E+00	6.51
	bhn_II037	-	hypothetical protein	-	-	0.00E+00	6.43
	bhn_I2475	-	hypothetical protein	-	-	0.00E+00	6.29
	bhn_I1132	*sufS*	arginine decarboxylase	-	[E]	3.55E+167	6.19
	bhn_III64	*bgl3A*	β-glucosidase	GH3	[G]	5.66E+53	6.13
	bhn_III70	-	HTH domain-containing protein	-	[K]	0.00E+00	6.10
	bhn_I0913	*pcrA*	ATPdependent DNA helicase	-	[L]	9.20E+223	6.09
	bhn_III76	-	lysozyme	GH25	[M]	2.98E+19	6.04
	bhn_I1599	-	hypothetical protein	-	-	3.39E+289	6.02
	bhn_I2699	*dapC*	aspartate/tyrosine/aromatic aminotransferase	-	[E]	1.27E+187	5.98
	bhn_I2510	-	betalactamase	-	[R]	1.19E+184	5.91
	bhn_I2226	*cbp94A*	cellobiose phosphorylase	GT36	[G]	1.10E+86	5.83
	bhn_I1337	*pepT*	peptidase T	-	[E]	5.97E+117	5.81
	bhn_I0865	*polC*	DNA polymerase III subunit alpha	-	[L]	1.68E+301	5.78
	bhn_I2381	*mcpL*	methylaccepting chemotaxis protein	-	[N]	1.71E+265	5.73
	bhn_I1982	*fusA*	translation elongation factor G	-	[J]	6.94E+125	5.70
	bhn_I0922	-	restriction system component	-	[V]	2.11E+160	5.69
	bhn_II063	-	ATP-dependent DNA helicase UvrD/REP family	-	[L]	9.41E+138	5.67
	bhn_I1844	-	sugar ABC transporter permease	-	[G]	4.72E+248	5.66
	bhn_III59	-	UvrD/REP helicase	-	[L]	0.00E+00	5.64
	bhn_I0888	*infB*	translation initiation factor IF2	-	[J]	6.40E+140	5.61
	bhn_I1924	-	hypothetical protein	-	-	5.29E+222	5.61
	bhn_I0127	-	amino acid ABC transporter substrate-binding protein	-	[E]	2.20E+197	5.58
	bhn_I0643	-	CAAX amino terminal protease family protein	-	[R]	5.46E+182	5.58
	bhn_I1953	*pl11A*	polysaccharide lyase	-	[G]	5.93E+136	5.51
	bhn_I0177	-	hypothetical protein	-	-	1.50E+48	5.49
	bhn_I1885	-	hypothetical protein	-	-	1.78E+122	5.46
	bhn_I1912	-	Wall associated protein	-	-	1.20E+90	5.36

	bhn_III19	-	hypothetical protein	-	-	4.34E+203	5.30
	bhn_II098	-	hypothetical protein	-	-	3.89E+258	5.30
	bhn_I0546	-	glycosyl transferase GT2	GT2	[R]	7.11E+143	5.29
	bhn_I0466	*fkbH*	FkbH domain-containing protein	-	[Q]	1.17E+102	5.22
	bhn_I2118	-	hypothetical protein	-	-	6.31E+105	5.21
	bhn_I2457	*pseF*	pseudaminic acid CMP-transferase	-	[M]	5.42E+52	5.19
	bhn_I2152	*pilF*	TPR domain-containing protein	-	[N]	4.92E+47	5.13
	bhn_I0977	*nikA*	nickel/peptide ABC transporter substrate-binding protein	-	[E]	2.65E+64	5.11
	bhn_I1339	*mutS*	DNA mismatch repair protein	-	[L]	4.21E+70	5.11
	bhn_I0976	-	oligo-endopeptidase F	-	[E]	3.31E+90	5.09
	bhn_I2349	-	glycosyl transferase GT2/GT4 family	GT2/GT4	[M]	3.59E+58	5.07
	bhn_I2734	-	sugar ABC transporter substrate-binding protein	-	[G]	5.04E+111	5.07
	bhn_I1871	-	helicase SNF2 family	-	[K]	6.73E+58	5.04
	bhn_III66	-	Na^{+}/H^{+} antiporter	-	[P]	1.85E-03	5.04
	bhn_I0835	*ccpA*	catabolite control protein A	-	[K]	8.09E+76	5.02
	bhn_I2695	*phoB*	two component system response regulator	-	[T]	1.21E+94	5.01
	bhn_III52	-	aminotransferase	-	[E]	5.69E+130	5.00
	bhn_III51	-	ACT domain-containing protein	-	[R]	4.41E+121	4.99
	bhn_I1064	*feoB*	ferrous iron transport protein B	-	[P]	6.63E+42	4.98
	bhn_I2625	-	Mn^{2+}/Zn^{2+} ABC transporter substrate-binding protein	-	[P]	8.65E+95	4.97
	bhn_I0187	*gh115A*	α-glucuronidase	GH115	[G]	4.39E+53	4.94
	bhn_I0477	*mviN*	polysaccharide biosynthesis protein	-	[R]	2.42E+44	4.94
	bhn_II068	-	hypothetical protein	-	-	7.89E+81	4.94
	bhn_I2413	*ugl88B*	unsaturated glucuronyl hydrolase	-	[R]	4.34E+58	4.93
	bhn_I0488	-	MBOAT family acyltransferase	-	[M]	2.01E+51	4.92
	bhn_I0383	-	hypothetical protein	-	-	2.85E+48	4.91
	bhn_I1340	*mutL*	DNA mismatch repair protein	-	[L]	7.78E+32	4.91
	bhn_I1527	-	hypothetical protein	-	-	3.45E+38	4.91
	bhn_I0900	*prfC*	peptide chain release factor 3	-	[J]	9.01E+56	4.88
	bhn_I0813	-	sugar ABC transporter substrate-binding protein	-	[G]	1.63E+40	4.87
	bhn_I1647	*pheA*	prephenate dehydratase/chorismate mutase	-	[E]	2.07E+52	4.87
	bhn_I0709	-	hypothetical protein	-	-	3.25E+61	4.86
	bhn_I1118	*topB*	DNA topoisomerase III	-	[L]	7.62E+50	4.83
	bhn_II059	-	hypothetical protein	-	-	8.33E+103	4.82
	bhn_I0696	-	GGDEF/TPR domain-containing protein	-	[T]	6.51E+37	4.81
	bhn_I2433	-	glycosyl transferase GT4 family	GT4	[M]	2.83E+37	4.81

	bhn_II061	-	hypothetical protein	-	-	1.30E+63	4.78
	bhn_I0132	*matE*	multidrug transporter	-	[V]	7.67E+40	4.77
	bhn_I1239	-	TetR family transcriptional regulator	-	[K]	2.84E+91	4.77
	bhn_III32	*parA*	plasmid partitioning protein	-	[D]	3.71E+45	4.77
	bhn_I2254	*paaC*	3-hydroxybutyryl-CoA dehydrogenase	-	[I]	1.94E+57	4.75
	bhn_I0316	-	helicase	-	[K]	8.23E+39	4.72
	bhn_II005	-	hypothetical protein	-	-	4.66E+56	4.72
	bhn_I0077	*pyk*	pyruvate kinase	-	[G]	6.00E+55	4.70
	bhn_I1829	*priA*	primosomal protein N'	-	[L]	5.83E+47	4.70
	bhn_I2326	*tagD*	glycerol-3-phosphate cytidylyltransferase	-	[M]	1.47E+44	4.69
	bhn_I2454	-	glycosyl transferase GT2 family	GT2	[M]	4.98E+50	4.69
	bhn_I2544	*murC*	UDPN-acetylmuramatealanine ligase	-	[M]	3.45E+29	4.62
	bhn_I2728	-	GGDEF/EAL domain-containing protein	-	[T]	8.99E+59	4.62
	bhn_I0971	*xylF*	xylose ABC transporter substrate-binding protein	-	[G]	1.74E+42	4.62
	bhn_III12	-	acyltransferase	-	[I]	4.53E+31	4.61
	bhn_I2293	-	ribonuclease	-	[R]	3.64E+29	4.57
	bhn_I0448	-	glycosyl transferase 4	GT4	[M]	1.05E+29	4.56
	bhn_I0067	-	FHA domain-containing protein	-	[T]	3.03E+30	4.55
	bhn_I2441	*pseC*	aminotransferase	-	[M]	2.33E+59	4.55
	bhn_I1972	*aga36C*	α-galactosidase	GH27	[G]	4.38E+28	4.54
	bhn_II046	-	hypothetical protein	-	-	8.23E+122	4.54
	bhn_I1004	*xylF*	xylose ABC transporter substrate-binding protein	-	[G]	1.34E+30	4.54
	bhn_I0437	-	AbiH-like protein	-	-	1.07E+29	4.52
	bhn_I0512	-	mannose1phosphate guanylyltransferase	-	[M]	1.30E+24	4.52
	bhn_I0697	*valS*	valyltRNA synthetase	-	[J]	4.95E+41	4.52
	bhn_I1868	-	MgtC family protein	-	[S]	5.30E+40	4.52
	bhn_I0178	-	sugar ABC transporter permease protein	-	[G]	2.52E+36	4.51
	bhn_I0084	*gh27A*	α-galactosidase	GH27	[G]	8.45E+34	4.50
	bhn_I1110	-	carboxylesterase	-	[I]	2.04E+39	4.50
	bhn_I0911	*recN*	DNA repair protein	-	[L]	9.46E+20	4.49
	bhn_I2570	-	two component system response regulator	-	[T]	1.56E+37	4.49
	bhn_I0134	-	hypothetical protein	-	-	1.62E+30	4.48
	bhn_I0210	*xylB*	xylulokinase	-	[G]	2.99E+21	4.48
	bhn_I0049	*sufE*	FeS assembly protein	-	[C]	2.75E+44	4.46
	bhn_I0970	*ntrC*	AraC family transcriptional regulator	-	[K]	2.30E+31	4.46
	bhn_I1485	-	YkuD domain-containing protein	-	[S]	1.78E+22	4.46

	bhn_I1964	*purL*	phosphoribosylformylglycinamidine synthase	-	[F]	9.84E+33	4.46
	bhn_I2621	-	GGDEF domain-containing protein	-	[T]	5.74E+24	4.46
	bhn_I1145	*mcpF*	methyl-accepting chemotaxis protein	-	[N]	2.09E+18	4.42
	bhn_I1285	*xylG*	ABC transporter ATP-binding protein	-	[R]	1.94E+24	4.42
	bhn_I1006	-	two component system response regulator	-	[T]	1.43E+38	4.41
	bhn_I0172	-	sugar ABC transporter substrate-binding protein	-	[G]	4.06E+27	4.39
	bhn_I1156	-	peptidase M16	-	[R]	8.65E+33	4.39
	bhn_I2552	-	hypothetical protein	-	-	1.18E+27	4.39
	bhn_I2608	-	FRG domain-containing protein	-	-	1.17E+42	4.39
	bhn_I1968	-	GDSL-like lipase/acylhydrolase	-	[E]	1.31E+34	4.38
	bhn_I2389	-	lipoprotein NLPA family	-	[P]	1.88E+27	4.38
	bhn_I2682	-	serine recombinase	-	[L]	1.97E+27	4.38
	bhn_I1486	*ilvB*	acetolactate synthase catalytic subunit	-	[E]	2.95E+24	4.37
	bhn_I0151	-	hypothetical protein	-	-	1.22E+22	4.36
	bhn_I0681	*agn53A*	arabinogalactan endo-1,4-β-galactosidase	GH53	[G]	1.71E+22	4.36
	bhn_II064	-	DNA polymerase III delta subunit	-	[L]	3.49E+27	4.36
	bhn_I0749	-	leucyl aminopeptidase	-	[E]	2.69E+20	4.35
	bhn_I2176	*hlyB*	ABC transporter ATPbinding/permease protein	-	[V]	4.38E+28	4.35
	bhn_I2539	*mfd*	transcriptionrepair coupling factor	-	[L]	7.16E+26	4.35
	bhn_I0407	-	histidine kinase	-	[T]	1.80E+24	4.34
	bhn_II021	-	ATPase involved in pili biogenesis	-	[U]	3.65E+37	4.34
	bhn_II114	-	hypothetical protein	-	-	3.65E+37	4.33
	bhn_I1806	-	carboxylesterase	-	[R]	9.19E+22	4.31
	bhn_I1920	-	hypothetical protein	-	-	1.12E+27	4.31
	bhn_I0449	-	glycosyl transferase	GT4	[M]	1.32E+25	4.31
	bhn_I0516	-	hypothetical protein	-	-	6.16E+26	4.30
	bhn_I1209	-	hypothetical protein	-	-	2.61E+38	4.30
	bhn_II090	-	AraC family transcriptional regulator	-	-	1.52E+34	4.30
	bhn_I0139	*amt*	ammonium transporter	-	[P]	3.50E+24	4.29
	bhn_I2495	-	serine/threonine protein kinase	-	[R]	1.86E+28	4.29
	bhn_I0655	-	aldo/keto reductase	-	[C]	3.29E+23	4.28
	bhn_I2665	*cpdB*	5'-nucleotidase	-	[F]	5.10E+22	4.28
	bhn_I1488	-	hypothetical protein	-	-	2.36E+16	4.27
	bhn_I2197	*msbA*	ABC transporter ATP-binding/permease protein	-	[V]	3.30E+18	4.27
	bhn_I2576	-	SIR2 family protein	-	[K]	3.96E+22	4.27
	bhn_II048	-	hypothetical protein	-	-	1.57E+25	4.26

	bhn_II004	-	restriction endonuclease	-	[V]	2.62E+26	4.26
	bhn_I0307	*est1B*	feruloyl esterase	CE1	[R]	2.56E+23	4.25
	bhn_I0537	-	hypothetical protein	-	-	1.19E+26	4.25
	bhn_I0597	-	glycosyl transferase family 2	GT2	[M]	1.91E+20	4.25
	bhn_I2213	*guaB*	inosine-5'-monophosphate dehydrogenase	-	[F]	4.07E+25	4.25
	bhn_II022	-	cell surface protein	-	-	5.73E+43	4.24
	bhn_I0100	*fucO*	iron-containing alcohol dehydrogenase	-	[C]	2.64E+24	4.24
	bhn_I2181	*hisC*	histidinolphosphate transaminase	-	[E]	8.21E+21	4.24
	bhn_I1592	*atpD*	ATP synthase F1 beta subunit	-	[C]	3.07E+24	4.23
	bhn_I1217	*hsdR*	type I restrictionmodification system R subunit	-	[V]	2.14E+17	4.22
	bhn_I1611	-	AraC family transcriptional regulator	-	[T]	1.93E+17	4.22
	bhn_I2074	-	two component system histidine kinase	-	[T]	2.96E+24	4.22
	bhn_I0242	-	GNAT family acetyltransferase	-	-	4.38E+27	4.22
	bhn_I1479	*spoT*	RelA/SpoT domain-containing protein	-	[T]	3.94E+13	4.21
	bhn_I1412	*hypE*	hydrogenase expression/formation protein	-	[O]	3.60E+17	4.20
	bhn_I0822	*rfbB*	NAD dependent epimerase/dehydratase	-	[M]	9.46E+22	4.19
	bhn_II115	-	hypothetical protein	-	-	1.05E+29	4.19
	bhn_I0661	*udk*	adenylylsulfate kinase	-	[P]	9.73E+22	4.19
	bhn_I0904	*xseA*	exo-deoxyribonuclease VII	-	[L]	2.08E+23	4.19
	bhn_I1576	*minD*	ATP-binding protein	-	[D]	1.19E+28	4.17
	bhn_I2212	-	GGDEF/EAL domain-containing protein	-	[T]	6.70E+20	4.17
	bhn_I2243	-	N6 DNA methylase	-	[L]	8.09E+16	4.15
	bhn_III86	*hicA*	DNA repair protein	-	[S]	3.82E+41	4.15
	bhn_I0092	*gh105B*	unsaturated rhamnogalacturonyl hydrolase	GH88/GH105	[G]	1.26E+16	4.15
	bhn_I0207	*comE*	RNA-binding protein S1	-	[K]	1.66E+19	4.14
	bhn_I1604	*gh31A*	glycoside hydrolase family 31	GH31	[G]	5.45E+35	4.14
	bhn_I0682	*purO*	IMP cyclohydrolase	-	[F]	4.51E+24	4.14
	bhn_I0497	-	glycosyl transferase GT4	GT4	[M]	4.68E+28	4.13
	bhn_I0133	-	membrane protein	-	[S]	7.09E+19	4.11
	bhn_I2738	*amy13A*	α-amylase	GH13	[G]	7.67E+18	4.11
	bhn_I1646	*ppx*	exopolyphosphatase	-	[P]	1.02E+15	4.10
	bhn_I1699	*alaS*	alanyltRNA synthetase	-	[J]	2.48E+21	4.10
	bhn_I1054	-	hypothetical protein	-	-	1.20E+37	4.09
	bhn_I1818	-	serine/threonine protein kinase	-	[R]	2.31E+21	4.09
	bhn_I2139	*xsa43D*	xylosidase/arabinofuranosidase	GH43	[G]	9.78E+18	4.09
	bhn_I2386	*cobT*	nicotinatenucleotide-di-methyl-benzimidazole phosphoribosyltransferase	-	[H]	4.87E+17	4.09

	bhn_I2351	-	glycosyl transferase GT2 family	GT2	[R]	2.66E+25	4.07
	bhn_I0730	*arch*	serine/threonine protein kinase	-	[R]	2.68E+17	4.07
	bhn_I1254	*cobQ*	cobyric acid synthase	-	[H]	1.72E+15	4.07
	bhn_I0006	*gyrA*	DNA gyrase subunit A	-	[L]	3.33E+14	4.06
	bhn_I1112	*glpA*	FAD-dependent oxidoreductase	-	[R]	2.82E+17	4.06
	bhn_I1863	*ispH*	4-hydroxy-3-methyl-but-2-enyl diphosphate reductase	-	[I]	6.96E+19	4.06
	bhn_I1238	*msbA*	ABC transporter ATP-binding protein	-	[V]	1.37E+14	4.04
	bhn_I1558	-	cell surface protein	-	-	5.14E+17	4.03
	bhn_I1926	*msbA*	ABC transporter ATP-binding/permease protein	-	[V]	5.85E+11	4.03
	bhn_I1308	*aga27A*	α-galactosidase	GH27	[G]	1.12E+16	4.02
	bhn_I0959	-	MFS transporter	-	[G]	5.87E+16	4.02
	bhn_I1607	-	sugar ABC transporter substrate-binding protein	-	[G]	1.13E+17	4.02
	bhn_I2303	-	ATPase	-	[R]	1.69E+18	4.02
	bhn_I1010	*urtB*	sugar ABC transporter permease protein	-	[G]	7.31E+21	4.00
	bhn_I1160	*glyS*	glycyl-tRNA synthetase	-	[J]	9.27E+17	3.99
	bhn_I0530	-	hypothetical protein	-	-	1.18E+13	3.98
	bhn_I0534	-	hypothetical protein	-	-	1.21E+12	3.98
	bhn_I1913	-	hypothetical protein	-	-	6.09E+13	3.98
	bhn_I1032	-	ATPase	-	[R]	1.41E+12	3.96
	bhn_I1785	-	phospholipase	-	[R]	3.61E+19	3.96
	bhn_I0695	-	helicase	-	[K]	2.81E+20	3.95
	bhn_I1976	*purA*	adenylosuccinate synthetase	-	[F]	5.06E+15	3.95
	bhn_II089	-	archaeal ATPase family protein	-	[R]	9.95E+24	3.95
	bhn_I0857	-	hypothetical protein	-	-	1.22E+14	3.94
	bhn_I2450	-	hypothetical protein	-	-	2.66E+13	3.94
	bhn_III67	-	hypothetical protein	-	-	1.78E+06	3.94
	bhn_I1721	*matE*	sugar transporter	-	[V]	1.43E+12	3.93
	bhn_I2158	-	glycosyl transferase GT4 family	GT4	[S]	1.78E+11	3.93
	bhn_I1222	-	membrane protein	-	-	4.67E+16	3.92
	bhn_I1413	*hypD*	hydrogenase expression/formation protein	-	[O]	2.77E+14	3.92
	bhn_I0056	*fliS*	flagellar protein	-	[N]	3.73E+19	3.91
	bhn_I0286	*gh31B*	α-glucosidase	GH31	[G]	4.66E+14	3.91
	bhn_I0385	-	hypothetical protein	-	-	4.06E+15	3.91
	bhn_III18	-	anti-sigma factor antagonist	-	[T]	4.27E+20	3.91
	bhn_I0615	*rpoC*	DNAdirected RNA polymerase subunit α/β'	-	[K]	6.20E+18	3.88
	bhn_I1554	*guaB*	inorganic pyrophosphatase	-	[C]	7.52E+12	3.88

