

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.

The auxiliary replicons of *Butyrivibrio proteoclasticus*

**A Thesis presented in fulfilment of the Doctorate of Philosophy degree
at Massey University, Palmerston North, New Zealand.**

**Carl Yeoman
2009**

Table of Contents

List of Figures.....	xi
List of Tables.....	xiii
Abstract.....	xvi
Acknowledgements.....	xvii
Dedication.....	xviii
Abbreviations.....	xix

Chapter 1	Literature review.....	1
1.1	Ruminant animals.....	1
1.2	Cellulose, Hemicellulose and Lignin.....	1
1.3	The reticulo-rumen.....	2
1.4	<i>Butyrivibrio</i>	4
1.6	<i>Butyrivibrio proteoclasticus</i> B316 ^T	5
1.7	Genome sequencing.....	7
1.8	Gene prediction, annotation and analysis.....	10
1.9	Plasmids	13
1.10	Megaplasms	14
1.11	Miniature, Minor or Secondary chromosomes	18
1.12	Plasmid replication.....	19
1.13	Plasmid conjugative transfer.....	23
1.14	Plasmid vectors.....	24
1.15	Shuttle vectors.....	25
Chapter 2	Materials and methods.....	27
2.1	Materials.....	27
2.1.1	Agarose.....	27
2.1.2	Antibiotics.....	27
2.1.3	Bacterial strains.....	27
2.1.4	Buffers and solutions	28
2.1.4.1	Acridine orange curing solution.....	28

2.1.4.2	Acriflavine curing solution.....	28
2.1.4.3	Alkaline lysis solutions I, II and III.....	28
2.1.4.4	Ammonium acetate solution.....	29
2.1.4.5	CE buffer.....	29
2.1.4.6	Chloroform.....	29
2.1.4.7	Colony hybridisation lysis solution.....	29
2.1.4.8	Conjugation buffer.....	30
2.1.4.9	Denaturation solution.....	30
2.1.4.10	Diethylpyrocarbonate.....	30
2.1.4.11	Depurination solution.....	30
2.1.4.12	Deoxynucleoside triphosphates solution.....	30
2.1.4.13	EC buffer.....	30
2.1.4.14	EDTA-Sarkosyl solution.....	30
2.1.4.15	EDTA solution.....	30
2.1.4.16	Electroporation buffer.....	31
2.1.4.17	Ethanol.....	31
2.1.4.18	Ethidium bromide.....	31
2.1.4.19	Ethidium bromide curing solution.....	31
2.1.4.20	Glycerol solutions.....	31
2.1.4.21	Hybridisation solution.....	31
2.1.4.22	IPTG.....	32
2.1.4.23	Isopropanol.....	32
2.1.4.24	Liquid nitrogen.....	32
2.1.4.25	Microarray pre-hybridisation buffer.....	32
2.1.4.26	Microarray wash solutions.....	32
2.1.4.27	Mineral solution.....	33
2.1.4.28	NaCl solution.....	33
2.1.4.29	Neutralisation solution.....	33
2.1.4.30	Novobiocin curing solution.....	33
2.1.4.31	Pfennigs heavy metal solution.....	33
2.1.4.32	Phenol.....	33
2.1.4.33	Phenol:Chloroform.....	33
2.1.4.34	Phenol:Chloroform:Isoamyl alcohol.....	34
2.1.4.35	PMSF stock solution.....	34

