

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

Genome Sequencing of Rumen Bacteria Involved in Lignocellulose Digestion

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

Master of Science

in the

Institute of Molecular Biosciences

At Massey University, Palmerston North,

New Zealand.



MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA

Nikola Palevich

2011

Abstract

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

Acknowledgements

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

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

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

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

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

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

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

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

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

Table of Contents

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

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

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

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

Figures

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

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

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

Tables

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

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

Abbreviations

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