	bhn_I2707	-	hypothetical protein	-	-	1.77E+12	3.87
	bhn_I0601	-	glycosyl transferase GT2 family	GT2	[R]	1.03E+13	3.87
	bhn_I0117	-	P-loop-containing ATPase	-	[R]	1.88E+11	3.86
	bhn_I0636	-	capsule biosynthesis protein	-	[M]	7.43E+15	3.85
	bhn_I0192	*xyl3A*	β-xylosidase	GH3	[G]	4.58E+12	3.85
	bhn_I2538	*msbA*	ABC transporter ATPbinding/permease protein	-	[V]	2.27E+12	3.85
	bhn_I0115	-	ROK family protein	-	[K]	3.46E+14	3.85
	bhn_I2180	-	small multidrug export	-	[S]	3.59E+17	3.84
	bhn_I0868	-	β-lactamase	-	[V]	1.22E+11	3.83
	bhn_I0195	*rfbF*	UTP-glucose1phosphate uridylyl-transferase	-	[G]	2.74E+09	3.81
	bhn_I0412	*pseB*	UDP-Nacetylglucosamine 4,6-de-hydratase	-	[M]	2.90E+10	3.81
	bhn_I0770	-	phage integrase family protein	-	[L]	1.10E+10	3.81
	bhn_I1732	*nikA*	nickel/peptide ABC transporter ATPbinding protein	-	[E]	2.96E+13	3.81
	bhn_I2015	-	glycosyl transferase GT4 family	GT4	[M]	3.62E+11	3.81
	bhn_I2530	-	hypothetical protein	-	-	1.08E+11	3.81
	bhn_I2361	-	glycosyl transferase GT4 family	GT4	[S]	1.57E+15	3.80
	bhn_I2571	-	two component system histidine kinase	-	[T]	2.09E+13	3.79
	bhn_II018	-	hypothetical protein	-	-	5.09E+14	3.79
	bhn_I0450	*glgA*	glycosyl transferase	GT4	[M]	5.07E+09	3.79
	bhn_I2156	*nuoI*	4-Fe_4S-binding domain-containing protein	-	[S]	8.21E+10	3.79
	bhn_I0082	-	methyltransferase domain-containing protein	-	[R]	1.68E+10	3.78
	bhn_I1321	-	cell surface protein	-	[M]	8.68E+18	3.78
	bhn_I1805	-	hypothetical protein	-	-	6.55E+12	3.78
	bhn_I0545	-	SAM-dependent methyltransferase	-	-	2.29E+10	3.77
	bhn_I1059	*trpE*	anthranilate synthase component I	-	[E]	9.84E+11	3.77
	bhn_I0183	-	acetyl-xylan esterase	GH2/DUF303	[G]	9.96E+15	3.77
	bhn_I0068	-	hypothetical protein	-	[I]	1.97E+10	3.75
	bhn_I1168	*oadG*	sodium ionpumping decarboxylase gamma subunit	-	-	3.51E+10	3.75
	bhn_I2364	*flgL*	flagellar hook-associated protein	-	[N]	4.45E+07	3.75
	bhn_I2663	*metG*	methionyl-tRNA synthetase	-	[J]	4.20E+09	3.75
	bhn_I2692	*rumA*	23S rRNA (uracil5)-methyltransferase	-	[J]	1.90E+12	3.75
	bhn_I1408	*metA*	homoserine O-succinyltransferase	-	[E]	4.20E+10	3.74
	bhn_I1506	-	FAD-dependent oxidoreductase	-	[R]	8.38E+11	3.74
	bhn_I1370	-	ribosomal protein	-	[J]	4.36E+12	3.73
	bhn_I0520	*pds*	flavin containing amine oxidoreductase	-	[H]	1.03E+10	3.70
	bhn_I0528	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	-	[I]	1.81E+09	3.70

	bhn_I0885	*nusA*	transcription elongation factor	-	[K]	1.58E+10	3.70
	bhn_I0907	*dxs*	1-deoxy-D-xylulose-5-phosphate synthase	-	[H]	4.47E+11	3.70
	bhn_I0908	*tly*	rRNA methyltransferase	-	[J]	3.47E+09	3.70
	bhn_I1713	-	hypothetical protein	-	-	9.20E+11	3.70
	bhn_I2382	*speE*	spermine synthase	-	[R]	1.10E+08	3.70
	bhn_I2428	-	MBOAT family acyltransferase	-	[M]	2.25E+07	3.70
	bhn_I0863	*rseP*	peptidase M50 family	-	[M]	3.05E+09	3.68
	bhn_I2128	-	two component system response regulator	-	[T]	4.40E+10	3.68
	bhn_I0590	*modB*	ABC transporter permease	-	[G]	3.07E+08	3.68
	bhn_I1574	*ilvE*	branchedchain amino acid aminotransferas	-	[E]	7.56E+13	3.68
	bhn_I2137	-	FAD-dependent oxidoreductase	-	[R]	3.75E+11	3.68
	bhn_I2771	*gh30A*	glycoside hydrolase family 30	GH30	[G]	5.91E+13	3.68
	bhn_III003	-	diguanylate cyclase	-	[T]	2.11E+45	3.66
	bhn_III87	*hicB*	DNA repair protein	-	[S]	2.18E+15	3.66
	bhn_I0231	-	nickel/peptide ABC transporter permease protein	-	[E]	3.82E+10	3.64
	bhn_I1618	-	GGDEF/HAMP domain-containing protein	-	[T]	6.83E+08	3.64
	bhn_II072	-	phage integrase family protein	-	[L]	1.49E+16	3.64
	bhn_I2300	-	xylose isomerase domain-containing protein	-	[G]	9.92E+12	3.63
	bhn_III85	-	hypothetical protein	-	-	7.80E+11	3.63
	bhn_I0376	-	hypothetical protein	-	-	9.16E+11	3.63
	bhn_I1894	-	CobQ/CobB/MinD/ParA nucleotide binding domain-containing protein	-	[D]	3.82E+10	3.63
	bhn_I2043	-	GNAT family acetyltransferase	-	[K]	6.61E+11	3.62
	bhn_I1661	*thrC*	threonine synthase	-	[E]	2.19E+06	3.61
	bhn_I0519	-	GDSLlike lipase/acylhydrolase	-	-	1.75E+07	3.61
	bhn_I2404	*rnfA*	electron transporter	-	[C]	3.75E+11	3.61
	bhn_I0176	-	hypothetical protein	-	-	5.11E+09	3.60
	bhn_I0953	-	acetyltransferase	-	[R]	2.21E+10	3.60
	bhn_I2023	*hpaA*	AraC family transcriptional regulator	-	[K]	3.31E+08	3.60
	bhn_I2045	-	hypothetical protein	-	-	1.47E+08	3.60
	bhn_I0511	*eryC1*	aminotransferase	-	[M]	1.71E+08	3.58
	bhn_I1369	*rpmA*	ribosomal protein L27	-	[J]	9.90E+10	3.57
	bhn_I2311	-	hypothetical protein	-	-	2.03E+08	3.57
	bhn_I0654	*aga36A*	α-galactosidase	GH27	[G]	5.20E+10	3.56
	bhn_I0727	*LeuS*	leucyl-tRNA synthetase	-	[J]	2.04E+10	3.54
	bhn_I1348	*rodA*	rod shape-determining protein	-	[D]	1.04E+07	3.54
	bhn_I2640	*asnS*	asparaginyl-tRNA synthetase	-	[J]	4.44E+09	3.54

	bhn_I0278	-	hypothetical protein	-	-	1.11E+08	3.54
	bhn_I2283	-	hypothetical protein	-	-	1.87E+08	3.54
	bhn_I1927	*msbB*	ABC transporter ATP-binding/permease protein	-	[V]	9.02E+09	3.54
	bhn_I0258	-	ATC hydrolase family protein	-	-	1.83E+07	3.53
	bhn_I2281	*coaBC*	phosphopantothenoylcysteine decarboxylase	-	[H]	6.52E+08	3.52
	bhn_I0680	*amy13D*	α-amylase	GH13	[G]	1.39E+06	3.51
	bhn_I1959	*hpaA*	AraC family transcriptional regulator	-	[T]	1.27E+07	3.51
	bhn_I2238	*pbp1A*	penicillinbinding protein 1A family	GT51	[M]	8.06E+09	3.51
	bhn_I0635	-	hypothetical protein	-	-	4.81E+10	3.50
	bhn_I1882	*gspF*	type II secretion system protein F	-	[N]	7.29E+08	3.50
	bhn_I1775	-	hypothetical protein	-	-	2.46E+07	3.49
	bhn_II040	-	hypothetical protein	-	-	1.05E+08	3.49
	bhn_III36	-	MerR family transcriptional regulator	-	[K]	4.23E+10	3.49
	bhn_II060	-	ThiF family protein	-	[H]	2.67E+09	3.48
	bhn_I2010	-	glycosyl transferase group 2 family protein	GT2	[R]	4.47E+08	3.47
	bhn_I0215	*proC*	pyrroline-5-carboxylate reductase	-	[E]	9.43E+08	3.46
	bhn_I0796	*cysE*	serine acetyltransferase	-	[E]	4.29E+06	3.46
	bhn_I1831	*recJ*	single-stranded-DNA-specific exonuclease	-	[L]	3.83E+07	3.46
	bhn_I1880	-	GGDEF domain-containing protein	-	[T]	9.09E+07	3.46
	bhn_I2054	*cbiO*	cobalt ABC transporter ATP-binding protein	-	[P]	2.14E+08	3.46
	bhn_I2779	*yesM*	two component system histidine kinase	-	[T]	4.68E+07	3.46
	bhn_I0582	*sdhA*	serine dehydratase	-	[E]	7.32E+09	3.45
	bhn_I0288	-	O-acetyl-homoserine aminocarboxypropyltransferase	-	[E]	6.66E-06	3.44
	bhn_I0599	*pseC*	lipopolysaccharide biosynthesis protein	-	[M]	1.98E+07	3.43
	bhn_I2317	-	carbamoyltransferase family protein	-	[O]	2.41E+09	3.43
	bhn_II002	-	hypothetical protein	-	-	1.83E+06	3.42
	bhn_I0169	*xsa43A*	xylosidase/arabinofuranosidase	GH43/CBM6	[G]	1.89E+09	3.41
	bhn_I0899	*pnp*	polyribonucleotide nucleotidyltransferase	-	[J]	5.89E+07	3.41
	bhn_I1690	-	hypothetical protein	-	-	4.18E-05	3.41
	bhn_I2568	*phnU*	sugar ABC transporter permease protein	-	[G]	1.63E-05	3.41
	bhn_I2315	-	glycosyl transferase family 8	GT8	-	2.02E+10	3.40
	bhn_I0969	*cztS*	histidine kinase	-	[T]	1.08E-06	3.39
	bhn_I2141	*thiC*	thiamine biosynthesis protein	-	[H]	9.44E+07	3.39
	bhn_I2427	-	hypothetical protein	-	-	2.62E+06	3.39
	bhn_I1792	*menG*	SAM-dependent methyltransferase	-	[Q]	3.07E+08	3.38
	bhn_II062	-	hypothetical protein	-	-	5.20E+12	3.37

	bhn_I0572	-	hypothetical protein	-	-	9.53E+06	3.37
	bhn_I1315	*ssnA*	amidohydrolase	-	[F]	1.21E-05	3.37
	bhn_I0356	-	VTC domain-containing protein	-	-	1.32E+07	3.36
	bhn_I2643	-	hypothetical protein	-	-	5.67E+07	3.36
	bhn_I0978	*nikB*	nickel/peptide ABC transporter substrate-binding protein	-	[R]	2.39E+06	3.36
	bhn_I2187	*eps*	polysaccharide biosynthesis protein	-	[M]	3.14E+06	3.36
	bhn_I2799	*trmE*	tRNA modification GTPase	-	[R]	1.25E-05	3.36
	bhn_I0758	*clpB*	ATPdependent chaperone protein	-	[O]	8.57E+08	3.35
	bhn_I1035	*degV*	DegV family protein	-	[S]	3.57E+06	3.35
	bhn_I1392	*degV*	di-hydroxyacetone kinase	-	[S]	2.72E-05	3.35
	bhn_I0514	-	hypothetical protein	-	-	9.69E+07	3.34
	bhn_I1989	-	GGDEF domain-containing protein	-	[T]	3.31E+08	3.34
	bhn_I0493	-	acyltransferase	-	[S]	2.49E+07	3.32
	bhn_I0906	*hepT*	geranylgeranyl pyrophosphate synthase	-	[H]	3.02E-05	3.32
	bhn_I2681	-	acyltransferase	-	-	2.31E+06	3.32
	bhn_I0343	*thiD*	phosphomethylpyrimidine kinase	-	[H]	6.58E-06	3.31
	bhn_I0551	*wcaJ*	exopolysaccharide biosynthesis polyprenyl glycosyl-phosphotransferase	-	[M]	1.40E+06	3.30
	bhn_I1406	-	hypothetical protein	-	-	1.74E+07	3.30
	bhn_I2247	*cbiQ*	cobalt ABC transporter permease	-	[P]	1.58E-05	3.30
	bhn_I0094	-	hypothetical protein	-	-	1.78E+06	3.30
	bhn_I1615	-	MBOAT family acyltransferase	-	[M]	5.17E-06	3.30
	bhn_I2628	*chr*	chromate transporter	-	[P]	9.12E+07	3.30
	bhn_I0578	*ytkD*	NUDIX hydrolase	-	[L]	9.88E+07	3.30
	bhn_I0840	*bga2C*	β-galactosidase	GH2	[G]	4.62E+06	3.30
	bhn_I1113	*trxB*	pyridine nucleotidedisulfide oxidoreductase	-	[R]	5.47E-05	3.30
	bhn_I2697	*degP*	trypsin-like serine protease	-	[O]	1.21E-05	3.30
	bhn_II042	-	hypothetical protein	-	-	5.81E+12	3.30
	bhn_I1826	*ehuA*	amino acid ABC transporter ATP-binding protein	-	[E]	3.28E-05	3.29
	bhn_I0598	*rfbG*	CDP-glucose 4,6-dehydratase	-	[M]	2.22E+06	3.28
	bhn_I1336	*glgX2*	glycogen debranching enzyme	GH13	[G]	3.69E+06	3.27
	bhn_I2262	*ccpA*	LacI family transcription regulator	-	[K]	2.12E+06	3.26
	bhn_I2465	-	polysaccharide biosynthesis protein	-	-	1.57E-04	3.26
	bhn_I2065	*xsa43C*	xylosidase/arabinofuranosidase	GH43	[G]	1.33E+06	3.26
	bhn_I0753	*purD*	phosphoribosylamineglycine ligase	-	[F]	3.97E+07	3.25
	bhn_I1278	*ffh*	signal recognition particle protein	-	[U]	2.72E+06	3.25
	bhn_I0923	-	restriction endonuclease	-	[V]	5.46E-06	3.24

	bhn_I1550	-	hypothetical protein	-	-	2.12E+06	3.24
	bhn_I1873	-	carbohydrate binding protein	CBM2/6xCBM6	-	4.24E+09	3.24
	bhn_I2338	-	glycosyl transferase GT4 family	GT4	[S]	5.59E+06	3.23
	bhn_I0185	*agu67A*	α-D-glucuronidase	GH67	[G]	2.48E+08	3.23
	bhn_I1017	-	isoprenylcysteine carboxyl methyltransferase family protein	-	[S]	3.38E+07	3.23
	bhn_I1200	-	hypothetical protein	-	-	1.19E-06	3.22
	bhn_I1472	*topA*	DNA topoisomerase I	-	[L]	7.75E+06	3.22
	bhn_I2635	*ubiG*	SAM-dependent methyltransferase	-	[Q]	2.17E-05	3.22
	bhn_I2651	*murF*	UDPN-acetylmuramoyltripeptide-D-alanyl-D-alanine ligase	-	[M]	3.53E-05	3.22
	bhn_I1570	-	peptidase U32 family	-	[O]	6.51E+07	3.22
	bhn_I1193	-	β-lactamase family protein	-	[V]	1.01E+06	3.21
	bhn_I1407	*est4D*	polysaccharide deacetylase	CE4	[G]	9.58E-05	3.21
	bhn_I0564	-	hypothetical protein	-	-	1.08E-05	3.20
	bhn_I1018	-	transposase	-	[L]	1.80E-05	3.20
	bhn_I0504	*wecB*	glycosyl transferase 26	GT26	[M]	1.17E+06	3.20
	bhn_I0154	*aspS*	aspartyl-tRNA synthetase	-	[J]	6.09E+06	3.20
	bhn_I1695	*cysS*	cysteinyl-tRNA synthetase	-	[J]	4.73E-05	3.20
	bhn_I2634	-	aminoglycoside 6-adenylyltransferase	-	[E]	8.54E+06	3.19
	bhn_I0677	*nagA*	N-acetylglucosamine-6-phosphate deacetylase	-	[G]	7.23E-05	3.19
	bhn_I0924	-	hypothetical protein	-	-	4.65E-05	3.18
	bhn_I2355	-	glycosyl transferase GT4 family	GT4	[S]	1.49E+06	3.18
	bhn_I0121	-	Na/Pi cotransporter	-	[P]	7.83E+06	3.17
	bhn_I0329	-	ATPase	-	[R]	2.21E-04	3.17
	bhn_I0515	-	radical SAM protein	-	[R]	9.10E-05	3.17
	bhn_I0580	-	hypothetical protein	-	-	1.05E+06	3.17
	bhn_I1043	*nikC*	nickel/peptide ABC transporter permease protein	-	[E]	1.10E-04	3.17
	bhn_I1827	-	amino acid ABC transporter permease protein	-	[E]	1.05E-04	3.17
	bhn_I1955	-	hypothetical protein	-	-	7.21E+08	3.17
	bhn_I2437	-	nucleotide sugar dehydrogenase	-	[M]	7.30E-05	3.17
	bhn_I0337	*bchM*	SAM-dependent methyltransferase	-	[Q]	1.55E-05	3.15
	bhn_I1655	*pheS*	phenylalanyl-tRNA synthetase alpha chain	-	[J]	1.41E-04	3.15
	bhn_I0606	-	hypothetical protein	-	-	3.60E+07	3.14
	bhn_I2505	-	acyl-ACP thioesterase	-	[I]	9.44E-05	3.14
	bhn_III004	*xerC*	phage integrase family protein	-	[L]	4.97E+17	3.14
	bhn_I1782	-	methyl-accepting chemotaxis protein	-	[N]	1.17E-05	3.13
	bhn_I2345	-	WavE family lipopolysaccharide synthesis protein	-	-	5.48E-05	3.13