2.1.4.36	Potassium acetate solution.....	34
2.1.4.37	R1 salts solution.....	34
2.1.4.38	Reducing agent.....	34
2.1.4.39	Rumen fluid.....	34
2.1.4.40	Saline:EDTA solution.....	35
2.1.4.41	SDS solution.....	35
2.1.4.42	SDS curing solution.....	35
2.1.4.43	Saline sodium citrate.....	35
2.1.4.44	Sodium acetate solution.....	35
2.1.4.45	SSPE solution.....	35
2.1.4.46	STE buffer.....	35
2.1.4.47	TAE buffer.....	36
2.1.4.48	TBE buffer.....	36
2.1.4.49	TE buffers.....	36
2.1.4.50	TES buffer.....	36
2.1.4.51	Trace element solution.....	36
2.1.4.52	TRIzol.....	37
2.1.4.53	Vitamin solution.....	37
2.1.4.54	Volatile fatty acid solution.....	38
2.1.4.55	Wash solution for PFGE.....	38
2.1.4.56	X-Gal.....	38
2.1.4.57	Xylan solution.....	39
2.1.5	Enzymes.....	39
2.1.5.1	Calf Intestinal Alkaline Phosphatase.....	39
2.1.5.2	Lysozyme.....	39
2.1.5.3	Proteinase K.....	39
2.1.5.4	Restriction endonucleases.....	39
2.1.5.5	Ribonuclease A.....	39
2.1.5.6	T4 DNA ligase.....	40
2.1.5.7	T4 DNA polymerase.....	40
2.1.6	Gel migration size standards.....	40
2.1.7	Glassware.....	40
2.1.8	Laboratory equipment.....	41
2.1.9	Media.....	43

2.1.9.1	BY+ medium.....	43
2.1.9.2	DM Arabinose medium.....	44
2.1.9.3	GYT medium.....	45
2.1.9.4	Lauria-Bertini medum.....	45
2.1.9.5	M704 medium.....	45
2.1.9.6	RGM medium.....	46
2.1.9.7	SOC medium.....	47
2.1.10	Microarrays.....	47
2.1.11	Software.....	48
2.1.12	Vectors.....	48
2.2	Methods.....	52
2.2.1	Growth Conditions.....	52
2.2.2	Culture purity.....	52
2.2.2.1	Wet mounts.....	52
2.2.2.2	Gram stain.....	52
2.2.3	Growth curves.....	53
2.2.3.1	Thoma slide counts.....	53
2.2.4	Pulsed-field gel-electrophoresis.....	54
2.2.4.1	DNA extraction for PFGE.....	54
2.2.4.2	RE digestion of PFGE plugs.....	54
2.2.4.3	PFGE.....	55
2.2.5	Genome sequencing and gap closure.....	56
2.2.5.1	DNA extraction.....	56
2.2.5.2	Phenol:Chloroform extraction.....	57
2.2.5.3	Nucleic acid quantification.....	57
2.2.5.4	DNA concentration.....	57
2.2.5.5	Primer design.....	58
2.2.5.6	PCR.....	58
2.2.5.7	Multiplex PCR.....	59
2.2.5.8	Inverse PCR.....	60
2.2.5.9	RE digestion of DNA.....	61
2.2.5.10	DNA ligation.....	61
2.2.5.11	Long range PCR.....	61

2.2.5.12	Agarose gel-electrophoresis.....	62
2.2.5.13	Agarose gel elution.....	63
2.2.5.14	PCR clean up.....	63
2.2.5.15	DNA sequencing.....	64
2.2.6	Gene finding and annotation.....	64
2.2.7	DNA sequence analysis.....	65
2.2.8	Amino acid sequence analysis.....	66
2.2.9	Plasmid curing.....	66
2.2.9.1	Colony hybridisation.....	66
2.2.9.2	Southern hybridisation.....	67
2.2.9.3	Culture storage.....	68
2.2.10	Determining plasmid copy number.....	68
2.2.10.1	Preparing electrocompetent cells.....	71
2.2.10.2	TOPO cloning.....	71
2.2.10.3	Plasmid mini preparations.....	71
2.2.10.4	Realtime qPCR.....	72
2.2.11	Co-culture vs Monoculture microarray analysis.....	73
2.2.11.1	Growth conditions.....	73
2.2.11.2	Microarray analysis.....	74
2.2.11.3	Avoiding RNase contamination.....	75
2.2.11.4	RNA extraction.....	75
2.2.11.5	RNA purification.....	76
2.2.11.6	RNA quality analysis.....	76
2.2.11.7	Concentrating RNA or cDNA samples.....	77
2.2.11.8	First strand cDNA synthesis.....	77
2.2.11.9	cDNA purification.....	78
2.2.11.10	Enumeration of organisms in co-culture.....	78
2.2.11.11	cDNA labelling.....	79
2.2.11.12	Microarray hybridisation.....	79
2.2.11.13	Data acquisition.....	80
2.2.11.14	Quality control and normalisation.....	81
2.2.11.15	Statistical analysis.....	82