	bhn_I0441	*pseH*	GNAT family acetyltransferase	-	[J]	9.65E-05	3.12
	bhn_I0562	*rnz*	ribonuclease Z	-	[R]	2.06E-04	3.12
	bhn_III13	-	hypothetical protein	-	-	7.17E+07	3.12
	bhn_I1207	-	hypothetical protein	-	-	7.13E-05	3.11
	bhn_I1855	-	hypothetical protein	-	-	1.21E-04	3.11
	bhn_I0331	-	RNA polymerase sigma factor sigma70 family	-	[K]	3.30E-05	3.10
	bhn_I1152	-	hypothetical protein	-	-	1.95E-03	3.09
	bhn_I1705	-	phage integrase family protein	-	[L]	1.81E-05	3.09
	bhn_I2356	-	glycosyl transferase GT2 family	GT2	[M]	4.19E-04	3.09
	bhn_I1201	-	hypothetical protein	-	-	7.31E-05	3.08
	bhn_I2263	-	MFS transporter	-	[G]	6.91E-05	3.08
	bhn_I2591	*ebfC*	DNAbinding protein	-	[S]	3.39E-04	3.06
	bhn_I0843	-	HTH domain-containing protein	-	[K]	3.55E-04	3.06
	bhn_I0467	*acpP*	acyl carrier protein	-	[I]	2.19E-04	3.05
	bhn_I0989	-	glycosyl transferase GT2 family	GT2	[M]	6.23E-04	3.05
	bhn_I0410	*wecB*	UDPN-acetylglucosamine 2-epimerase	-	[M]	3.69E-03	3.04
	bhn_I0834	-	HD family phosphohydrolase	-	[K]	9.65E-05	3.04
	bhn_I1967	-	mannose-6-phosphate isomerase	GH1	[G]	2.31E+06	3.04
	bhn_I2236	*typA*	GTP-binding protein	-	[T]	1.47E-05	3.04
	bhn_I0257	-	hypothetical protein	-	-	9.03E-05	3.03
	bhn_I1427	*tsf*	elongation factor Ts	-	[J]	5.76E-04	3.03
	bhn_I1147	-	CoA-binding domain-containing protein	-	[R]	5.58E-04	3.02
	bhn_I1626	*xth*	endonuclease/exonuclease/phosphatase	-	[P]	1.74E-04	3.02
	bhn_II058	-	hypothetical protein	-	-	9.45E+06	3.02
	bhn_I0095	*iolC*	5-dehydro-2-deoxygluconokinase	-	[G]	2.33E-04	3.02
	bhn_I1693	*trmH*	RNA methyltransferase	-	[J]	3.56E-04	3.02
	bhn_I0536	*lyc25E*	lysozyme	-	[M]	3.60E+09	3.01
	bhn_I0548	*pgm*	phosphoglucomutase	-	[G]	2.44E+06	3.01
	bhn_I0402	*plsC1*	acylphosphate glycerol 3-phosphate acyltransferase	-	[I]	2.83E-04	3.01
	bhn_I1492	*vanY*	D-alanyl-D-alanine carboxypeptidase	-	[M]	1.98E-04	3.00
	bhn_I1575	-	hypothetical protein	-	-	1.59E-03	3.00
	bhn_I1866	-	HTH/SIS domain-containing protein	-	[K]	1.57E-03	3.00
	bhn_III75	-	β-lactamase	-	[R]	4.56E+29	3.00
	bhn_I2080	*map*	methionine aminopeptidase	-	[J]	3.19E-05	2.98
	bhn_I0144	-	GGDEF domain-containing protein	-	[T]	6.92E-05	2.98
	bhn_I2588	-	hypothetical protein	-	-	1.20E-04	2.98

	bhn_I2032	-	ABC transporter ATP-binding protein	-	[V]	2.29E-04	2.97
	bhn_I2346	*kpsF*	sugar phosphate isomerase	-	[M]	4.11E-05	2.97
	bhn_I2667	*phnE*	phosphonate ABC transporter permease protein	-	[P]	4.02E-03	2.97
	bhn_I2724	-	hypothetical protein	-	-	1.86E-03	2.96
	bhn_I0482	*dred*	transposase	-	[L]	4.91E-04	2.95
	bhn_I1756	*msbA*	ABC transporter ATP-binding protein	-	[V]	2.03E-04	2.95
	bhn_I2400	*cyt*	nicotinamidenucleotide adenylyltransferase	-	[H]	1.35E-03	2.95
	bhn_I0260	-	membrane protein	-	-	3.58E-04	2.95
	bhn_I2700	-	AsnC family transcriptional regulator	-	[K]	9.16E-04	2.95
	bhn_I1015	*phn*	transcriptional regulator	-	[K]	4.09E-04	2.94
	bhn_I1523	*smpB*	single-stranded DNA-binding protein	-	[O]	1.48E-04	2.94
	bhn_I2260	*pstC*	sugar ABC transporter permease	-	[G]	4.22E-03	2.94
	bhn_I2745	*ccA*	tRNA nucleotidyltransferase	-	[J]	1.28E-04	2.94
	bhn_I2058	*rpoA*	DNAdirected RNA polymerase alpha subunit	-	[K]	4.91E-04	2.93
	bhn_I0781	-	hypothetical protein	-	-	5.46E-04	2.92
	bhn_I1186	-	hypothetical protein	-	-	6.63E-04	2.92
	bhn_I2131	-	sugar ABC transporter permease protein	-	[G]	3.24E-03	2.92
	bhn_I2613	*potA*	sugar ABC transporter ATP-binding protein	-	[G]	1.33E-03	2.92
	bhn_I2208	*betI*	TetR family transcriptional regulator	-	[K]	2.36E-03	2.92
	bhn_I0453	*galU*	CTP-phosphocholine cytidylyltransferase	-	[M]	4.09E+06	2.91
	bhn_I0721	-	DHH domain-containing protein	-	[P]	1.18E-04	2.91
	bhn_I1082	*mutt*	NUDIX hydrolase	-	[L]	2.15E-03	2.91
	bhn_I0035	*rluA*	pseudouridine synthase	-	[J]	1.64E-03	2.90
	bhn_I0861	*pgsA*	CDPdiglyceride synthetase	-	[I]	1.74E-04	2.90
	bhn_I0114	*hprK*	HPr(Ser) kinase/phosphatase	-	[T]	2.34E-03	2.89
	bhn_I1355	*lspA*	lipoprotein signal peptidase	-	[M]	1.75E-03	2.89
	bhn_I2732	-	flavodoxin	-	[C]	3.29E-05	2.88
	bhn_I0929	-	WYL domain-containing protein	-	[K]	3.53E-03	2.87
	bhn_I1723	-	hypothetical protein	-	-	4.58E-03	2.87
	bhn_I2744	-	PEGA domain-containing protein	-	-	1.50E-04	2.87
	bhn_I0494	*ytvI*	polysaccharide biosynthesis protein	-	-	1.03E-04	2.86
	bhn_I0955	*matE*	multidrug transporter	-	[V]	1.67E-04	2.86
	bhn_I0345	-	radical SAM protein	-	[H]	4.28E-04	2.85
	bhn_I0573	*thrS*	threonyl-tRNA synthetase	-	[J]	1.85E-05	2.85
	bhn_I1158	*recO*	DNA recombination protein	-	[L]	2.97E-03	2.85
	bhn_I1195	-	hypothetical protein	-	-	4.98E-05	2.84

	bhn_I0399	-	hypothetical protein	-	-	4.30E-03	2.84
	bhn_I0547	-	glycosyl transferase GT2	GT2	[R]	5.91E-04	2.84
	bhn_I2251	*etfA*	electron transfer flavoprotein alpha subunit	-	[C]	7.01E-03	2.84
	bhn_I0565	*gcp*	O-sialoglycoprotein endopeptidase	-	[O]	1.20E-03	2.83
	bhn_I0367	*porD*	$4Fe_4S$ di-cluster domain-containing protein	-	-	2.84E-03	2.82
	bhn_I1140	*aguA*	agmatine deiminase	-	[E]	1.26E-04	2.82
	bhn_I1224	*recF*	ATPase AAA family	-	[R]	1.23E-03	2.82
	bhn_I0492	-	hypothetical protein	-	-	3.96E-04	2.81
	bhn_I0809	-	iron ABC transporter permease protein	-	[P]	4.06E-03	2.81
	bhn_I0894	*pgm*	phosphoglucomutase/phosphomannomutase	-	[G]	1.59E-03	2.81
	bhn_I1386	*cwlD*	N-acetylmuramoyl-L-alanine amidase	-	[M]	7.75E-03	2.81
	bhn_I1547	*polA*	DNA polymerase I	-	[L]	4.84E-03	2.81
	bhn_I1877	*pgk*	phosphoglycerate kinase	-	[G]	1.93E-03	2.81
	bhn_I2008	*pspE*	rhodanese-like protein	-	[P]	9.60E-04	2.81
	bhn_II006	-	DNA/RNA non-specific endonuclease	-	-	2.89E+06	2.81
	bhn_II041	-	hypothetical protein	-	-	7.26E+06	2.81
	bhn_I0078	*gmk*	guanylate kinase	-	[F]	5.63E-03	2.79
	bhn_III50	-	response regulator domain-containing protein	-	[T]	1.09E-03	2.78
	bhn_I2595	-	aldose 1-epimerase family protein	-	[G]	8.01E-03	2.78
	bhn_I2038	-	auxin efflux carrier family transporter	-	[R]	4.30E-03	2.77
	bhn_I1272	*ftsY*	cell division protein	-	[U]	3.73E-03	2.77
	bhn_I1645	*ppk*	polyphosphate kinase	-	[P]	2.35E-04	2.77
	bhn_I0484	-	hypothetical protein	-	-	2.63E-03	2.76
	bhn_I2434	*est4B*	polysaccharide deacetylase	CE4	[G]	8.83E-03	2.76
	bhn_I2456	*pseI*	pseudaminic acid synthase	-	[M]	4.41E-03	2.76
	bhn_I0821	*udk*	phosphoribulokinase/uridine kinase	-	[F]	2.94E-03	2.75
	bhn_I1794	-	hypothetical protein	-	-	6.53E-03	2.74
	bhn_I1247	-	hypothetical protein	-	-	3.78E-04	2.73
	bhn_I1504	-	hypothetical protein	-	-	1.48E-03	2.73
	bhn_I2201	*tig*	trigger factor	-	[O]	6.00E-03	2.73
	bhn_I1287	*urtB*	ABC transporter permease protein	-	[R]	4.80E-03	2.72
	bhn_I1314	*deoD1*	purinenucleoside phosphorylase	-	[F]	9.23E-03	2.72
	bhn_II124	-	hypothetical protein	-	-	6.97E-04	2.72
	bhn_I0675	-	hypothetical protein	-	-	6.89E-03	2.71
	bhn_II015	-	hypothetical protein	-	-	2.27E-03	2.71
	bhn_I0579	*xyn10B*	endo-1,4-β-xylanase	GH10	[G]	6.33E-04	2.70

	bhn_I1190	*met*	restriction-modification system DNA methyltransferase	-	[V]	5.49E-04	2.70
	bhn_I1733	*msbA*	ABC transporter ATP-binding protein	-	[C]	1.47E-02	2.70
	bhn_I2375	-	hypothetical protein	-	-	2.41E-02	2.70
	bhn_I2424	*gltA*	glutamate synthase	-	[E]	2.81E-03	2.70
	bhn_I2782	-	hypothetical protein	-	-	7.81E-03	2.69
	bhn_I0142	*proA*	glutamate-5-semialdehyde dehydrogenase	-	[E]	3.92E+06	2.69
	bhn_I1502	-	hypothetical protein	-	-	7.81E-03	2.69
	bhn_I1997	*matE*	MatE efflux family protein/ATPase	-	[V]	5.61E-03	2.69
	bhn_I2332	*pgl28B*	polygalacturonase	GH28	[M]	1.31E-02	2.69
	bhn_II118	-	hypothetical protein	-	-	2.14E-04	2.69
	bhn_I0795	*comE*	comEA protein	-	[L]	8.72E-03	2.69
	bhn_I0350	-	hypothetical protein	-	-	4.28E-04	2.68
	bhn_I1710	*matE*	multidrug transporter	-	[V]	1.38E-02	2.68
	bhn_I1961	*cel5C*	endo-1,4-β-glucanase	GH5/CBM2	[G]	1.35E-02	2.68
	bhn_I2715	-	radical SAM domain-containing protein	-	[L]	2.45E-04	2.68
	bhn_I1675	-	GGDEF/EAL domain-containing protein	-	[T]	2.09E-03	2.67
	bhn_I1182	-	hypothetical protein	-	[O]	1.70E-02	2.66
	bhn_I1394	-	acyltransferase	-	[I]	1.99E-02	2.66
	bhn_I1616	*trpB*	tryptophan synthase beta subunit	-	[E]	6.16E-03	2.66
	bhn_I2630	*rluA*	pseudouridine synthase	-	[J]	3.65E-03	2.64
	bhn_I1022	-	GGDEF/EAL domain-containing protein	-	[T]	2.28E-04	2.64
	bhn_I0052	*glgA*	glycosyl transferase GT4 family	GT4	[M]	1.57E-04	2.62
	bhn_I0447	-	CDP-glycerol glycerophosphotransferase	-	[M]	3.34E-03	2.62
	bhn_I1845	*nikD*	nickel/peptide ABC transporter ATP-binding protein	-	[E]	1.08E-02	2.62
	bhn_I1930	*uvrA*	excinuclease ABC subunit A	-	[L]	1.39E-03	2.61
	bhn_I1061	-	WavE family lipopolysaccharide synthesis protein	-	-	1.28E-02	2.61
	bhn_I1385	*dnaJ*	molecular chaperone	-	[O]	2.96E-02	2.61
	bhn_I1393	-	RND family transporter	-	[R]	7.28E-04	2.60
	bhn_I1738	*trpD*	anthranilate phosphoribosyltransferase	-	[E]	3.10E-02	2.58
	bhn_I2430	*argF*	ornithine carbamoyltransferase	-	[E]	1.43E-03	2.58
	bhn_II120	-	hypothetical protein	-	-	4.97E-03	2.58
	bhn_I2721	-	sugar ABC transporter substrate-binding protein	-	[G]	1.19E-02	2.56
	bhn_I1060	*amy13B*	α-amylase	GH13	[G]	2.71E-02	2.56
	bhn_I1420	-	FeS oxidoreductase	-	[C]	2.38E-02	2.56
	bhn_I1142	-	hypothetical protein	-	-	3.82E-02	2.56
	bhn_I0423	*isdE*	sugar ABC transporter substrate-binding protein	-	[G]	1.29E-03	2.54

	bhn_I2323	-	hypothetical protein	-	-	1.68E-03	2.54
	bhn_I1729	*nikD*	nickel/peptide ABC transporter ATP-binding protein	-	[E]	3.11E-03	2.54
	bhn_I2492	-	hypothetical protein	-	-	4.67E-04	2.54
	bhn_I2352	*htrA*	serine protease	-	[O]	4.07E-03	2.53
	bhn_I0672	*flgK*	flagellar hookassociated protein	-	[N]	5.08E-03	2.52
	bhn_I1754	*xylG*	ABC transporter ATP-binding protein	-	[R]	1.59E-03	2.52
	bhn_I0554	-	RelA/SpoT domain-containing protein	-	[S]	7.25E-02	2.52
	bhn_I1933	*trpD*	anthranilate phosphoribosyltransferase	-	[E]	5.88E-03	2.50
	bhn_I1489	*recQ*	DEAD/DEAH box helicase	-	[L]	1.04E-02	2.49
	bhn_I0217	*hsdS*	type I restrictionmodification system S subunit	-	[V]	4.46E-03	2.49
	bhn_I2228	-	hypothetical protein	-	-	4.17E-02	2.48
	bhn_I0310	-	hypothetical protein	-	-	8.20E-02	2.47
	bhn_I1354	*aroB*	3-dehydroquinate synthase	-	E]	1.24E-02	2.46
	bhn_I2290	*gh31C*	glycoside hydrolase family 31	GH31	[G]	4.13E-02	2.44
	bhn_I2395	*matE*	multidrug transporter	-	[V]	1.21E-02	2.44
	bhn_I2730	-	ATPase AAA	-	[O]	4.59E-03	2.43
	bhn_I1706	-	hypothetical protein	-	-	4.47E-02	2.43
	bhn_I1956	-	sugar ABC transporter substrate-binding protein	-	[G]	8.33E-03	2.42
	bhn_I0644	*amy13B*	α-amylase	GH13	[G]	3.49E-03	2.42
	bhn_I1301	-	Fe-S oxidoreductase	-	[C]	2.11E-03	2.42
	bhn_I0762	-	hypothetical protein	-	-	2.05E-04	2.41
	bhn_I2551	*bga2A*	β-galactosidase	GH2	[G]	4.09E-04	2.41
	bhn_I1839	-	GGDEF/EAL domain-containing protein	-	[T]	4.16E-03	2.39
	bhn_I2383	*cobD*	cobalamin biosynthesis protein	-	[H]	1.49E-02	2.39
	bhn_I0487	-	hypothetical protein	-	-	6.06E-03	2.38
	bhn_I0435	-	hypothetical protein	-	-	2.91E-02	2.37
	bhn_I0446	-	hypothetical protein	-	-	5.28E-03	2.37
	bhn_II070	*dnaX*	DNA polymerase III subunit gamma/tau	-	[L]	2.47E-05	2.37
	bhn_I0199	*htrA*	serine protease HtrA	-	[O]	5.29E-03	2.36
	bhn_I1312	*hvy*	heavy metal translocating P-type ATPase	-	[P]	2.63E-03	2.36
	bhn_I2365	*flgK*	flagellar hook-associated protein	-	[N]	6.89E-03	2.34
	bhn_I0542	-	polysaccharide ABC transporter permease	-	[G]	3.54E-02	2.34
	bhn_II009	*parA*	partitioning protein	-	[D]	1.63E-05	2.33
	bhn_I0216	-	cell wall hydrolase	-	[M]	9.39E-03	2.32
	bhn_I0566	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	-	[I]	1.98E-02	2.32
	bhn_I0800	-	LPxTG motif-containing cell surface protein	-	-	1.17E-03	2.32

	bhn_I2138	*leuA*	2-isopropylmalate synthase	-	[E]	3.21E-02	2.32
	bhn_I1535	-	DEAD/DEAH box helicase	-	[L]	5.31E-03	2.31
	bhn_II069	-	DNA polymerase III delta subunit	-	[L]	1.89E-03	2.31
	bhn_I1079	*rhaB*	rhamnulokinase	-	[G]	1.63E-02	2.30
	bhn_I1986	-	histidine kinase	-	[T]	2.22E-02	2.30
	bhn_I2426	*asnB*	exosortase A system-associated amidotransferase	-	[E]	1.49E-02	2.27
	bhn_I1216	*hsdS*	type I restriction-modification system S subunit	-	[V]	8.63E-03	2.27
	bhn_I0113	*uvrC*	excinuclease ABC subunit C	-	[L]	6.55E-03	2.27
	bhn_I2192	*cobO*	cob(I)yrinic acid a,c-diamide adenosyltransferase	-	[H]	3.10E-02	2.25
	bhn_I0771	-	hypothetical protein	-	-	9.21E-02	2.24
	bhn_I1597	*atpB*	ATP synthase F0 A subunit	-	[C]	4.72E-02	2.23
	bhn_I2314	-	hypothetical protein	-	-	9.16E-03	2.23
	bhn_I2741	*tilS*	tRNA(Ile)-lysidine synthetase	-	[D]	1.22E-02	2.23
	bhn_I0881	-	hypothetical protein	-	-	5.10E-02	2.22
	bhn_I0676	-	NAD-dependent aldehyde dehydrogenase	-	[C]	1.97E-02	2.21
	bhn_I0732	-	GGDEF/CHASE domain-containing protein	-	[T]	1.29E-02	2.19
	bhn_I1701	-	GNAT family acetyltransferase	-	[J]	8.76E-02	2.18
	bhn_I2299	-	glycosyl transferase GT4 family	GT4	[S]	3.50E-02	2.18
	bhn_I0214	-	RNA methyltransferase TrmH family	-	[J]	6.05E-02	2.16
	bhn_I1087	*fusA*	translation elongation factor G	-	[J]	1.51E-02	2.15
	bhn_I2067	-	two component system histidine kinase	-	[T]	1.80E-02	2.15
	bhn_I1403	*phoR*	two component system histidine kinase	-	-	2.61E-03	2.15
	bhn_II038	-	hypothetical protein	-	-	1.40E-05	2.14
	bhn_I1945	*sleB*	cell wall hydrolase	-	[S]	6.40E-02	2.14
	bhn_III69	-	hypothetical protein	-	-	2.46E-02	2.14
	bhn_I1127	*msbA*	ABC transporter ATP-binding/permease protein	-	[V]	1.68E-02	2.14
	bhn_I1741	*mraW*	S-adenosyl-methyltransferase	-	[M]	5.14E-02	2.13
	bhn_I0194	*lyc25C*	lysozyme	GH25	[G]	2.32E-02	2.13
	bhn_I1261	-	hypothetical protein	-	-	3.69E-02	2.13
	bhn_I2694	*phoR*	two component system histidine kinase	-	[T]	4.17E-02	2.12
	bhn_I1837	*hisS*	histidyl-tRNA synthetase	-	[J]	5.40E-02	2.11
	bhn_I1969	-	glycoside hydrolase GH130 family	GH130/GH43	[G]	2.15E-02	2.11
	bhn_I0745	*pyrF*	orotidine 5'-phosphate decarboxylase	-	[F]	9.14E-02	2.10
	bhn_I1734	*ftsZ*	cell division protein	-	[D]	6.23E-02	2.09
	bhn_I0275	*hvy*	heavy metal translocating P-type ATPase	-	[P]	2.18E-02	2.09
	bhn_I1236	*nifB*	peptidyl-prolyl cis-trans isomerase FKBP-type	-	[O]	6.62E-02	2.09

	bhn_II087	*cpf1*	CRISPR-associated protein	-	-	2.87E-05	2.09
	bhn_I2221	*gh31C*	glycoside hydrolase family 31	GH31	[G]	6.89E-03	2.07
	bhn_I1850	-	HD family phosphohydrolase	-	[T]	9.56E-02	2.04
	bhn_I1417	*echE*	Ech hydrogenase subunit	-	[C]	7.71E-02	2.04
	bhn_I1998	-	hypothetical protein	-	-	6.45E-02	2.03
	bhn_I0627	-	hypothetical protein	-	-	7.95E-02	2.03
	bhn_I0054	-	hypothetical protein	-	-	2.08E-02	2.02
	bhn_I2369	-	hypothetical protein	-	-	2.71E-02	2.02
	bhn_I1737	*murD*	UDP-N-acetylmuramoylalanine-D-glutamate ligase	-	[M]	5.04E-02	2.02
	bhn_I0004	*recF*	DNA replication protein	-	[L]	0.00E+00	N/A
	bhn_I0050	-	hydrolase TatD	-	[L]	0.00E+00	N/A
	bhn_I0276	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0302	-	MFS transporter	-	[G]	0.00E+00	N/A
	bhn_I0314	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0359	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0384	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0429	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0539	*rfbA*	glucose-1-phosphate thymidylyltransferase	-	[M]	0.00E+00	N/A
	bhn_I0571	-	HTH domain-containing protein	-	[K]	0.00E+00	N/A
	bhn_I0623	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0626	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0736	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0816	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I0947	-	transposase	-	-	0.00E+00	N/A
	bhn_I1063	*feoA*	iron transporter	-	[P]	0.00E+00	N/A
	bhn_I1187	*xerD*	tyrosine recombinase	-	[L]	0.00E+00	N/A
	bhn_I1271	*sMC*	chromosome segregation protein	-	[D]	0.00E+00	N/A
	bhn_I1309	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I1328	*iorA*	indolepyruvate ferredoxin oxidoreductase	-	[C]	0.00E+00	N/A
	bhn_I1367	*yhbY*	RNA-binding protein	-	[J]	0.00E+00	N/A
	bhn_I1368	*cgtA*	GTPase	-	[R]	0.00E+00	N/A
	bhn_I1379	*hisG*	ATP phosphoribosyltransferase	-	[E]	0.00E+00	N/A
	bhn_I1480	*apt*	adenine phosphoribosyltransferase	-	[F]	0.00E+00	N/A
	bhn_I1490	*yajC*	preprotein translocase	-	[U]	0.00E+00	N/A
	bhn_I1510	*gvpN*	phosphate starvation protein	-	[T]	0.00E+00	N/A
	bhn_I1549	*gph*	sugar/sodium symporter	-	[G]	0.00E+00	N/A

	bhn_I1603	*apbE*	thiamin biosynthesis lipoprotein	-	[H]	0.00E+00	N/A
	bhn_I1681	*argC*	N-acetyl-gamma-glutamyl-phosphate reductase	-	[E]	0.00E+00	N/A
	bhn_I1774	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I1931	-	N-(5'-phosphoribosyl)anthranilate isomerase	-	[E]	0.00E+00	N/A
	bhn_I1952	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I1962	*carB*	carbamoyl-phosphate synthase large subunit	-	[E]	0.00E+00	N/A
	bhn_I2031	-	ABC transporter permease	-	[V]	0.00E+00	N/A
	bhn_I2061	*rpsM*	30*S* ribosomal protein S13	-	[J]	0.00E+00	N/A
	bhn_I2084	*rpmD*	50*S* ribosomal protein L30	-	[J]	0.00E+00	N/A
	bhn_I2089	*rpsN*	30*S* ribosomal protein S14	-	[J]	0.00E+00	N/A
	bhn_I2093	*rpsQ*	30*S* ribosomal protein S17	-	[J]	0.00E+00	N/A
	bhn_I2094	*rpmC*	50*S* ribosomal protein L29	-	[J]	0.00E+00	N/A
	bhn_I2097	*rplV*	50*S* ribosomal protein L22	-	[J]	0.00E+00	N/A
	bhn_I2103	*rpsJ*	30*S* ribosomal protein S10	-	[J]	0.00E+00	N/A
	bhn_I2169	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I2417	-	glycosyl transferase GT2 family	GT2	[M]	0.00E+00	N/A
	bhn_I2431	-	glycosyl transferase GT4 family	GT4	[M]	0.00E+00	N/A
	bhn_I2614	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_I2780	*phnS*	ABC transporter substrate-binding protein	-	[P]	0.00E+00	N/A
	bhn_I2804	*rpmH*	ribosomal protein L34	-	[J]	0.00E+00	N/A
	bhn_II023	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II025	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II026	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II031	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II032	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II033	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II034	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II035	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II039	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II045	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II047	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II054	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II055	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II056	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II057	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II066	*narQ*	nitrate/nitrite sensor protein	-	-	0.00E+00	N/A

	bhn_II079	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II099	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II101	-	HTH domain-containing protein	-	[K]	0.00E+00	N/A
	bhn_II102	-	RTX toxins and related Ca^{2+}-binding protein	-	-	0.00E+00	N/A
	bhn_II103	-	lytic transglycosylase	-	-	0.00E+00	N/A
	bhn_II104	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II105	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II125	-	phosphoadenosine phosphosulfate reductase	-	[E]	0.00E+00	N/A
	bhn_II126	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II127	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II128	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II129	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II130	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II131	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II132	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II133	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II134	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II137	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II138	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II140	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_II144	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_III002	-	radical SAM domain-containing protein	-	[R]	0.00E+00	N/A
	bhn_III006	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_III009	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_III11	-	PAP2 family protein	-	[I]	0.00E+00	N/A
	bhn_III15	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_III20	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_III27	-	hypothetical protein	-	-	0.00E+00	N/A
	bhn_III28	*hypA*	hydrogenase nickel insertion protein	-	[R]	0.00E+00	N/A
	bhn_III31	*repB*	replication initiation protein	-	[L]	0.00E+00	N/A
	bhn_III39	*fabD*	malonyl CoA-acyl carrier protein transacylase	-	[I]	0.00E+00	N/A
	bhn_III40	*fabG*	3-oxoacyl-(acyl-carrier-protein) reductase	-	[I]	0.00E+00	N/A
	bhn_III41	*fabF*	3-oxoacyl-(acyl-carrier-protein) synthase	-	[I]	0.00E+00	N/A
	bhn_III43	*fabZ*	β-hydroxyacyl-(acyl-carrier-protein) dehydratase	-	[I]	0.00E+00	N/A
	bhn_III44	*accC*	acetyl-CoA carboxylase biotin carboxylase	-	[I]	0.00E+00	N/A
	bhn_III45	*accD*	acetyl-CoA carboxylase carboxyl transferase beta subunit	-	[I]	0.00E+00	N/A

	bhn_III54	-	GIY-YIG domain-containing protein	-	[L]	0.00E+00	N/A
	bhn_III56	-	radical SAM domain-containing protein	-	[R]	1.63E+07	N/A
	bhn_III58	-	hypothetical protein	-	-	8.53E+185	N/A
	bhn_III61	*corA*	magnesium transporter	-	[P]	5.51E-03	N/A
	bhn_III72	*est4A*	polysaccharide deacetylase	CE4	[G]	0.00E+00	N/A
	bhn_III73	-	histidine phosphatase superfamily protein	-	[G]	1.72E+06	N/A
	bhn_III74	-	TauE family protein	-	[R]	0.00E+00	N/A
	bhn_III84	-	AAA domain-containing protein	-	-	5.18E+29	N/A
MB2003 co-culture growth on pectin	bhn_I2409	-	DNA polymerase	-	[L]	0.00E+00	8.26
	bhn_I0091	*hpaA*	AraC family transcriptional regulator	-	[K]	0.00E+00	6.27
	bhn_I2657	-	Rrf2 family transcriptional regulator	-	[K]	0.00E+00	6.14
	bhn_I2752	-	hypothetical protein	-	-	0.00E+00	6.10
	bhn_I2703	-	hypothetical protein	-	-	0.00E+00	5.90
	bhn_I0828	-	hypothetical protein	-	-	7.13E-79	5.86
	bhn_I1084	-	hypothetical protein	-	-	0.00E+00	5.79
	bhn_III22	-	hypothetical protein	-	-	0.00E+00	5.73
	bhn_I2353	*fliC*	flagellin	-	[N]	9.27E-55	5.61
	bhn_I2777	-	ABC transporter substrate-binding protein	-	[P]	0.00E+00	5.58
	bhn_I2645	-	hypothetical protein	-	-	0.00E+00	5.58
	bhn_I0777	-	hypothetical protein	-	-	0.00E+00	5.48
	bhn_I1532	*arf51C*	α-L-arabinofuranosidase	GH51/CBM4	[G]	0.00E+00	5.05
	bhn_I1014	-	methyl-accepting chemotaxis protein	-	[N]	0.00E+00	4.97
	bhn_I1233	-	α/β hydrolase family protein	-	[R]	0.00E+00	4.96
	bhn_I2526	-	hypothetical protein	-	-	0.00E+00	4.88
	bhn_I2775	*cysA*	ABC transporter ATP-binding protein	-	[E]	0.00E+00	4.87
	bhn_I2343	-	hypothetical protein	-	-	0.00E+00	4.84
	bhn_I2642	-	hypothetical protein	-	-	0.00E+00	4.81
	bhn_I2222	-	sugar ABC transporter substrate-binding protein	-	[G]	1.89E-300	4.79
	bhn_I2776	*modB*	ABC transporter permease	-	[P]	0.00E+00	4.53
	bhn_I0036	-	hypothetical protein	-	-	5.56E-59	4.53
	bhn_I0280	-	hypothetical protein	-	-	0.00E+00	4.26
	bhn_I1641	-	hypothetical protein	-	-	2.99E-318	4.21
	bhn_I2104	-	SAM-dependent methyltransferase	-	[R]	1.82E-290	4.16
	bhn_III29	-	hypothetical protein	-	-	2.32E-125	4.04
	bhn_I1974	*phn*	LysR family transcriptional regulator	-	[K]	4.62E-214	3.86

	bhn_I2789	-	hypothetical protein	-	[V]	4.55E-284	3.85
	bhn_I2554	*sleB*	cell wall hydrolase	-	[M]	5.59E-149	3.79
	bhn_I0851	-	QueT family transporter	-	[S]	8.05E-194	3.69
	bhn_I0111	*pyrG*	CTP synthase	-	[F]	2.32E-125	3.67
	bhn_I2041	*bioC*	methyltransferase	-	[H]	2.65E-50	3.67
	bhn_I0340	*fabG*	short chain dehydrogenase/reductase	-	[I]	2.79E-193	3.66
	bhn_I1351	*pilT*	twitching motility protein	-	[N]	2.27E-121	3.56
	bhn_I2150	-	ATPase	-	[R]	7.11E-109	3.52
	bhn_I2553	-	hypothetical protein	-	-	4.47E-147	3.52
	bhn_I2354	*fliC*	flagellin	-	[N]	2.67E-117	3.52
	bhn_I1458	*flgF*	flagellar basal-body rod protein	-	[N]	2.52E-48	3.48
	bhn_I1244	*fucO*	iron-containing alcohol dehydrogenase	-	[C]	4.02E-84	3.46
	bhn_I2763	-	transglycosylase associated protein	-	[S]	2.00E-79	3.41
	bhn_I2560	-	MecA family protein	-	[O]	1.57E-123	3.40
	bhn_I2503	-	CAAX amino terminal protease family protein	-	[R]	6.70E-85	3.36
	bhn_I0847	*cel9B*	cellodextrinase	GH9/CelD	[G]	1.74E-31	3.35
	bhn_I1097	*rho*	transcription termination factor	-	[K]	1.34E-68	3.32
	bhn_I1242	-	hypothetical protein	-	-	7.04E-85	3.27
	bhn_I1457	-	ion transport protein	-	[P]	1.92E-97	3.25
	bhn_I2410	-	hypothetical protein	-	-	2.94E-89	3.19
	bhn_III79	-	ATP-binding protein	-	[R]	6.17E-06	3.09
	bhn_I0366	-	cAMP-binding protein	-	[T]	4.60E-54	3.07
	bhn_II052	-	resolvase	-	-	1.40E-17	2.95
	bhn_I2335	-	hypothetical protein	-	-	1.14E-53	2.90
	bhn_III30	-	ATPase AAA	-	[R]	1.62E-10	2.88
	bhn_II110	-	hypothetical protein	-	-	2.28E-43	2.80
	bhn_I2561	*nifJ*	pyruvate-flavodoxin oxidoreductase	-	[C]	2.30E-05	2.76
	bhn_I2661	*rimI*	GNAT family acetyltransferase	-	[K]	1.15E-33	2.76
	bhn_I1263	*ackA*	acetate kinase	-	[C]	1.12E-30	2.72
	bhn_II119	-	phosphocarrier HPr family	-	[G]	2.23E-20	2.35
	bhn_I0605	-	hypothetical protein	-	-	4.98E-09	2.30
	bhn_I1192	-	GGDEF/EAL domain-containing protein	-	[T]	4.51E-16	2.22
	bhn_I0131	-	rRNA methyltransferase	-	[J]	7.17E-17	2.17
	bhn_I1101	-	GGDEF domain-containing protein	-	[T]	1.07E-11	2.12
	bhn_I0327	-	hypothetical protein	-	-	1.53E-13	2.03

Genes were deemed "significantly expressed" if all three thresholds were met: FDR calculated using the Benjamini-Hochberg procedure *Q*-values < 0.05, KW analysis of variance with $p < 0.05$, and ≥ 2-fold $\log_2$ difference in expression. CAZy designations have been assigned to genes containing the relevant domains. COG classifications: [C], Energy production and conversion; [D], Cell cycle control, cell division, chromosome partitioning; [E], Amino acid transport and metabolism; [F], Nucleotide transport and metabolism; [G], Carbohydrate transport and metabolism; [H], Coenzyme transport and metabolism; [I], Lipid transport and metabolism; [J], Translation, ribosomal structure and biogenesis; [K], Transcription; [L], Replication, recombination and repair; [M], Cell wall-membrane-envelope biogenesis; [N], Cell motility; [O], Posttranslational modification, protein turnover, chaperones; [P], Inorganic ion transport and metabolism; [Q], Secondary metabolites biosynthesis, transport and catabolism; [R], General function prediction only; [S], Function unknown; [T], Signal transduction mechanisms; [U], Intracellular trafficking, secretion, and vesicular transport; [V], Defence mechanisms.

References

Abbott, D. W., Ficko-Blean, E., van Bueren, A. L., Rogowski, A., Cartmell, A., Coutinho, P. M., Henrissat, B., Gilbert, H. J., and Boraston, A. B. (2009). Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules. *Biochemistry,* 48 (*43*), 10395-10404.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. Selected Papers of Hirotugu Akaike. pp. 199-213. New York. Springer.

Alexandratos, N., and Bruinsma, J. (2012). World agriculture towards 2030/2050: the 2012 revision. *ESA Working Paper No. 12-03*. Rome: FAO.

Alghisi, P., and Favaron, F. (1995). Pectin-degrading enzymes and plant-parasite interactions. *European Journal of Plant Pathology,* 101 (*4*), 365-375.

Allen, M. S. (1996). Physical constraints on voluntary intake of forages by ruminants. *Journal of Animal Science,* 74 (*12*), 3063-3075.

Altermann, E. (2012). Tracing lifestyle adaptation in prokaryotic genomes. *Frontiers in Microbiology,* 3 (*48*), 1-17.

Altermann, E., and Klaenhammer, T. R. (2003). GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *OMICS A Journal of Integrative Biology,* 7 (*2*), 161-169.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology,* 215 (*3*), 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25 (*17*), 3389-3402.

Alvira, P., Tomás-Pejó, E., Ballesteros, M., and Negro, M. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresource Technology,* 101 (*13*), 4851-4861.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology,* 11 (*10*), R106.

Anderson, K., and Blair, B. (1996). Regulation of the cellulolytic activity of *Eubacterium cellulosolvens* 5494: a review. *SAAS Bulletin, Biochemistry and Biotechnology,* 9 (*1*), 57-62.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology,* 26 (*1*), 32-46.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. from Babraham Bioinformatics. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., and Croning, M. D. R. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Research,* 29 (*1*), 37-40.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., and Eppig, J. T. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics,* 25 (*1*), 25-29.

Aspinall, G. O. (1980). Chemistry of cell wall polysaccharides. In J. Preiss (Ed.), The biochemistry of plants: A comprehensive treatise. Vol. 3 Carbohydrates: Structure and function, pp. 473-501. New York, USA. Acadaemic Press.

Atalla, R. H., Hackney, J. M., Uhlin, I., and Thompson, N. S. (1993). Hemicelluloses as structure regulators in the aggregation of native cellulose. *International Journal of Biological Macromolecules,* 15 (*2*), 109-112.

Attwood, G. T., Klieve, A. V., Ouwerkerk, D., and Patel, B. K. C. (1998). Ammonia-hyperproducing bacteria from New Zealand ruminants. *Applied and Environmental Microbiology,* 64 (*5*), 1796.

Attwood, G. T., and Reilly, K. (1996). Characterization of proteolytic activities of rumen bacterial isolates from forage-fed cattle. *Journal of Applied Bacteriology,* 81 (*5*), 545-552.

Attwood, G. T., and Reilly, K. (1998). Detection of *Clostridium proteoclasticum* and closely related strains in the rumen by competitive PCR. *Applied and Environmental Microbiology,* 64 (*3*), 907-913.

Attwood, G. T., Reilly, K., and Patel, B. K. C. (1996). *Clostridium proteoclasticum* sp. nov., a novel proteolytic bacterium from the bovine rumen. *International Journal of Systematic Bacteriology,* 46 (*3*), 753-758.

Avguštin, G., Wallace, R. J., and 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.

Avguštin, G., Wright, F., and Flint, H. J. (1994). Genetic diversity and phylogenetic relationships among strains of *Prevotella* (*Bacteroides*) *ruminicola* from the rumen. *International Journal of Systematic Bacteriology,* 44 (*2*), 246-255.

Ayers, W. A. (1958). Nutrition and physiology of *Ruminococcus flavefaciens*. *Journal of Bacteriology,* 76 (*5*), 504.

Bacic, A., Harris, P. J., and Stone, B. A. (1988). Structure and function of plant cell walls. In J. Preiss (Ed.), Biochemistry of Plants. Vol. 14, pp. 297-371. New York, USA. Academic Press.

Balch, D., Balch, C., and Rowland, S. (1954). The influence of the method of determination of lignin on the lignin ratio technique for digestibility in the cow. *Journal of the Science of Food and Agriculture,* 5 (*12*), 584-588.

Bansal, A. K., and Meyer, T. E. (2002). Evolutionary analysis by whole-genome comparisons. *Journal of Bacteriology,* 184 (*8*), 2260-2272.

Bartholomew, J. W., and Mittwer, T. (1952). The Gram stain. *Bacteriological Reviews,* 16 (*1*), 1-29.

Bayer, E. A., Lamed, R., White, B. A., and Flint, H. J. (2008). From cellulosomes to cellulosomics. *The Chemical Record,* 8 (*6*), 364-377.