Chapter	3	Sequencing of <i>B. proteoclasticus</i> B316 ^T episomes	83
	3.1	Introduction.....	93
	3.2	Identifying episomal contigs within the Phase I genome sequence.....	93
	3.3	Gap closure of episomal DNAs	97
	3.3.1	Multiplex PCR.....	98
	3.3.2	BAC clone screening.....	93
	3.3.3	Conventional and long-range PCR.....	93
	3.3.4	Inverse PCR.....	94
	3.3.5	454 sequencing.....	94
	3.3.6	Reassembly.....	94
	3.4	Confirmation of two ribosomal RNA operons.....	97
	3.5	Confirmation of assembly.....	100
	3.6	Quality control of replicon sequences.....	103
	3.7	Discussion.....	106
	3.8	Summary.....	110
Chapter 4		pCY360.....	111
	4.1	Introduction.....	111
	4.2	Sequence analysis of pCY360.....	111
	4.3	pCY360 origin of replication.....	112
	4.4	Conjugative transfer-related proteins.....	113
	4.5	pCY360s predicted impact on the membrane and extracellular environment.....	118
	4.6	pCY360 contains genes of the Minimal Gene Set.....	118
	4.7	Predicted contributions to enzymatic pathways.....	119
	4.8	Transposases.....	119
	4.9	Phylogenetic relationship of repB genes.....	121
	4.10	Codon usage of <i>B. proteoclasticus</i> replicons.....	122
	4.11	Attempts to cure pCY360.....	127
	4.11.1	Determining maximal sublethal levels.....	127
	4.11.2	Determining generation times under curing conditions.....	127
	4.11.3	Evaluation of megaplasmid loss.....	127

4.12	Copy number of pCY360.....	131
4.12.1	qPCR optimisation.....	131
4.12.2	Ensuring reaction integrity.....	131
4.12.3	Absolute copy numbers.....	132
4.13	Microarray analysis of coculture with the rumen methanogen, <i>Methanobrevibacter ruminantium</i>	136
4.13.1	Microarray hybridisation and scanning.....	138
4.13.2	Microarray analysis.....	143
4.14	The distribution of large replicons in other <i>Butyrivibrio</i> / <i>Pseudobutyrvibrio</i> species.....	160
4.14.1	Relatedness of pCY360 to auxiliary replicons from other <i>Butyrivibrio</i> species.....	160
4.15	Discussion.....	162
4.15.1	Replication of pCY360.....	162
4.15.2	Minimal gene set ORFs.....	162
4.15.3	Unique enzymatic contributions.....	164
4.15.4	<i>oriT</i> and Mob proteins.....	165
4.15.5	pCY360 can potentially influence the membrane topology.....	167
4.15.6	Transposases.....	167
4.15.7	RepB phylogeny and codon usage.....	168
4.15.8	Is pCY360 an essential part of the <i>B. proteoclasticus</i> genome?.....	168
4.15.9	Microarray analysis.....	169
4.15.10	Distribution and relatedness of auxiliary replicons in other <i>Butyrivibrio</i> spp.....	173
4.16	Summary.....	174
Chapter 5	The secondary chromosome of <i>Butyrivibrio proteoclasticus</i>	175
5.1	Introduction.....	175
5.2	Sequence analysis of BPc2.....	176
5.3	The BPc2 origin of replication.....	177
5.4	Energy metabolism.....	181
5.5	Putative detoxification functions encoded by BPc2.....	184