Bayer, E. A., Shoham, Y., and Lamed, R. (2013). Lignocellulose-decomposing bacteria and their enzyme systems. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt and F. Thompson (Eds.), The Prokaryotes: Prokaryotic physiology and biochemistry. 4th ed., pp. 215-266. Berlin, Germany. Springer.

Beauchemin, K., Rode, L., and Sewalt, V. (1995). Fibrolytic enzymes increase fiber digestibility and growth rate of steers fed dry forages. *Canadian Journal of Animal Science,* 75 (*4*), 641-644.

Beauchemin, K. A., Colombatto, D., Morgavi, D. P., and Yang, W. Z. (2003). Use of exogenous fibrolytic enzymes to improve feed utilization by ruminants. *Journal of Animal Science,* 81 (*14*), E37-E47.

Beg, Q. K., Kapoor, M., Mahajan, L., and Hoondal, G. S. (2001). Microbial xylanases and their industrial applications: a review. *Applied Microbiology and Biotechnology,* 56 (*3*), 326-338.

Belfort, M., and Roberts, R. J. (1997). Homing endonucleases: keeping the house in order. *Nucleic Acids Research,* 25 (*17*), 3379-3388.

Benjamini, Y., and 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 (*1*), 289-300.

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2015). GenBank. *Nucleic Acids Research,* 43 (*Database issue*), D30-D35.

Berger, L., Fahey, G. C., Bourquin, L. D., and Titgemeyer, E. C. (1994). Modification of forage quality after harvest. In J. G. C. Fahey (Ed.), Forage Cell Wall Structure and Digestibility. pp. 922-966. Madison, WI, USA. American Society of Agronomy.

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biology and Biochemistry,* 43 (*7*), 1450-1455.

Berntsson, R. P.-A., Smits, S. H., Schmitt, L., Slotboom, D.-J., and Poolman, B. (2010). A structural classification of substrate-binding proteins. *FEBS Letters,* 584 (*12*), 2606-2617.

Beveridge, T. J. (2001). Use of the Gram stain in microbiology. *Biotechnic & Histochemistry,* 76 (*3*), 111-118.

Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances,* 30 (*6*), 1575-1588.

Biely, P., and Côté, G. (2005). Microbial hemicellulolytic carbohydrate esterases. In C. Hou (Ed.), Handbook of industrial biocatalysis. pp. 431-456. Boca Raton, USA. CRC, Taylor and Francis.

Binate, S., N'dri, D., Toka, M., and Kouame, L. P. (2008). Purification and characterization of two β-glucosidases from termite workers *Macrotemes bellicosus* (Termitidae: Macrotermitinae). *Journal of Applied Biosciences,* 10 (*1*), 461-470.

Black, G., Rixon, J., Clarke, J., Hazlewood, G., Theodorou, M., Morris, P., and Gilbert, H. (1996). Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. *Biochemical Journal,* 319 (*2*), 515-520.

Blackburn, T., and Hobson, P. (1962). Further studies on the isolation of proteolytic bacteria from the sheep rumen. *Journal of General Microbiology,* 29 (*1*), 69-81.

Blake, A. W., McCartney, L., Flint, J. E., Bolam, D. N., Boraston, A. B., Gilbert, H. J., and Knox, J. P. (2006). Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *Journal of Biological Chemistry,* 281 (*39*), 29321-29329.

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., and Hugenholtz, P. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics,* 8 (*1*), 209.

Bolam, D., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M., Boraston, A., Hazlewood, G., and Gilbert, H. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochemical Journal,* 331 (*3*), 775-781.

Bond, J. J., Dunne, J. C., Kwan, F. Y., Li, D., Zhang, K., Leahy, S. C., Kelly, W. J., Attwood, G. T., and Jordan, T. W. (2012). Carbohydrate transporting membrane proteins of the rumen bacterium, *Butyrivibrio proteoclasticus*. *Journal of Proteomics,* 75 (*11*), 3138-3144.

Bonfield, J. K., Beal, K. F., Betts, M. J., and Staden, R. (2002). Trev: a DNA trace editor and viewer. *Bioinformatics,* 18 (*1*), 194-195.

Boraston, A., Bolam, D., Gilbert, H., and Davies, G. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal,* 382 (*3*), 769-781.

Borneman, W. S., Hartley, R. D., Morrison, W. H., Akin, D. E., and Ljungdahl, L. G. (1990). Feruloyl and *p*-coumaroyl esterase from anaerobic fungi in relation to plant cell wall degradation. *Applied Microbiology and Biotechnology,* 33 (*3*), 345-351.

Bozzola, J. J., and Russell, L. D. (1999). Electron microscopy: principles and techniques for biologists. Jones & Bartlett Learning.

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs,* 27 (*4*), 325-349.

Brice, R. E., and Morrison, I. M. (1982). The degradation of isolated hemicelluloses and lignin-hemicellulose complexes by cell-free, rumen hemicellulases. *Carbohydrate research,* 101 (*1*), 93-100.

Brock, F., Forsberg, C., and Buchanan-Smith, J. (1982). Proteolytic activity of rumen microorganisms and effects of proteinase inhibitors. *Applied and Environmental Microbiology,* 44 (*3*), 561-569.

Brown, R. M., and Saxena, I. M. (2000). Cellulose biosynthesis: a model for understanding the assembly of biopolymers. *Plant Physiology and Biochemistry,* 38 (*1*), 57-67.

Brulc, J. M., Antonopoulos, D. A., Miller, M. E. B., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., Edwards, R. E., Frank, E. D., Emerson, J. B., and Wacklin, P. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences,* 106 (*6*), 1948-1953.

Brumm, P., Mead, D., Boyum, J., Drinkwater, C., Deneke, J., Gowda, K., Stevenson, D., and Weimer, P. (2011). Functional annotation of *Fibrobacter succinogenes* S85 carbohydrate active enzymes. *Applied Biochemistry and Biotechnology,* 163 (*5*), 649-657.

Bryant, M. P. (1986). Genus IV. *Butyrivibrio*. In P. H. A. Sneath, N. S. Mair, M. E. Sharpe and J. G. Holt (Eds.), Bergeys Manual of Systematic Bacteriology. Vol. 2, pp. 641-643. Baltimore, USA. Williams and Williams.

Bryant, M. P., and Robinson, I. M. (1962). Some nutritional characteristics of predominant culturable ruminal bacteria. *Journal of Bacteriology,* 84 (*4*), 605-614.

Bryant, M. P., and Small, N. (1956a). The anaerobic monotrichous butyric acid-producing curved rod-shaped bacteria of the rumen. *Journal of Bacteriology,* 72 (*1*), 16-21.

Bryant, M. P., and Small, N. (1956b). Characteristics of two new genera of anaerobic curved rods isolated from the rumen of cattle. *Journal of Bacteriology,* 72 (*1*), 22.

Caffall, K. H., and Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate research,* 344 (*14*), 1879-1900.

Callaway, T., Dowd, S., Edrington, T., Anderson, R., Krueger, N., Bauer, N., Kononoff, P., and Nisbet, D. (2010). Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. *Journal of Animal Science,* 88 (*12*), 3977-3983.

Campbell, G., and Bedford, M. (1992). Enzyme applications for monogastric feeds: A review. *Canadian Journal of Animal Science,* 72 (*3*), 449-466.

Campbell, M. M., and 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.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research,* 37 (*Database issue*), D233-D238.

Cantarel, B. L., Lombard, V., and Henrissat, B. (2012). Complex carbohydrate utilization by the healthy human microbiome. *PLoS One,* 7 (*6*), e28742.

Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annual Review of Plant Biology,* 47 (*1*), 445-476.

Carpita, N. C., Defernez, M., Findlay, K., Wells, B., Shoue, D. A., Catchpole, G., Wilson, R. H., and McCann, M. C. (2001). Cell wall architecture of the elongating maize coleoptile. *Plant Physiology,* 127 (*2*), 551-565.

Carpita, N. C., and Gibeaut, D. M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal,* 3 (*1*), 1-30.

Carr, S. M. (1999). Ruminant digestion in *Bos taurus*. In N. A. Campbell, L. G. Lawrence G. Mitchell and J. B. Reece (Eds.), Biology. 3rd ed. Addison-Wesley Longman

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics,* 25 (*1*), 119-120.

Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P., and Karp, P. D. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research,* 40 (*Database issue*), D742-753.

Chandel, A. K., and Singh, O. V. (2011). Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of 'Biofuel'. *Applied Microbiology and Biotechnology,* 89 (*5*), 1289-1303.

Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., and Andersen, M. R. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research,* 33 (*20*), 179.

Chen, J. Q., and Weimer, P. J. (2001). Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. *Microbiology,* 147 (*1*), 21-30.

Cheng, K.-J., and Costerton, J. W. (1980). The formation of microcolonies by rumen bacteria. *Canadian Journal of Microbiology,* 26 (*9*), 1104-1113.

Cheng, K., and Costerton, J. (1977). Ultrastructure of *Butyrivibrio fibrisolvens*: a gram-positive bacterium. *Journal of Bacteriology,* 129 (*3*), 1506-1512.

Cheng, K., Forsberg, C., Minato, H., and Costerton, J. (1991). Microbial ecology and physiology of feed degradation within the rumen. In T. Tsuda, Y. Sasaki and R. Kawashima (Eds.), Physiological aspects of digestion and metabolism in ruminants: Proceedings of the seventh international symposium on ruminant physiology. Vol. 7, pp. 595-624. London, UK. Academic Press.

Chesson, A., and Forsberg, C. (1997). Polysaccharide degradation by rumen microorganisms. In P. N. Hobson and C. S. Stewart (Eds.), The rumen microbial ecosystem. 2nd ed., pp. 329-381. London, UK. Chapman and Hall.

Christov, L. P., and Prior, B. A. (1993). Esterases of xylan-degrading microorganisms: production, properties, and significance. *Enzyme and Microbial Technology,* 15 (*6*), 460-475.

Cid, M., Pedersen, H. L., Kaneko, S., Coutinho, P. M., Henrissat, B., Willats, W. G., and Boraston, A. B. (2010). Recognition of the helical structure of β-1, 4-galactan by a new family of carbohydrate-binding modules. *Journal of Biological Chemistry,* 285 (*46*), 35999-36009.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology,* 18 (*1*), 117.

Clauss, M., Hume, I. D., and Hummel, J. (2010). Evolutionary adaptations of ruminants and their potential relevance for modern production systems. *Animal Behaviour,* 4 (*7*), 979-992.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A., McGarrell, D. M., Marsh, T., and Garrity, G. M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research,* 37 (*Database issue*), D141-D145.

Collins, M., Lawson, P., Willems, A., Cordoba, J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., and Farrow, J. (1994). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *International Journal of Systematic Bacteriology,* 44 (*4*), 812-826.

Colombatto, D., Morgavi, D., Furtado, A., and Beauchemin, K. (2003). Screening of exogenous enzymes for ruminant diets: Relationship between biochemical characteristics and *in vitro* ruminal degradation. *Journal of Animal Science,* 81 (*10*), 2628-2638.

Conover, W. J. (1980). Practical nonparametric statistics. 3rd ed. John Wiley & Sons.

Cooper, R. (1984). Metabolism of methylglyoxal in microorganisms. *Annual Reviews in Microbiology,* 38 (*1*), 49-68.

Correia, M. A., Pires, V. M., Gilbert, H. J., Bolam, D. N., Fernandes, V. O., Alves, V. D., Prates, J. A., Ferreira, L. M., and Fontes, C. M. (2009). Family 6 carbohydrate-binding modules display multiple β1, 3-linked glucan-specific binding interfaces. *FEMS Microbiology Letters,* 300 (*1*), 48-57.

Correia, M. A., Prates, J. A., Brás, J., Fontes, C. M., Newman, J. A., Lewis, R. J., Gilbert, H. J., and Flint, J. E. (2008). Crystal structure of a cellulosomal family 3 carbohydrate esterase from *Clostridium thermocellum* provides insights into the mechanism of substrate recognition. *Journal of Molecular Biology,* 379 (*1*), 64-72.

Cosgrove, D. J. (2000). Expansive growth of plant cell walls. *Plant Physiology and Biochemistry,* 38 (*1*), 109-124.

Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nature Reviews: Molecular Cell Biology,* 6 (*11*), 850-861.

Cotta, M. A. (1993). Utilization of xylooligosaccharides by selected ruminal bacteria. *Applied and Environmental Microbiology,* 59 (*11*), 3557-3563.

Cotta, M. A., and Zeltwanger, R. L. (1995). Degradation and utilization of xylan by the ruminal bacteria *Butyrivibrio fibrisolvens* and *Selenomonas ruminantium*. *Applied and Environmental Microbiology,* 61 (*12*), 4396-4402.

Coughlan, M., and Hazlewood, G. P. (1993). β-1, 4-D-xylan-degrading enzyme systems: biochemistry, molecular biology and applications. *Biotechnology and Applied Biochemistry,* 17 (*3*), 259-289.

Coughlan, M. P. (1985). The properties of fungal and bacterial cellulases with comment on their production and application. *Biotechnology and Genetic Engineering Reviews,* 3 (*1*), 39-110.

Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology,* 328 (*2*), 307-317.

Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics,* 11 (*1*), 485.

Creevey, C. J., Kelly, W. J., Henderson, G., and Leahy, S. C. (2014). Determining the culturability of the rumen bacterial microbiome. *Microbial Biotechnology,* 7 (*5*), 467-479.

Cruywagen, C., and Van Zyl, W. (2008). Effects of a fungal enzyme cocktail treatment of high and low forage diets on lamb growth. *Animal Feed Science and Technology,* 145 (*1*), 151-158.

Dado, R., and Allen, M. (1995). Intake limitations, feeding behavior, and rumen function of cows challenged with rumen fill from dietary fiber or inert bulk. *Journal of Dairy Science,* 78 (*1*), 118-133.

Dale, H. E., Stewart, R. E., and Brody, S. (1954). Rumen temperature. I. Temperature gradients during feeding and fasting. In D. W. Bruner (Ed.), The Cornell Veterinarian. Vol. 44, pp. 368-374. New York, USA. Cornell University.

Dam, P., Kataeva, I., Yang, S.-J., Zhou, F., Yin, Y., Chou, W., Poole, F. L., Westpheling, J., Hettich, R., and Giannone, R. (2011). Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Research,* 39 (*8*), 3240-3254.

Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews,* 72 (*2*), 317-364.

Davies, G. J., Gloster, T. M., and Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology,* 15 (*6*), 637-645.

de Menezes, A. B., Lewis, E., O'Donovan, M., O'Neill, B. F., Clipson, N., and Doyle, E. M. (2011). Microbiome analysis of dairy cows fed pasture or total mixed ration diets. *FEMS Microbiology Ecology,* 78 (*2*), 256-265.

Dehority, B. (1966). Characterization of several bovine rumen bacteria isolated with a xylan medium. *Journal of Bacteriology,* 91 (*5*), 1724-1729.

Dehority, B. (1973). Hemicellulose degradation by rumen bacteria. *Federation Proceedings,* 32 (*7*), 1819-1825.

Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. (1999). Alignment of whole genomes. *Nucleic Acids Research,* 27 (*11*), 2369-2376.

Dibbayawan, T., Cox, G., Cho, K. Y., and Dwarte, D. M. (1985). Cell wall and plasma membrane architecture of *Butyrivibrio spp*. *Journal of Ultrastructure Research,* 90 (*3*), 286-293.

Diez-Gonzalez, F., Bond, D. R., Jennings, E., and Russell, J. B. (1999). Alternative schemes of butyrate production in *Butyrivibrio fibrisolvens* and their relationship to acetate utilization, lactate production, and phylogeny. *Archives of Microbiology,* 171 (*5*), 324-330.

Dorsch, M., and Stackebrandt, E. (1992). Some modifications in the procedure of direct sequencing of PCR amplified 16*S* rDNA. *Journal of Microbiological Methods,* 16 (*4*), 271-279.

Ducros, V., Charnock, S. J., Derewenda, U., Derewenda, Z. S., Dauter, Z., Dupont, C., Shareck, F., Morosoli, R., Kluepfel, D., and Davies, G. J. (2000). Substrate specificity in glycoside hydrolase family 10 structural and kinetic analysis of the *Streptomyces lividans* xylanase 10A. *Journal of Biological Chemistry,* 275 (*30*), 23020-23026.

Dunne, J. C. (2010). A proteomic analysis of fibre degradation and assimilation by *Butyrivibrio proteoclasticus.* Doctor of Philosophy in Cell and Molecular Biosciences, Victoria University of Wellington, Wellington, NZ.

Dunne, J. C., Kelly, W. J., Leahy, S. C., Li, D., Bond, J. J., Peng, L., Attwood, G. T., and Jordan, T. W. (2015). The cytosolic oligosaccharide-degrading proteome of *Butyrivibrio proteoclasticus*. *Proteomes,* 3 (*4*), 347-368.

Dunne, J. C., Li, D., Kelly, W. J., Leahy, S. C., Bond, J. J., Attwood, G. T., and Jordan, T. W. (2011). Extracellular polysaccharide-degrading proteome of *Butyrivibrio proteoclasticus*. *Journal of Proteome Research,* 11 (*1*), 131-142.

Durso, L. M., Harhay, G. P., Smith, T. P., Bono, J. L., DeSantis, T. Z., Harhay, D. M., Andersen, G. L., Keen, J. E., Laegreid, W. W., and Clawson, M. L. (2010). Animal-to-animal variation in fecal microbial diversity among beef cattle. *Applied and Environmental Microbiology,* 76 (*14*), 4858-4862.

Duthoit, F., Godon, J.-J., and Montel, M.-C. (2003). Bacterial community dynamics during production of registered designation of origin Salers cheese as evaluated by 16*S* rRNA gene single-strand conformation polymorphism analysis. *Applied and Environmental Microbiology,* 69 (*7*), 3840-3848.

Ebringerová, A., Hromádková, Z., and Heinze, T. (2005). Hemicellulose. In T. Heinze (Ed.), Polysaccharides I: Structure, characterization and use. Vol. 186, pp. 1-67. Berlin, Germany. Springer.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology,* 7 (*10*), e1002195.

Edgar, R. C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics,* 8 (*1*), 18.

Edwards, J., Huws, S., Kim, E., Lee, M., Kingston-Smith, A., and Scollan, N. (2008). Advances in microbial ecosystem concepts and their consequences for ruminant agriculture. *Animal,* 2 (*5*), 653-660.

Edwards, J. E., McEwan, N. R., Travis, A. J., and Wallace, R. J. (2004). 16*S* rDNA library-based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek,* 86 (*3*), 263-281.

Escalante, A., Wacher, C., and Farrés, A. (2001). Lactic acid bacterial diversity in the traditional Mexican fermented dough pozol as determined by 16*S* rDNA sequence analysis. *International Journal of Food Microbiology,* 64 (*1*), 21-31.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research,* 8 (*3*), 175-185.

Faith, D. P., Minchin, P. R., and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio,* 69 (*1-3*), 57-68.

Fernández, M. H., and Vrba, E. S. (2005). A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews,* 80 (*2*), 269-302.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., and Mistry, J. (2013). Pfam: the protein families database. *Nucleic Acids Research,* 42 (*Database issue*), D222-D230.

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research,* gkr367.

Finn, R. D., Marshall, M., and Bateman, A. (2005). iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics,* 21 (*3*), 410-412.

Flicek, P., and Birney, E. (2009). Sense from sequence reads: methods for alignment and assembly. *Nature Methods,* 6 (*Supplement 11*), S6-S12.

Flint, H. J. (2004). Polysaccharide breakdown by anaerobic microorganisms inhabiting the mammalian gut. In A. I. Laskin, J. W. Bennett and G. M. Gadd (Eds.), Advances in Applied Microbiology. Vol. 56, pp. 89-120. London, UK. Elsevier.

Flint, H. J., Thomson, A. M., and Bisset, J. (1988). Plasmid-associated transfer of tetracycline resistance in *Bacteroides ruminicola*. *Applied and Environmental Microbiology,* 54 (*4*), 855-860.

Fondevila, M., and Dehority, B. (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.

Foroozandeh, A., Rezaeian, M., Balaly, G., and Alikhani, M. (2009). Relative contributions of ruminal bacteria, protozoa and fungi to degradation of forage fiber fractions. *Journal of Animal and Veterinary Advances,* 8 (*3*), 603-607.

Forster, R., Teather, R., Gong, J., and Deng, S.-J. (1996). 16*S* rDNA analysis of *Butyrivibrio fibrisolvens*: phylogenetic position and relation to butyrate-producing anaerobic bacteria from the rumen of white-tailed deer. *Letters in Applied Microbiology,* 23 (*4*), 218-222.

Forster, R. J., Gong, J., and Teather, R. M. (1997). Group-specific 16*S* rRNA hybridization probes for determinative and community structure studies of *Butyrivibrio fibrisolvens* in the rumen. *Applied and Environmental Microbiology,* 63 (*4*), 1256-1260.

Fouts, D. E., Szpakowski, S., Purushe, J., Torralba, M., Waterman, R. C., MacNeil, M. D., Alexander, L. J., and Nelson, K. E. (2012). Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. *PLoS One,* 7 (*11*), e48289.

Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *Journal of Molecular Biology,* 286 (*1*), 279-290.

Gardiner, K. (1991). Pulsed field gel electrophoresis. *Analytical chemistry,* 63 (*7*), 658-665.

Gilbert, H. J., Knox, J. P., and Boraston, A. B. (2013). Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology,* 23 (*5*), 669-677.

Gilkes, N., Henrissat, B., Kilburn, D., Miller, R., and Warren, R. (1991). Domains in microbial β-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiological Reviews,* 55 (*2*), 303-315.

Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R., Warren, R., and Kilburn, D. (1992). The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. *Journal of Biological Chemistry,* 267 (*10*), 6743-6749.

Giraldo, L., Tejido, M., Ranilla, M., Ramos, S., and Carro, M. (2008). Influence of direct-fed fibrolytic enzymes on diet digestibility and ruminal activity in sheep fed a grass hay-based diet. *Journal of Animal Science,* 86 (*7*), 1617-1623.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources,* 11 (*5*), 759-769.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology,* 11 (*8*), R86.

Golan, G., Shallom, D., Teplitsky, A., Zaide, G., Shulami, S., Baasov, T., Stojanoff, V., Thompson, A., Shoham, Y., and Shoham, G. (2004). Crystal structures of *Geobacillus stearothermophilus* α-glucuronidase complexed with its substrate and products: mechanistic implications. *Journal of Biological Chemistry,* 279 (*4*), 3014-3024.

González, I., Lê Cao, K.-A., Davis, M. J., and Déjean, S. (2012). Visualising associations between paired 'omics' data sets. *BioData mining,* 5 (*1*), 1-23.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology,* 57 (*1*), 81-91.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* 53 (*3-4*), 325-338.

Gower, J. C., and Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of classification,* 3 (*1*), 5-48.

Grabber, J. H., Ralph, J., and Hatfield, R. D. (2000). Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. *Journal of Agricultural and Food Chemistry,* 48 (*12*), 6106-6113.

Grabber, J. H., Ralph, J., Lapierre, C., and Barrière, Y. (2004). Genetic and molecular basis of grass cell-wall degradability. I. Lignin–cell wall matrix interactions. *Comptes Rendus Biologies,* 327 (*5*), 455-465.

Graham, H., and Balnave, D. (1995). Dietary enzymes for increasing energy availability. In R. Wallace and A. Chesson (Eds.), Biotechnology in Animal Feeds and Animal Feeding. pp. 295-309. New York, USA. VCH Inc.

Gray, K. A., Zhao, L., and Emptage, M. (2006). Bioethanol. *Current Opinion in Chemical Biology,* 10 (*2*), 141-146.

Greenacre, M. J. (1993). Correspondence analysis in practice. London. Academic Press.

Greve, L. C., Labavitch, J., and Hungate, R. (1984). α-L-arabinofuranosidase from *Ruminococcus albus* 8: purification and possible role in hydrolysis of alfalfa cell wall. *Applied and Environmental Microbiology,* 47 (*5*), 1135-1140.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research,* 33 (*Database issue*), D121-D124.

Hackmann, T. J., and Spain, J. N. (2010). Invited review: Ruminant ecology and evolution: Perspectives useful to ruminant livestock research and production. *Journal of Dairy Science,* 93 (*4*), 1320-1334.

Haenlein, G. (2001). Past, present, and future perspectives of small ruminant dairy research. *Journal of Dairy Science,* 84 (*9*), 2097-2115.

Harrison, P. W., Lower, R. P., Kim, N. K., and Young, J. P. (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in Microbiology,* 18 (*4*), 141-148.

Harvey, A. J., Hrmova, M., De Gori, R., Varghese, J. N., and Fincher, G. B. (2000). Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins: Structure, Function, and Bioinformatics,* 41 (*2*), 257-269.

Hatfield, R. (1989). Structural polysaccharides in forages and their degradability. *Agronomy Journal,* 81 (*1*), 39-46.

Hatfield, R., Ralph, J., and Grabber, J. (1999a). Cell wall structural foundations: Molecular basis for improving forage digestibilities. *Crop Science,* 39 (*1*), 27-37.

Hatfield, R., and Weimer, P. (1995). Degradation characteristics of isolated and *in situ* cell wall lucerne pectic polysaccharides by mixed ruminal microbes. *Journal of the Science of Food and Agriculture,* 69 (*2*), 185-196.

Hatfield, R. D. (1993). Cell wall polysaccharide interactions and degradability. In H. G. Jung, H. E. Yung, D. R. Buxton, R. D. Hatfield and J. Ralph (Eds.), Forage Cell Wall Structure and Digestibility. pp. 285-313. Madison, WI, USA. American Society of Agronomy-Crop Science Society of America-Soil Science Society of America (ASA-CSSA-SSSA).

Hatfield, R. D., Wilson, J. R., and Mertens, D. R. (1999b). Composition of cell walls isolated from cell types of grain sorghum stems. *Journal of the Science of Food and Agriculture,* 79 (*6*), 891-899.

Hazlewood, G. P., Theodorou, M. K., Hutchings, A., Jordan, D. J., and Galfre, G. (1986). Preparation and characterization of monoclonal antibodies to a *Butyrivibrio* sp. and their potential use in the identification of rumen *Butyrivibrios*, using an enzyme-linked immunosorbent assay. *Journal of General Microbiology,* 132 (*1*), 43-52.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Collaborators, G. R. C., and Janssen, P. H. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports,* 5 (*14567*), 1-13.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal,* 280 (*2*), 309-316.

Henrissat, B. (1998). Enzymology of cell-wall degradation. *Biochemical Society Transactions,* 26 (*1*), 153-156.

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.-P. (1989). Cellulase families revealed by hydrophobic cluster analysis. *Gene,* 81 (*1*), 83-95.

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology,* 7 (*5*), 637-644.

Hervé, C., Rogowski, A., Gilbert, H. J., and Paul Knox, J. (2009). Enzymatic treatments reveal differential capacities for xylan recognition and degradation in primary and secondary plant cell walls. *The Plant Journal,* 58 (*3*), 413-422.

Hespell, R., Kato, K., and Costerton, J. (1993). Characterization of the cell wall of *Butyrivibrio* species. *Canadian Journal of Microbiology,* 39 (*10*), 912-921.

Hespell, R., and Whitehead, T. (1990). Physiology and genetics of xylan degradation by gastrointestinal tract bacteria. *Journal of Dairy Science,* 73 (*10*), 3013-3022.

Hespell, R., Wolf, R., and Bothast, R. (1987). Fermentation of xylans by *Butyrivibrio fibrisolvens* and other ruminal bacteria. *Applied and Environmental Microbiology,* 53 (*12*), 2849-2853.

Hespell, R. B. (1992a). Fermentation of xylans by *Butyrivibrio fibrisolvens* and *Thermoanaerobacter* strain B6A: Utilization of uronic acids and xylanolytic activites. *Current Microbiology,* 25 (*4*), 189-195.

Hespell, R. B. (1992b). The genera *Butyrivibrio, Lachnospira*, and *Roseburia.* In A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K.-H. Schleifer (Eds.), The

Prokaryotes: A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications. pp. 2023 - 2033. New York, USA. Springer.
Hespell, R. B., and Cotta, M. A. (1995). Degradation and utilization by *Butyrivibrio fibrisolvens* H17c of xylans with different chemical and physical properties. *Applied and Environmental Microbiology,* 61 (*8*), 3042-3050.
Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D. S., Chen, F., and Zhang, T. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science,* 331 (*6016*), 463-467.
Himmel, M. E., Xu, Q., Luo, Y., Ding, S.-Y., Lamed, R., and Bayer, E. A. (2010). Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels,* 1 (*2*), 323-341.
Hobson, P. N., and Stewart, C. S. (1997). The rumen microbial ecosystem. 2nd ed. London, UK. Chapman and Hall.
Holden, L. A., Muller, L. D., and Fales, S. L. (1994). Estimation of intake in high producing Holstein cows grazing grass pasture. *Journal of Dairy Science,* 77 (*8*), 2332-2340.
Hollander, M., Wolfe, D. A., and Chicken, E. (2013). Nonparametric statistical methods. John Wiley & Sons.
Howard, B. (1957). Hydrolysis of the soluble pentosans of wheat flour and *Rhodymenia palmata* by ruminal micro-organisms. *Biochemical Journal,* 67 (*4*), 643.
Howard, R., Abotsi, E., Van Rensburg, E. J., and Howard, S. (2004). Lignocellulose biotechnology: issues of bioconversion and enzyme production. *African Journal of Biotechnology,* 2 (*12*), 602-619.
Hu, Y., Smith, D., Cheng, K.-J., and Forsberg, C. W. (1991). Cloning of a xylanase gene from *Fibrobacter succinogenes* 135 and its expression in *Escherichia coli*. *Canadian Journal of Microbiology,* 37 (*7*), 554-561.
Huang, L., and Forsberg, C. (1990). Cellulose digestion and cellulase regulation and distribution in *Fibrobacter succinogenes* subsp. *succinogenes* S85. *Applied and Environmental Microbiology,* 56 (*5*), 1221-1228.
Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics,* 26 (*5*), 680-682.
Hungate, R. (1950). The anaerobic mesophilic cellulolytic bacteria. *Bacteriological Reviews,* 14 (*1*), 1.
Hungate, R. E. (1966). The rumen and its microbes. New York, USA. Academic Press.
Hungate, R. E. (1975). The rumen microbial ecosystem. In R. F. Johnston, P. W. Frank and C. D. Michener (Eds.), Annual Review of Ecology and Systematics. Vol. 6, pp. 39-66. California, USA. Annual Reviews.
Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Palaniappan, K., Szeto, E., Pillay, M., Chen, I.-M. A., and Pati, A. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v. 4). *Standards in Genomic Sciences,* 10 (*1*), 1-6.
Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics,* 11 (*1*), 119.
Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics,* 5 (*3*), 299-314.
Ishii, T. (1997). Structure and functions of feruloylated polysaccharides. *Plant Science,* 127 (*2*), 111-127.
Iverson, W. G., and Millis, N. F. (1976). Characterization of *Streptococcus bovis* bacteriophages. *Canadian Journal of Microbiology,* 22 847-852.

Jain, R., Rivera, M. C., Moore, J. E., and Lake, J. A. (2003). Horizontal gene transfer accelerates genome innovation and evolution. *Molecular Biology and Evolution,* 20 (*10*), 1598-1602.

Jami, E., Israel, A., Kotser, A., and Mizrahi, I. (2013). Exploring the bovine rumen bacterial community from birth to adulthood. *The ISME Journal,* 7 (*6*), 1069-1079.

Jami, E., and Mizrahi, I. (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS One,* 7 (*3*), e33306.

Jarvis, M. C. (1984). Structure and properties of pectin gels in plant cell walls. *Plant, Cell & Environment,* 7 (*3*), 153-164.

Jensen, L. J., Friis, C., and Ussery, D. W. (1999). Three views of microbial genomes. *Research in Microbiology,* 150 (*9*), 773-777.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences: CABIOS,* 8 (*3*), 275-282.

Jun, H.-S., Qi, M., Gong, J., Egbosimba, E. E., and Forsberg, C. W. (2007). Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. *Journal of Bacteriology,* 189 (*19*), 6806-6815.

Jung, H.-J. G., and Lamb, J. F. (2003). Identification of lucerne stem cell wall traits related to *in vitro* neutral detergent fibre digestibility. *Animal Feed Science and Technology,* 110 (*1*), 17-29.

Jung, H., and Allen, M. (1995). Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. *Journal of Animal Science,* 73 (*9*), 2774-2790.

Jung, H. J. (2011). Forage digestibility: the intersection of cell wall lignification and plant tissue anatomy. Paper presented at the: *Proceedings of the 23rd annual Florida ruminant nutrition symposium*, University of Florida, Gainesville, Florida, USA.

Juturu, V., and Wu, J. C. (2012). Microbial xylanases: engineering, production and industrial applications. *Biotechnology Advances,* 30 (*6*), 1219-1227.

Juturu, V., and Wu, J. C. (2013). Insight into microbial hemicellulases other than xylanases: a review. *Journal of Chemical Technology and Biotechnology,* 88 (*3*), 353-363.

Kabel, M. A., van den Borne, H., Vincken, J.-P., Voragen, A. G., and Schols, H. A. (2007). Structural differences of xylans affect their interaction with cellulose. *Carbohydrate Polymers,* 69 (*1*), 94-105.

Kalmokoff, M., and Teather, R. (1997). Isolation and characterization of a bacteriocin (Butyrivibriocin AR10) from the ruminal anaerobe *Butyrivibrio fibrisolvens* AR10: evidence in support of the widespread occurrence of bacteriocin-like activity among ruminal isolates of *B. fibrisolvens*. *Applied and Environmental Microbiology,* 63 (*2*), 394-402.

Kalmokoff, M. L., Bartlett, F., and Teather, R. M. (1996). Are ruminal bacteria armed with bacteriocins? *Journal of Dairy Science,* 79 (*12*), 2297-2306.

Kamra, D. (2005). Rumen microbial ecosystem. *Current Science,* 89 (*1*), 124-135.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research,* 28 (*1*), 27-30.

Kav, A. B., Sasson, G., Jami, E., Doron-Faigenboim, A., Benhar, I., and Mizrahi, I. (2012). Insights into the bovine rumen plasmidome. *Proceedings of the National Academy of Sciences,* 109 (*14*), 5452-5457.

Kaznadzey, A., Alexandrova, N., Novichkov, V., and Kaznadzey, D. (2013). PSimScan: Algorithm and utility for fast protein similarity search. *PLoS One,* 8 (*3*), e58505.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., and Duran, C. (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., Asmundson, R., and Hopcroft, D. (1987). Isolation and characterization of a strictly anaerobic, cellulolytic spore former: *Clostridium chartatabidum* sp. nov. *Archives of Microbiology,* 147 (*2*), 169-173.

Kelly, W. J., Attwood, G. T., Cookson, A. L., Goodwin, L., Woyke, T., Henderson, G., Lambie, S. C., Perry, R., Teh, K., Creevey, C. J., and Leahy, S. C. (2014). The Hungate 1000 Project: A catalogue of reference genomes from the rumen microbiome. Available from: http://www.hungate1000.org.nz.

Kelly, W. J., Leahy, S. C., Altermann, E., Yeoman, C. J., Dunne, J. C., Kong, Z., Pacheco, D. M., Li, D., Noel, S. J., Moon, C. D., Cookson, A. L., and Attwood, G. T. (2010). The glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316^{T} highlights adaptation to a polysaccharide-rich environment. *PLoS One,* 5 (*8*), e11942.

Kenters, N., Henderson, G., Jeyanathan, J., Kittelmann, S., and Janssen, P. H. (2011). Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *Journal of Microbiological Methods,* 84 (*1*), 52-60.

Kepler, C. R., Hirons, K. P., McNeill, J. J., and Tove, S. B. (1966). Intermediates and products of the biohydrogenation of linoleic acid by *Butyrivibrio fibrisolvens*. *Journal of Biological Chemistry,* 241 (*6*), 1350-1354.

Kim, J.-B., Olek, A. T., and Carpita, N. C. (2000). Cell wall and membrane-associated exo-β-D-glucanases from developing maize seedlings. *Plant Physiology,* 123 (*2*), 471-486.

Kim, M., Morrison, M., and Yu, Z. (2011a). Phylogenetic diversity of bacterial communities in bovine rumen as affected by diets and microenvironments. *Folia Microbiologica,* 56 (*5*), 453-458.

Kim, M., Morrison, M., and Yu, Z. (2011b). Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiology Ecology,* 76 (*1*), 49-63.

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16*S* rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology,* 64 (*1*), 346-351.

Kim, M., and Yu, Z. (2012). Quantitative comparisons of select cultured and uncultured microbial populations in the rumen of cattle fed different diets. *Journal of Animal Science and Biotechnology,* 3 (*28*), 1-6.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* 16 (*2*), 111-120.

Kluskens, L., Van Alebeek, G., Voragen, A., de Vos, W., and van der Oost, J. (2003). Molecular and biochemical characterization of the thermoactive family 1 pectate lyase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemical Journal,* 370 (*2*), 651-659.

Koike, S., Pan, J., Kobayashi, Y., and Tanaka, K. (2003a). Kinetics of *in sacco* fiber-attachment of representative ruminal cellulolytic bacteria monitored by competitive PCR. *Journal of Dairy Science,* 86 (*4*), 1429-1435.

Koike, S., Yoshitani, S., Kobayashi, Y., and Tanaka, K. (2003b). Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiology Letters,* 229 (*1*), 23-30.

Kolenová, K., Ryabova, O., Vršanská, M., and Biely, P. (2010). Inverting character of family GH115 α-glucuronidases. *FEBS Letters,* 584 (*18*), 4063-4068.

Kolver, E., and De Veth, M. (2002). Prediction of ruminal pH from pasture-based diets. *Journal of Dairy Science,* 85 (*5*), 1255-1266.

Kondo, T., Mizuno, K., and Kato, T. (1990). Cell wall-bound *p*-coumaric and ferulic acids in Italian ryegrass. *Canadian Journal of Plant Science,* 70 (*2*), 495-499.

Kong, Y., Teather, R., and Forster, R. (2010). Composition, spatial distribution, and diversity of the bacterial communities in the rumen of cows fed different forages. *FEMS Microbiology Ecology,* 74 (*3*), 612-622.

Kong, Z. (2007). Microarray analysis of *Clostridium proteoclasticum* genes involved in hemicellulose degradation. Master of Science in Biological Sciences, University of Auckland, Palmerston North, NZ.

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America,* 102 (*7*), 2567-2572.

Konstantinidis, K. T., and Tiedje, J. M. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Current Opinion in Microbiology,* 10 (*5*), 504-509.

Kopecny, J., Logar, R. M., and Kobayashi, Y. (2001). Phenotypic and genetic data supporting reclassification of *Butyrivibrio fibrisolvens* isolates. *Folia Microbiologica,* 46 (*1*), 45-48.

Kopecny, J., Zorec, M., Mrazek, J., Kobayashi, Y., and Marinsek-Logar, R. (2003). *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrivibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. *International Journal of Systematic & Evolutionary Microbiology,* 53 (*1*), 201-209.

Kosugi, A., Murashima, K., and Doi, R. H. (2002). Xylanase and acetyl xylan esterase activities of XynA, a key subunit of the *Clostridium cellulovorans* cellulosome for xylan degradation. *Applied and Environmental Microbiology,* 68 (*12*), 6399-6402.

Krause, D. O., Denman, S. E., Mackie, R. I., Morrison, M., Rae, A. L., Attwood, G. T., and McSweeney, C. S. (2003). Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiology Reviews,* 27 (*5*), 663-693.

Kuhad, R. C., Gupta, R., Khasa, Y. P., Singh, A., and Zhang, Y.-H. P. (2011). Bioethanol production from pentose sugars: Current status and future prospects. *Renewable and Sustainable Energy Reviews,* 15 (*9*), 4950-4962.

Kulkarni, N., Shendye, A., and Rao, M. (1999). Molecular and biotechnological aspects of xylanases. *FEMS Microbiology Reviews,* 23 (*4*), 411-456.