5.6	Chemotaxis and flagella formation.....	186
5.7	Cofactor and vitamin uptake and metabolism.....	187
5.8	Ribosomal RNAs and Transfer RNAs.....	188
5.9	BPc2 contains genes described in the bacterial minimal gene set.....	189
5.10	Transposases.....	189
5.11	Maintenance of BPc2 under curing conditions.....	191
5.12	Copy number of BPc2.....	191
5.13	Microarray analysis of BPc2 gene expression in monoculture versus in coculture with <i>M. ruminantium</i>	195
5.14	Distribution of rRNA operons within the <i>Butyrivibrio</i> assemblage.....	200
5.15	Discussion.....	202
5.15.1	BPc2 replication.....	202
5.15.2	BPc2 is a secondary chromosome.....	202
5.15.3	The role of BPc2 in energy metabolism.....	203
5.15.4	The contribution of BPc2 to nitrogen metabolism.....	206
5.15.5	BPc2 genes involved in other forms of energy metabolism.....	208
5.15.6	The role of BPc2 in cellular homeostasis.....	209
5.15.7	Chemotaxis and flagella-related proteins.....	211
5.15.8	Vitamin and Cofactor metabolism.....	212
5.15.9	Other genes uniquely encoded by BPc2.....	214
5.15.10	Microarray analysis.....	215
5.16	Summary.....	218
Chapter 6	pCY186.....	219
6.1	Introduction.....	219
6.2	Sequence analysis of pCY186.....	220
6.3	The replication origin of pCY186.....	222
6.4	DNA metabolism.....	223
6.5	Restriction modification.....	223
6.6	pCY186 contains genes described in the minimal gene set..	223

6.7	Transposases.....	223
6.8	Attempts to cure pCY186.....	226
6.8.1	Morphological differences.....	226
6.8.2	Growth rate.....	227
6.9	Copy number.....	227
6.10	Microarray analysis.....	232
6.11	Discussion.....	235
6.11.1	Replication of pCY186.....	235
6.11.2	DNA metabolism.....	236
6.11.3	Restriction modification.....	237
6.11.4	ORFs from the minimal gene set.....	237
6.11.5	Transposases.....	238
6.11.6	Dispensibility.....	238
6.11.7	Microarray analysis.....	239
6.12	Summary.....	240
Chapter 7	General discussion, conclusion and Future directions.....	241
7.1	Introduction.....	241
7.2	Comparing replication machinery.....	242
7.3	Evolutionary aspects.....	248
7.4	Contributions to <i>B. proteoclasticus</i>	249
7.5	Distribution of auxiliary replicons amongst the <i>Butyrivibrio</i> / <i>Pseudobutyrvibrio</i> assemblage.....	251
7.6	Looking forward.....	254
Appendix I	Gram-stain.....	257
Appendix II	Primers.....	258
Appendix III	Supporting microarray data.....	263
Appendix IV	Programming codes.....	271
Appendix V	Gene list.....	281
References		299