Kung, L., Treacher, R., Nauman, G., Smagala, A., Endres, K., and Cohen, M. (2000). The effect of treating forages with fibrolytic enzymes on its nutritive value and lactation performance of dairy cows. *Journal of Dairy Science,* 83 (*1*), 115-122.

Kusmartono, Shimada, A., and Barry, T. N. (1997). Rumen digestion and rumen outflow rate in deer fed fresh chicory (*Cichorium intybus*) or perennial ryegrass (*Lolium perenne*). *Journal of Agricultural Science,* 128 (*1*), 87-94.

Lagaert, S., Beliën, T., and Volckaert, G. (2009). Plant cell walls: Protecting the barrier from degradation by microbial enzymes. *Seminars in Cell & Developmental Biology,* 20 (*9*), 1064-1073.

Lairson, L., Henrissat, B., Davies, G., and Withers, S. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Biochemistry,* 77 (*1*), 521.

Lane, D. J. (1991). 16*S*/23*S* rRNA Sequencing. In E. Stackebrandt and M. Goodfellow (Eds.), Nucleic acid techniques in bacterial systematics. pp. 115-147. Chichester, UK. John Wiley and Sons.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9 (*4*), 357-359.

Larue, R., Yu, Z., Parisi, V. A., Egan, A. R., and Morrison, M. (2005). Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. *Environmental Microbiology,* 7 (*4*), 530-543.

Le Nours, J., De Maria, L., Welner, D., Jørgensen, C. T., Christensen, L. L., Borchert, T. V., Larsen, S., and Lo Leggio, L. (2009). Investigating the binding of β-1, 4-galactan to *Bacillus licheniformis* β-1, 4-galactanase by crystallography and computational modeling. *Proteins: Structure, Function, and Bioinformatics,* 75 (*4*), 977-989.

Leahy, S., Kelly, W., Ronimus, R., Wedlock, N., Altermann, E., and Attwood, G. (2013). Genome sequencing of rumen bacteria and archaea and its application to methane mitigation strategies. *Animal,* 7 (*Supplement 2*), 235-243.

Leedle, J., and Hespell, R. B. (1980). Differential carbohydrate media and anaerobic replica plating techniques in delineating carbohydrate-utilizing subgroups in rumen bacterial populations. *Applied and Environmental Microbiology,* 39 (*4*), 709.

Levisson, M., Han, G. W., Deller, M. C., Xu, Q., Biely, P., Hendriks, S., Ten Eyck, L. F., Flensburg, C., Roversi, P., and Miller, M. D. (2012). Functional and structural characterization of a thermostable acetyl esterase from *Thermotoga maritima*. *Proteins: Structure, Function, and Bioinformatics,* 80 (*6*), 1545-1559.

Li, Y., Irwin, D. C., and Wilson, D. B. (2007). Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Applied and Environmental Microbiology,* 73 (*10*), 3165-3172.

Liab, K., Azadi, P., Collins, R., Tolan, J., Kim, J. S., and Eriksson, K. E. L. (2000). Relationships between activities of xylanases and xylan structures. *Enzyme and Microbial Technology,* 27 (*1*), 89-94.

Linder, Å., Bergman, R., Bodin, A., and Gatenholm, P. (2003). Mechanism of assembly of xylan onto cellulose surfaces. *Langmuir,* 19 (*12*), 5072-5077.

Liu, S.-L., Hessel, A., and Sanderson, K. E. (1993). Genomic mapping with I-*Ceu*I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. *Proceedings of the National Academy of Sciences,* 90 (*14*), 6874-6878.

López-Cortés, N., Reyes-Duarte, D., Beloqui, A., Polaina, J., Ghazi, I., Golyshina, O. V., Ballesteros, A., Golyshin, P. N., and Ferrer, M. (2007). Catalytic role of conserved HQGE motif in the CE6 carbohydrate esterase family. *FEBS Letters,* 581 (*24*), 4657-4662.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research,* 25 (*5*), 955-964.

Lueders, T., Manefield, M., and Friedrich, M. W. (2004). Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environmental Microbiology,* 6 (*1*), 73-78.

Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2011). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research,* 39 (*Database issue*), D52-D57.

Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker Jr, C. T., Saxman, P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M., and Tiedje, J. M. (2001). The RDP-II (Ribosomal Database Project). *Nucleic Acids Research,* 29 (*1*), 173-174.

Mannarelli, B. M. (1988). Deoxyribonucleic acid relatedness among strains of the species *Butyrivibrio fibrisolvens*. *International Journal of Systemic Bacteriology,* 38 (*4*), 340-347.

Mannarelli, B. M., Stack, R. J., Lee, D., and Ericsson, L. (1990). Taxonomic relatedness of *Butyrivibrio*, *Lachnospira*, *Roseburia*, and *Eubacterium* species as determined by DNA hybridization and extracellular-polysaccharide analysis. *International Journal of Systematic Bacteriology,* 40 (*4*), 370-378.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., and Lanczycki, C. J. (2012). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Research,* 41 (*Database issue*), D348–D352.

Marcus, S. E., Verhertbruggen, Y., Hervé, C., Ordaz-Ortiz, J. J., Farkas, V., Pedersen, H. L., Willats, W. G., and Knox, J. P. (2008). Pectic homogalacturonan masks abundant sets of xyloglucan epitopes in plant cell walls. *BMC Plant Biology,* 8 (*60*), 1-12.

Margherita, S., and Hungate, R. (1963). Serological analysis of *Butyrivibrio* from the bovine rumen. *Journal of Bacteriology,* 86 (*4*), 855-860.

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., and Huntemann, M. (2013). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Research,* 42 (*Database issue*), D560-567.

Markowitz, V. M., Mavromatis, K., Ivanova, N. N., Chen, I.-M. A., Chu, K., and Kyrpides, N. C. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics,* 25 (*17*), 2271-2278.

Marshall, P., and Lemieux, C. (1992). The I-*Ceu*I endonuclease recognizes a sequence of 19 base pairs and preferentially cleaves the coding strand of the *Chlamydomonas moewusii* chloroplast large subunit rRNA gene. *Nucleic Acids Research,* 20 (*23*), 6401-6407.

Martens, D. A., and Loeffelmann, K. L. (2002). Improved accounting of carbohydrate carbon from plants and soils. *Soil Biology and Biochemistry,* 34 (*10*), 1393-1399.

Martin, J. A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics,* 12 (*10*), 671-682.

Master, E. R., Rudsander, U. J., Zhou, W., Henriksson, H., Divne, C., Denman, S., Wilson, D. B., and Teeri, T. T. (2004). Recombinant expression and enzymatic characterization of PttCel9A, a KOR homologue from *Populus tremula x tremuloides*. *Biochemistry,* 43 (*31*), 10080-10089.

Mathlouthi, N., Saulnier, L., Quemener, B., and Larbier, M. (2002). Xylanase, β-glucanase, and other side enzymatic activities have greater effects on the viscosity of several feedstuffs than xylanase and β-glucanase used alone or in combination. *Journal of Agricultural and Food Chemistry,* 50 (*18*), 5121-5127.

Maynard, L., Loosli, J., Hintz, H., and Warner, R. (1979). Animal Nutrition. The carbohydrates and their metabolism. 7th ed., pp. 74-103. New York, USA. McGraw-Hill, Inc.

McAllister, T. A., Bae, H. D., Jones, G. A., and Cheng, K. J. (1994). Microbial attachment and feed digestion in the rumen. *Journal of Animal Science,* 72 (*11*), 3004-3018.

McCartney, L., Gilbert, H. J., Bolam, D. N., Boraston, A. B., and Knox, J. P. (2004). Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Analytical Biochemistry,* 326 (*1*), 49-54.

McClelland, M. (1981). The effect of sequence specific DNA methylation on restriction endonuclease cleavage. *Nucleic Acids Research,* 9 (*22*), 5859-5866.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., Vanderpool, C. K., and Tjaden, B. (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research,* 41 (*14*), e140.

McSweeney, C., Dulieu, A., and Bunch, R. (1998). *Butyrivibrio* spp. and other xylanolytic microorganisms from the rumen have cinnamoyl esterase activity. *Anaerobe,* 4 (*1*), 57-65.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development,* 15 (*6*), 589-594.

Michalet-Doreau, B., Fernandez, I., Peyron, C., Millet, L., and Fonty, G. (2001). Fibrolytic activities and cellulolytic bacterial community structure in the solid and liquid phases of rumen contents. *Reproduction Nutrition Development,* 41 (*2*), 187-194.

Michel, G., Barbeyron, T., Kloareg, B., and Czjzek, M. (2009). The family 6 carbohydrate binding modules have co-evolved with their appended catalytic modules towards similar substrate specificity. *Glycobiology,* 19 (*6*), 615-623.

Mielenz, J. R. (2001). Ethanol production from biomass: technology and commercialization status. *Current Opinion in Microbiology,* 4 (*3*), 324-329.

Miller, M. E. B., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., and Coutinho, P. M. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS One,* 4 (*8*), e6650.

Minato, H., Endo, A., Ootomo , Y., and Uemura, T. (1966). Ecological treatise on the rumen fermentation. *The Journal of General and Applied Microbiology,* 12 (*1*), 53-69.

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. In I. C. Prentice and E. v. d. Maarel (Eds.), Theory and models in vegetation science. Vol. 8, pp. 89-107. Netherlands. Springer.

Miron, J. (1991). The hydrolysis of lucerne cell-wall monosaccharide components by monocultures or pair combinations of defined ruminal bacteria. *Journal of Applied Bacteriology,* 70 (*3*), 245-252.

Miron, J., and Ben-Ghedalia, D. (1992). The degradation and utilization of wheat-straw cell-wall monosaccharide components by defined ruminal cellulolytic bacteria. *Applied Microbiology and Biotechnology,* 38 (*3*), 432-437.

Miron, J., and Ben-Ghedalia, D. (1993a). Digestion of cell-wall monosaccharides of ryegrass and alfalfa hays by the ruminal bacteria *Fibrobacter succinogenes* and *Butyrivibrio fibrisolvens*. *Canadian Journal of Microbiology,* 39 (*8*), 780-786.

Miron, J., and Ben-Ghedalia, D. (1993b). Digestion of structural polysaccharides of *Panicum* and vetch hays by the rumen bacterial strains *Fibrobacter succinogenes* BL2 and *Butyrivibrio fibrisolvens* D1. *Applied Microbiology and Biotechnology,* 39 (*6*), 756-759.

Miron, J., and Ben-Ghedalia, D. (1993c). Untreated and delignified cotton stalks as model substrates for degradation and utilization of cell-wall monosaccharide components by defined ruminal cellulolytic bacteria. *Bioresource Technology,* 43 (*3*), 241-247.

Miron, J., Duncan, S. H., and Stewart, C. S. (1994). Interactions between rumen bacterial strains during the degradation and utilization of the monosaccharides of barley straw cell-walls. *Journal of Applied Bacteriology,* 76 (*3*), 282-287.

Mitchell, R. A., Dupree, P., and Shewry, P. R. (2007). A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiology,* 144 (*1*), 43-53.

Möller, S., Croning, M. D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics,* 17 (*7*), 646-653.

Montanier, C., Van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., and Ghinet, M. G. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences,* 106 (*9*), 3065-3070.

Montgomery, L., Flesher, B., and Stahl, D. (1988). Transfer of *Bacteroides succinogenes* (Hungate) to *Fibrobacter* gen. nov. as *Fibrobacter succinogenes* comb. nov. and description of *Fibrobacter intestinalis* sp. nov. *International Journal of Systematic Bacteriology,* 38 430-435.

Moon, C. D., Pacheco, D. M., Kelly, W. J., Leahy, S. C., Li, D., Kopecny, J., and Attwood, G. T. (2008). Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate-producing ruminal bacterium. *International Journal of Systematic and Evolutionary Microbiology,* 58 (*9*), 2041-2045.

Moore, W. C., Johnson, J. L., and Holdeman, L. V. (1976). Emendation of *Bacteroidaceae* and *Butyrivibrio* and descriptions of *Desulfomonas* gen. nov. and ten new species in the genera *Desulfomonas*, *Butyrivibrio*, *Eubacterium*, *Clostridium*, and *Ruminococcus*. *International Journal of Systematic Bacteriology,* 26 (*2*), 238-252.

Morgavi, D., Kelly, W., Janssen, P., and Attwood, G. (2013). Rumen microbial (meta) genomics and its application to ruminant production. *Animal,* 7 (*Supplement 1*), 184-201.

Morgavi, D. P., Newbold, C. J., Beever, D. E., and Wallace, R. J. (2000). Stability and stabilization of potential feed additive enzymes in rumen fluid. *Enzyme and Microbial Technology,* 26 (*2*), 171-177.

Morrison, M. (1996). Do ruminal bacteria exchange genetic material? *Journal of Dairy Science,* 79 (*8*), 1476-1486.

Mrázek, J. (2009). Phylogenetic signals in DNA composition: limitations and prospects. *Molecular Biology and Evolution,* 26 (*5*), 1163-1169.

Mulder, N. J., Kersey, P., Pruess, M., and Apweiler, R. (2008). *In silico* characterization of proteins: UniProt, InterPro and Integr8. *Molecular Biotechnology,* 38 (*2*), 165-177.

Nakamura, A., Furuta, H., Maeda, H., Takao, T., and Nagamatsu, Y. (2002). Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean polysaccharides. *Bioscience, Biotechnology, and Biochemistry,* 66 (*5*), 1155-1158.

Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics,* 25 (*10*), 1335-1337.

Neinhuis, C., and Edelmann, H. (1996). Methanol as a rapid fixative for the investigation of plant surfaces by SEM. *Journal of Microscopy,* 184 (*1*), 14-16.

Noel, S. (2013). Cultivation and community composition analysis of plant-adherent rumen bacteria. Doctor of Philosophy in Microbiology and Genetics, Massey University, Palmerston North, New Zealand.

Nurizzo, D., Nagy, T., Gilbert, H. J., and Davies, G. J. (2002). The structural basis for catalysis and specificity of the *Pseudomonas cellulosa* α-glucuronidase, GlcA67A. *Structure,* 10 (*4*), 547-556.

O'Neill, M. A., and York, W. S. (2003). The composition and structure of plant primary cell walls. In J. K. C. Rose (Ed.), The plant cell wall. 1st ed., pp. 1-54. Boca Raton, FL, USA. Blackwell Publishing and CRC Press.

Ochiai, A., Itoh, T., Mikami, B., Hashimoto, W., and Murata, K. (2009). Structural determinants responsible for substrate recognition and mode of action in family 11 polysaccharide lyases. *Journal of Biological Chemistry,* 284 (*15*), 10181-10189.

Orpin, C. G., Mathiesen, S. D., Greenwood, Y., and Blix, A. S. (1985). Seasonal changes in the ruminal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *Applied and Environmental Microbiology,* 50 (*1*), 144-151.

Paës, G., Skov, L. K., O'Donohue, M. J., Rémond, C., Kastrup, J. S., Gajhede, M., and Mirza, O. (2008). The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit. *Biochemistry,* 47 (*28*), 7441-7451.

Pagani, I., Liolios, K., Jansson, J., Chen, I.-M. A., Smirnova, T., Nosrat, B., Markowitz, V. M., and Kyrpides, N. C. (2012). The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research,* 40 (*Database issue*), D571-D579.

Paillard, D., McKain, N., Chaudhary, L. C., Walker, N. D., Pizette, F., Koppova, I., McEwan, N. R., Kopecny, J., Vercoe, P. E., Louis, P., and Wallace, R. J. (2007a). Relation between phylogenetic position, lipid metabolism and butyrate production by different *Butyrivibrio*-like bacteria from the rumen. *Antonie Van Leeuwenhoek,* 91 (*4*), 417-422.

Paillard, D., McKain, N., Rincon, M. T., Shingfield, K. J., Givens, D. I., and Wallace, R. J. (2007b). Quantification of ruminal *Clostridium proteoclasticum* by real-time PCR using a molecular beacon approach. *Journal of Applied Microbiology,* 103 (*4*), 1251-1261.

Palevich, N. (2011). Genome sequencing of rumen bacteria involved in lignocellulose digestion. Master of Science in Genetics and Microbiology, Massey University, Palmerston North, New Zealand.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R. E., and Blot, M. (1999). Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences,* 96 (*7*), 3807-3812.

Patterson, H., Irvin, R., Costerton, J., and Cheng, K. (1975). Ultrastructure and adhesion properties of *Ruminococcus albus*. *Journal of Bacteriology,* 122 (*1*), 278-287.

Pauly, M., Albersheim, P., Darvill, A., and York, W. S. (1999). Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *The Plant Journal,* 20 (*6*), 629-639.

Peters, A., Lebzien, P., Meyer, U., Borchert, U., Bulang, M., and Flachowsky, G. (2010). Effect of exogenous fibrolytic enzymes on ruminal fermentation and nutrient digestion in dairy cows. *Archives of Animal Nutrition,* 64 (*3*), 221-237.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* 8 (*10*), 785-786.

Pinos-Rodríguez, J., González, S., Mendoza, G., Bárcena, R., Cobos, M., Hernández, A., and Ortega, M. (2002). Effect of exogenous fibrolytic enzyme on ruminal fermentation and digestibility of alfalfa and rye-grass hay fed to lambs. *Journal of Animal Science,* 80 (*11*), 3016-3020.

Pitta, D. W., Pinchak, W. E., Dowd, S. E., Osterstock, J., Gontcharova, V., Youn, E., Dorton, K., Yoon, I., Min, B. R., and Fulford, J. (2010). Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. *Microbial Ecology,* 59 (*3*), 511-522.

Pizzi, A., and Eaton, N. (1985). The structure of cellulose by conformational analysis. 2. The cellulose polymer chain. *Journal of Macromolecular Science: Part A - Chemistry,* 22 (*1*), 105-137.

Podell, S., and Gaasterland, T. (2007). DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biology,* 8 (*2*), R16.

Polan, C. E., McNeill, J. J., and Tove, S. B. (1964). Biohydrogenation of unsaturated fatty acids by rumen bacteria. *Journal of Bacteriology,* 88 (*4*), 1056-1064.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., and Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research,* 35 (*21*), 7188-7196.

Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research,* 33 (*Database issue*), D501-D504.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., and Clements, J. (2011). The Pfam protein families database. *Nucleic Acids Research,* 40 (*Database issue*), D290-D301.

Purushe, J., Fouts, D. E., Morrison, M., White, B. A., Mackie, R. I., Coutinho, P. M., Henrissat, B., and Nelson, K. E. (2010). Comparative genome analysis of *Prevotella ruminicola* and *Prevotella bryantii:* insights into their environmental niche. *Microbial Ecology,* 60 (*4*), 721-729.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Racine, J. S. (2012). RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics,* 27 (*1*), 167-172.

Rancour, D. M., Marita, J. M., and Hatfield, R. D. (2012). Cell wall composition throughout development for the model grass *Brachypodium distachyon*. *Frontiers in Plant Science,* 3 (*266*), 1-14.

Ransom-Jones, E., Jones, D. L., McCarthy, A. J., and McDonald, J. E. (2012). The *Fibrobacteres*: an important phylum of cellulose-degrading bacteria. *Microbial Ecology,* 63 (*2*), 267-281.

Reddy, T., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E. A., and Kyrpides, N. C. (2014). The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. *Nucleic Acids Research,* 43 (*Database issue*), D1099–D1106.

Reynolds, C., Dürst, B., Lupoli, B., Humphries, D., and Beever, D. (2004). Visceral tissue mass and rumen volume in dairy cows during the transition from late gestation to early lactation. *Journal of Dairy Science,* 87 (*4*), 961-971.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics,* 16 (*6*), 276-277.

Richardson, A., Calder, A., Stewart, C., and 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.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences,* 106 (*45*), 19126-19131.

Rincon, M. T., Ding, S.-Y., McCrae, S. I., Martin, J. C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E. A., and Flint, H. J. (2003). Novel organization and divergent dockerin specificities in the cellulosome system of *Ruminococcus flavefaciens*. *Journal of Bacteriology,* 185 (*3*), 703-713.

Robinson, M. D., and Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics,* 9 (*2*), 321-332.

Roché, C., Albertyn, H., van Gylswyk, N. O., and Kistner, A. (1973). The growth response of cellulolytic acetate-utilizing and acetate-producing *Butyrivibrios* to volatile fatty acids and other nutrients. *Journal of General Microbiology,* 78 (*2*), 253-260.