List of Figures

Figure 1.1	Phylogenetic placement of <i>Butyrivibrio</i> / <i>Pseudobutyrvibrio</i> spp. according to 16s rDNA	6
Figure 2.1	Multiplex PCR notation.....	59
Figure 2.2	iPCR notation.....	60
Figure 3.1	Phylogenetic placement of Par proteins.....	86
Figure 3.2	Self alignment of Contig 67 reveals overlapping ends.....	87
Figure 3.3	Phred quality scores of contig ends.....	92
Figure 3.4	Restriction enzymes and PCR primer locations in contig ends as used for iPCR.	95
Figure 3.5	Final sequence coverage of each replicon.....	98
Figure 3.6	Confirming the presence of two rRNA operons upon BPc2.	99
Figure 3.7	Restriction mapping of replicons.....	102
Figure 3.8	Re-sequencing of potential frame shift in ORF 187.....	104
Figure 3.9	Re-sequencing of a stop codon in ORF 335.....	105
Figure 4.1	ORF and encoded protein composition (pCY360).....	114
Figure 4.2	Map of pCY360 predicted ORF function.....	115
Figure 4.3	Conserved sequence motifs in Gram positive relaxases.....	116
Figure 4.4	<i>OriT</i> candidate sequences.....	117
Figure 4.5	Phylogenetic tree of RepB proteins.....	124
Figure 4.6	Codon usage comparison.....	126
Figure 4.7	Growth curves of <i>B. proteoclasticus</i> exposed to maximal sub- lethal limits of curing agents or temperature.....	129
Figure 4.8	Optimisation of qPCR reactions.....	133
Figure 4.9	Standard curves for qPCR reactions.....	134
Figure 4.10	<i>Confirmation</i> of qPCR amplification specificities of the chromosome primer set (A) and pCY360 primer set (B).....	135
Figure 4.11	Observation of interspecies interactions between <i>B. proteoclasticus</i> and <i>Methanobrevibacter ruminantium</i>	137

Figure 4.12	Electropherograms and electrophoretic gel-translation of total RNA extracts.....	139
Figure 4.13	RT-qPCR amplification efficiencies and specificities.....	142
Figure 4.14	Low quality regions of microarray slides ‘bad-flagged’ prior to analysis.....	144
Figure 4.15	Distribution of background and foreground intensities.....	145
Figure 4.16	Spatial distribution plots of background and foreground intensities.....	146
Figure 4.17	MA plots for all analysed features.....	148
Figure 4.18	Density plots (Raw data).....	150
Figure 4.19	Density plots by slide (normalised data).....	151
Figure 4.20	Distribution of feature intensities by category.....	152
Figure 4.21	Differential regulation.....	159
Figure 4.22	Distribution and relatedness of large extra-chromosomal replicons in <i>Butyrivibrio</i> / <i>Pseudobutyrvibrio</i> spp.....	161
Figure 5.1	ORF and encoded protein composition (BPc2).....	179
Figure 5.2	<i>In-silico</i> map of BPc2.....	180
Figure 5.3	N-Glycan degradation capacity of <i>B. proteoclasticus</i>	186
Figure 5.4	Optimisation of qPCR reactions.....	192
Figure 5.5	Confirmation of qPCR amplification specificities of BPc2 primer set.....	193
Figure 5.6	Standard curves of qPCR reactions.....	194
Figure 5.7	Differential regulation of BPc2-located genes during co-culture of <i>B. proteoclasticus</i> with <i>M. ruminantium</i>	200
Figure 5.8	Distribution of rRNA operons amongst megaplasmid-like replicons in <i>Butyrivibrio</i> / <i>Pseudobutyrvibrio</i> spp.....	201
Figure 6.1	ORF and encoded protein composition (pCY186).....	221
Figure 6.2	Compositional map of pCY186.....	224
Figure 6.3	Analaysis of Δ pCY186.....	228
Figure 6.4	Optimisation of qPCR reactions.....	229
Figure 6.5	Confirmation of qPCR amplification specificites.....	230
Figure 6.6	Standard curve for the qPCR reactions.....	231

Figure 6.7	Differential pCY186 gene regulation during co-culture with <i>M. Ruminantium</i>	234
Figure 7.1	Comparison of the replication loci of <i>B. proteoclasticus</i> ' auxiliary replicons.....	245
Figure 7.2	Comparison of replication origins.....	247
Figure 7.3	ACT comparison of gene synteny of <i>B. proteoclasticus</i> ' auxiliary replicons.....	252
Figure 7.4	COG category distributions.....	253
Figure A1	Typical Gram-stain of wild-type <i>B. proteoclasticus</i> grown in M704 broth media.....	257
Figure A2	Growth curve of <i>B. proteoclasticus</i> in BY+ media supplemented with 0.2% Xylan.....	263
Figure A3	GC analysis of H ₂ and CH ₄ in co-culture.....	263
Figure A4	Microarray scans.....	266