Rose, P. W., Bi, C., Bluhm, W. F., Christie, C. H., Dimitropoulos, D., Dutta, S., Green, R. K., Goodsell, D. S., Prlić, A., and Quesada, M. (2013). The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Research,* 41 (*Database issue*), D475-D482.

Rumney, C. J., Duncan, S. H., Henderson, C., and Stewart, C. (1995). Isolation and characteristics of a wheatbran-degrading *Butyrivibrio* from human faeces. *Letters in Applied Microbiology,* 20 (*4*), 232-236.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics,* 16 (*10*), 944-945.

Rzhetsky, A., and Nei, M. (1992). A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution,* 9 (*5*), 945-967.

Saito, H., and Miura, K.-I. (1963). Preparation of transforming deoxyribonucleic acid by phenol treatment. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Nucleic Acids and Related Subjects,* 72 (*1*), 619-629.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution,* 4 (*4*), 406-425.

Sakai, T., Sakamoto, T., Hallaert, J., and Vandamme, E. J. (1993). Pectin, pectinase, and protopectinase: Production, properties, and applications. In S. L. Neidleman (Ed.), Advances in Applied Microbiology. Vol. 39, pp. 213-294. London, UK. Academic Press.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning. 2nd ed. New York. Cold spring harbor laboratory press.

Sarkar, N., Ghosh, S. K., Bannerjee, S., and Aikat, K. (2012). Bioethanol production from agricultural wastes: An overview. *Renewable Energy,* 37 (*1*), 19-27.

Saulnier, L., and Thibault, J. F. (1999). Ferulic acid and diferulic acids as components of sugar-beet pectins and maize bran heteroxylans. *Journal of the Science of Food and Agriculture,* 79 (*3*), 396-402.

Sawanon, S., and Kobayashi, Y. (2006). Synergistic fibrolysis in the rumen by cellulolytic *Ruminococcus flavefaciens* and non-cellulolytic *Selenomonas ruminantium*: Evidence in defined cultures. *Animal Science Journal,* 77 (*2*), 208-214.

Scalbert, A., Monties, B., Lallemand, J.-Y., Guittet, E., and Rolando, C. (1985). Ether linkage between phenolic acids and lignin fractions from wheat straw. *Phytochemistry,* 24 (*6*), 1359-1362.

Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology,* 61 (*1*), 263-289.

Scheutz, F., and Strockbine, N. A. (2005). *Escherichia*. In W. B. Whitman (Ed.), Bergey's Manual of Systematics of Archaea and Bacteria. pp. 1-49. John Wiley & Sons.

Schingoethe, D. J., Stegeman, G. A., and Treacher, R. J. (1999). Response of lactating dairy cows to a cellulase and xylanase enzyme mixture applied to forages at the time of feeding. *Journal of Dairy Science,* 82 (*5*), 996-1003.

Schloss, P. D., and Westcott, S. L. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16*S* rRNA gene sequence analysis. *Applied and Environmental Microbiology,* 77 (*10*), 3219-3226.

Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W. C., Richter, A. R., and White, O. (2007). TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Research,* 35 (*Database issue*), D260-D264.

Selinger, L., Forsberg, C., and Cheng, K.-J. (1996). The rumen: a unique source of enzymes for enhancing livestock production. *Anaerobe,* 2 (*5*), 263-284.

Sewell, G., Aldrich, H., Williams, D., Mannarelli, B., Wilkie, A., Hespell, R., Smith, P., and Ingram, L. (1988). Isolation and characterization of xylan-degrading strains of *Butyrivibrio fibrisolvens* from a Napier grass-fed anaerobic digester. *Applied and Environmental Microbiology,* 54 (*5*), 1085-1090.

Shallom, D., and Shoham, Y. (2003). Microbial hemicellulases. *Current Opinion in Microbiology,* 6 (*3*), 219-228.

Shane, B. S., Gouws, L., and Kistner, A. (1969). Cellulolytic bacteria occurring in the rumen of sheep conditioned to low-protein teff hay. *Journal of General Microbiology,* 55 (*3*), 445-457.

Sharp, P. M., and Li, W.-H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research,* 15 (*3*), 1281-1295.

Shevchik, V. E., and Hugouvieux-Cotte-Pattat, N. (2003). PaeX, a second pectin acetylesterase of *Erwinia chrysanthemi* 3937. *Journal of Bacteriology,* 185 (*10*), 3091-3100.

Shibuya, N., and Nakane, R. (1984). Pectic polysaccharides of rice endosperm cell walls. *Phytochemistry,* 23 (*7*), 1425-1429.

Shimoyama, T., Kato, S., Ishii, S. i., and Watanabe, K. (2009). Flagellum mediates symbiosis. *Science,* 323 (*5921*), 1574-1574.

Shin, H.-D., McClendon, S., Vo, T., and Chen, R. R. (2010). *Escherichia coli* binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Applied and Environmental Microbiology,* 76 (*24*), 8150-8159.

Shoemaker, N. B., Wang, G.-R., and Salyers, A. A. (1992). Evidence for natural transfer of a tetracycline resistance gene between bacteria from the human colon and

bacteria from the bovine rumen. *Applied and Environmental Microbiology,* 58 (*4*), 1313-1320.

Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999). A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity. *Structure,* 7 (*7*), 853-864.

Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *Journal of Biological Chemistry,* 275 (*52*), 41137-41142.

Sirohi, S. K., Singh, N., Dagar, S. S., and Puniya, A. K. (2012). Molecular tools for deciphering the microbial community structure and diversity in rumen ecosystem. *Applied Microbiology and Biotechnology,* 95 (*5*), 1135-1154.

Sneath, P. H., and Sokal, R. R. (1962). Numerical taxonomy. *Nature,* 193 (*4818*), 855-860.

Sørensen, H. R., Meyer, A. S., and Pedersen, S. (2003). Enzymatic hydrolysis of water-soluble wheat arabinoxylan. 1. Synergy between α-L-arabinofuranosidases, endo-1, 4-β-xylanases, and β-xylosidase activities. *Biotechnology and Bioengineering,* 81 (*6*), 726-731.

Stack, R. J. (1988). Neutral sugar composition of extracellular polysaccharides produced by strains of *Butyrivibrio fibrisolvens*. *Applied and Environmental Microbiology,* 54 (*4*), 878-883.

Staden, R., Beal, K. F., and Bonfield, J. K. (1999). The staden package. In S. Misener and S. A. Krawetz (Eds.), Bioinformatics Methods and Protocols. Vol. 132, pp. 115-130. New York, USA. Humana Press.

Statistics New Zealand. (2015). 'Infoshare - Agriculture.' Retrieved 15 March, 2014. Available from: http://www.stats.govt.nz/infoshare.

Stewart, C., Flint, H., and Bryant, M. (1997). The rumen bacteria. In P. N. Hobson and C. S. Stewart (Eds.), The rumen microbial ecosystem. 2nd ed., pp. 10-72. London, UK. Chapman and Hall.

Subramaniyan, S., and Prema, P. (2002). Biotechnology of microbial xylanases: enzymology, molecular biology, and application. *Critical Reviews in Biotechnology,* 22 (*1*), 33-64.

Suen, G., Stevenson, D. M., Bruce, D. C., Chertkov, O., Copeland, A., Cheng, J.-F., Detter, C., Detter, J. C., Goodwin, L. A., and Han, C. S. (2011a). Complete genome of the cellulolytic ruminal bacterium *Ruminococcus albus* 7. *Journal of Bacteriology,* 193 (*19*), 5574-5575.

Suen, G., Weimer, P. J., Stevenson, D. M., Aylward, F. O., Boyum, J., Deneke, J., Drinkwater, C., Ivanova, N. N., Mikhailova, N., Chertkov, O., Goodwin, L. A., Currie, C. R., Mead, D., and Brumm, P. J. (2011b). The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS One,* 6 (*4*), e18814.

Tabka, M., Herpoël-Gimbert, I., Monod, F., Asther, M., and Sigoillot, J. (2006). Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme and Microbial Technology,* 39 (*4*), 897-902.

Tajima, K., Aminov, R. I., Nagamine, T., Ogata, K., Nakamura, M., Matsui, H., and Benno, Y. (1999). Rumen bacterial diversity as determined by sequence analysis of 16*S* rDNA libraries. *FEMS Microbiology Ecology,* 29 (*2*), 159-169.

Tajima, K., Arai, S., Ogata, K., Nagamine, T., Matsui, H., Nakamura, M., Aminov, R. I., and Benno, Y. (2000). Rumen bacterial community transition during adaptation to high-grain diet. *Anaerobe,* 6 (*5*), 273-284.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* 30 (*12*), 2725-2729.

Tardy, F., Nasser, W., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. (1997). Comparative analysis of the five major *Erwinia chrysanthemi* pectate lyases: enzyme characteristics and potential inhibitors. *Journal of Bacteriology,* 179 (*8*), 2503-2511.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research,* 28 (*1*), 33-36.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014). RefSeq microbial genomes database: New representation and annotation strategy. *Nucleic Acids Research,* 42 (*Database issue*), D553-D559.

Teather, R. M. (1982). Isolation of plasmid DNA from *Butyrivibrio fibrisolvens*. *Applied and Environmental Microbiology,* 43 (*2*), 298-302.

Teeri, T. T. (1997). Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology,* 15 (*5*), 160-167.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., and Durkin, A. S. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences, USA,* 102 (*39*), 13950-13955.

Thompson, J. D., Gibson, T., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. In A. D. Baxevanis (Ed.), Current Protocols in Bioinformatics. Vol. Unit 2.3, pp. 2.3.1–2.3.22. Hoboken, NJ, USA. John Wiley and Sons.

Till, M., Goldstone, D., Card, G., Attwood, G. T., Moon, C. D., and Arcus, V. L. (2014). Structural analysis of the GH43 enzyme Xsa43E from *Butyrivibrio proteoclasticus*. *Acta Crystallographica F: Structural Biology Communications,* 70 (*9*), 1193-1198.

Tilman, D., and Clark, M. (2014). Global diets link environmental sustainability and human health. *Nature,* 515 (*7528*), 518-522.

Titi, H., and Lubbadeh, W. (2004). Effect of feeding cellulase enzyme on productive responses of pregnant and lactating ewes and goats. *Small Ruminant Research,* 52 (*1*), 137-143.

Tittsler, R. P., and Sandholzer, L. A. (1936). The use of semi-solid agar for the detection of bacterial motility. *Journal of Bacteriology,* 31 (*6*), 575.

Tjaden, B. (2015). *De novo* assembly of bacterial transcriptomes from RNA-seq data. *Genome Biology,* 16 (*1*), 1-10.

Topakas, E., Kyriakopoulos, S., Biely, P., Hirsch, J., Vafiadi, C., and Christakopoulos, P. (2010). Carbohydrate esterases of family 2 are 6-*O*-deacetylases. *FEBS Letters,* 584 (*3*), 543-548.

Topakas, E., Vafiadi, C., and Christakopoulos, P. (2007). Microbial production, characterization and applications of feruloyl esterases. *Process Biochemistry,* 42 (*4*), 497-509.

van Bueren, A. L., Morland, C., Gilbert, H. J., and Boraston, A. B. (2005). Family 6 carbohydrate binding modules recognize the non-reducing end of β-1, 3-linked

glucans by presenting a unique ligand binding surface. *Journal of Biological Chemistry,* 280 (*1*), 530-537.

van der Heide, T., and Poolman, B. (2002). ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO reports,* 3 (*10*), 938-943.

Van Gylswyk, N. O., Hippe, H., and Rainey, F. A. (1996). *Pseudobutyrivibrio ruminis* gen. nov., sp. nov., a butyrate-producing bacterium from the rumen that closely resembles *Butyrivibrio fibrisolvens* in phenotype. *International Journal of Systematic Bacteriology,* 46 (*2*), 559-563.

Van Soest, P. J. (1994). Nutritional ecology of the ruminant. 2nd ed. London, UK. Cornell University Press.

Van Soest, P. v., Robertson, J., and Lewis, B. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *Journal of Dairy Science,* 74 (*10*), 3583-3597.

Varner, J. E., and Lin, L.-S. (1989). Plant cell wall architecture. *Cell,* 56 (*2*), 231-239.

Vazana, Y., Moraïs, S., Barak, Y., Lamed, R., and Bayer, E. A. (2012). Designer cellulosomes for enhanced hydrolysis of cellulosic substrates. *Methods in Enzymology,* 510 429-452.

Vesth, T., Lagesen, K., Acar, O., and Ussery, D. (2013). CMG-biotools, a free workbench for basic comparative microbial genomics. *PLoS One,* 8 (*4*), e60120.

Villas-Bôas, S. G., Esposito, E., and Mitchell, D. A. (2002). Microbial conversion of lignocellulosic residues for production of animal feeds. *Animal Feed Science and Technology,* 98 (*1*), 1-12.

Villas-Bôas, S. G., Noel, S., Lane, G. A., Attwood, G., and Cookson, A. (2006). Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media. *Analytical Biochemistry,* 349 (*2*), 297-305.

Vogel, J. (2008). Unique aspects of the grass cell wall. *Current Opinion in Plant Biology,* 11 (*3*), 301-307.

Vrba, E. S., and Schaller, G. B. (2000). Antelopes, deer, and relatives: Fossil record, behavioural ecology, systematics, and conservation. New Haven, CT, USA Yale University Press.

Waghorn, G. C. (2002). Can forages match concentrate diets for dairy production? Proceedings of the New Zealand Society of Animal Production. Vol. 62, pp. 261-266. Palmerston North, NZ. New Zealand Society of Animal Production.

Waghorn, G. C., Burke, J. L., and Kolver, E. S. (2007). Principles of feeding value. In P. V. Rattray, I. M. Brookes and A. M. Nicol (Eds.), Pasture and supplements for grazing animals. Vol. 14, pp. 35-59. Hamilton, NZ. New Zealand Society of Animal Production.

Waghorn, G. C., and Clark, D. A. (2004). Feeding value of pastures for ruminants. *New Zealand Veterinary Journal,* 52 (*6*), 320-331.

Wallace, R., Wallace, S., McKain, N., Nsereko, V., and Hartnell, G. (2001). Influence of supplementary fibrolytic enzymes on the fermentation of corn and grass silages by mixed ruminal microorganisms *in vitro*. *Journal of Animal Science,* 79 (*7*), 1905-1916.

Wallace, R. J., Chaudhary, L. C., McKain, N., McEwan, N. R., Richardson, A. J., Vercoe, P. E., Walker, N. D., and Paillard, D. (2006). *Clostridium proteoclasticum*: a ruminal bacterium that forms stearic acid from linoleic acid. *FEMS Microbiology Letters,* 265 (*2*), 195-201.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association,* 58 (*301*), 236-244.

Warton, D. I., Wright, S. T., and Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution,* 3 (*1*), 89-101.

Webb, E. C. (1992). Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes. 6th ed. Academic Press.

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.

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16*S* ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology,* 173 (*2*), 697-703.

Whitford, M. F., Forster, R. J., Beard, C. E., Gong, J., and Teather, R. M. (1998). Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16*S* rRNA genes. *Anaerobe,* 4 (*3*), 153-163.

Widdel, F., Kohring, G. W., and Mayer, F. (1983). Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. *Archives of Microbiology,* 134 (*4*), 286-294.

Wilkie, K. C. B. (1985). New perspectives on non-cellulosic cell-wall polysaccharides (hemicellulose and pectic substances) of land plants. In C. T. Brett and J. R. Hillman (Eds.), Biochemistry of plant cell walls. Vol. 28, pp. 1-39. Cambridge, UK. Cambridge University Press.

Willems, A., Amat-Marco, M., and Collins, M. D. (1996). Phylogenetic analysis of *Butyrivibrio* strains reveals three distinct groups of species within the *Clostridium* subphylum of the gram-positive bacteria. *International Journal of Systematic Bacteriology,* 46 (*1*), 195-199.

Williams, A. G., and Withers, S. E. (1992). Induction of xylan-degrading enzymes in *Butyrivibrio fibrisolvens*. *Current Microbiology,* 25 (*5*), 297-303.

Williams, Y. J., Walker, G. P., Wales, W. J., and Doyle, P. T. (2001). Effect of pasture type and intake on the rumen pH of grazing dairy cows in spring. *Australian Journal of Dairy Technology,* 56 (*2*), 163.

Womble, D. D. (1999). GCG: The Wisconsin package of sequence analysis programs. In S. Misener and S. A. Krawetz (Eds.), Bioinformatics Methods and Protocols. pp. 3-22. New York, USA. Humana Press.

Woodfield, D. (1999). Genetic improvements in New Zealand forage cultivars. Paper presented at the: *Proceedings of the New Zealand Grassland Association*.

Wu, S., Baldwin, R. L., Li, W., Li, C., Connor, E. E., and Li, R. W. (2012). The bacterial community composition of the bovine rumen detected using pyrosequencing of 16*S* rRNA genes. *Metagenomics,* 1 (*Article ID 235571*), 1-11.

Xia, Z.-x., and Mathews, F. S. (1990). Molecular structure of flavocytochrome b_2 at 24 Å resolution. *Journal of Molecular Biology,* 212 (*4*), 837-863.

Yang, H. J., and Xie, C. Y. (2010). Assessment of fibrolytic activities of 18 commercial enzyme products and their abilities to degrade the cell wall fraction of corn stalks in *in vitro* enzymatic and ruminal batch cultures. *Animal Feed Science and Technology,* 159 (*3-4*), 110-121.

Yang, W., Beauchemin, K., and Rode, L. (2000). A comparison of methods of adding fibrolytic enzymes to lactating cow diets. *Journal of Dairy Science,* 83 (*11*), 2512-2520.

Yapo, B. M. (2011). Pectic substances: From simple pectic polysaccharides to complex pectins-A new hypothetical model. *Carbohydrate Polymers,* 86 (*2*), 373-385.

Yeoman, C. (2009). The auxiliary replicons of *Butyrivibrio proteoclasticus.* Doctorate of Philosophy, Massey University, Palmerston North, New Zealand.

Yeoman, C. J., Kelly, W. J., Rakonjac, J., Leahy, S. C., Altermann, E., and Attwood, G. T. (2011). The large episomes of *Butyrivibrio proteoclasticus* B316^{T} have arisen through intragenomic gene shuttling from the chromosome to smaller *Butyrivibrio*-specific plasmids. *Plasmid,* 66 (*2*), 67-78.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research,* 40 (*Web server issue*), W445-W451.

Yoshida, K., and Komae, K. (2006). A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant and Cell Physiology,* 47 (*11*), 1541-1554.

Yu, P., McKinnon, J. J., Maenz, D. D., Olkowski, A. A., Racz, V. J., and Christensen, D. A. (2003). Enzymic release of reducing sugars from oat hulls by cellulase, as influenced by *Aspergillus* ferulic acid esterase and *Trichoderma* xylanase. *Journal of Agricultural and Food Chemistry,* 51 (*1*), 218-223.

Zablackis, E., Huang, J., Muller, B., Darvill, A. G., and Albersheim, P. (1995). Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. *Plant Physiology,* 107 (*4*), 1129-1138.

Zapala, M. A., and Schork, N. J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences,* 103 (*51*), 19430-19435.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using *De Bruijn* graphs. *Genome Research,* 18 (*5*), 821-829.

Zhou, J., Bao, L., Chang, L., Liu, Z., You, C., and Lu, H. (2012). β-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans. *Letters in Applied Microbiology,* 54 (*2*), 79-87.

Zhou, M., and Zhu, F. (2011). Protein biological function clustering with GO and KO. Paper presented at the: *5th International Conference on Bioinformatics and Biomedical Engineering (iCBBE).*

ZoBell, D. R., Wiedmeier, R., Olson, K., and Treacher, R. (2000). The effect of an exogenous enzyme treatment on production and carcass characteristics of growing and finishing steers. *Animal Feed Science and Technology,* 87 (*3*), 279-285.

Zuroff, T. R., and Curtis, W. R. (2012). Developing symbiotic consortia for lignocellulosic biofuel production. *Applied Microbiology and Biotechnology,* 93 (*4*), 1423-1435.