List of Tables

Table 2.1	Bacterial strains used.....	27
Table 2.2	Components of alkaline lysis solution.....	29
Table 2.3	Microarray wash solution compositions.....	32
Table 2.4	TE buffer components.....	36
Table 2.5	Trace element solution components.....	37
Table 2.6	Vitamin solution components.....	38
Table 2.7	Gel migration standards.....	40
Table 2.8	Centrifuge specifications.....	41
Table 2.9	Centrifuge tubes and suppliers.....	41
Table 2.10	Shakers, incubators and water baths.....	42
Table 2.11	BY+ medium.....	44
Table 2.12	DMA medium.....	45
Table 2.13	M704 medium.....	46
Table 2.14	RGM medium.....	47
Table 2.15	Arabidopsis control probes.....	48
Table 2.16	Vectors.....	48

Table 2.17	Software details.....	49
Table 2.18	PCR master mix.....	59
Table 2.19	RE reaction constituents.....	61
Table 2.20	Long range PCR reaction constituents.....	62
Table 2.21	Realtime PCR reaction constituents.....	73
Table 2.22	First strand cDNA synthesis reaction constituents.....	77
Table 3.1	Identification of contigs encoding plasmid replication-related proteins	85
Table 3.2	Contigs assigned to each replicon.....	89
Table 3.3	Sequence and physical gaps within replicon DNAs and their methods of closure.....	96
Table 4.1	pCY360 genes of the bacterial minimal gene set.....	120
Table 4.2	pCY360 ORFs that encode enzymes assigned a unique EC number.....	123
Table 4.3	Maximum sublethal limits.....	128
Table 4.4	Percentage contribution of cDNAs from either <i>B. proteoclasticus</i> or <i>M. ruminantium</i> to Co-culture RNA extracts.....	143
Table 4.5	Observed transcriptional differences in co-culture.....	155
Table 4.6	pCY360 ORFs found to be significantly (FDR < 0.05) up- or down-regulated greater than 2 fold.....	158
Table 5.1	BPc2-encoded proteins involved in energy metabolism.....	183
Table 5.2	BPc2-encoded proteins involved in detoxification.....	187
Table 5.3	BPc2-encoded proteins involved in chemotaxis.....	188
Table 5.4	BPc2-encoded proteins involved in cofactor and vitamin metabolism.....	189
Table 5.5	Genes of the bacterial minimal gene set found on BPc2.....	191
Table 5.6	Differential regulation of BPc2 ORFs during co-culture.....	198
Table 6.1	pCY186 ORFs described in the bacterial minimal gene set..	225

Table 6.2	pCY186 ORFs significantly upregulated in co-culture.....	233
Table 7.1	Sequence identity of replication machinery.....	245
Table A1	Primer details.....	258
Table A2	Spectrophotometric analysis of total RNA samples to determine purity and concentration.....	264
Table A3	Mixing schedule for mono-cultures.....	264
Table A4	Microarray scan levels.....	265
Table A5	Microarray grid settings.....	269
Table A6	Feature weightings.....	270
Table A7	Gene list.....	281

Abstract

Butyrivibrio proteoclasticus B316^T is the most recently described species of the *Butyrivibrio* / *Pseudobutyrvibrio* assemblage and now the first to have its genome sequenced. The genome of this organism was found to be spread across four replicons: a 3.5 Mb major chromosome and three additional large replicons: 186, 302 and 361 Kb in size. This thesis describes the sequencing, analysis, annotation and initial characterisation of all three *B. proteoclasticus* auxiliary replicons. Most significantly, these analyses revealed that the 302-Kb replicon is a second chromosome. This small chromosome, named BPc2, encodes essential systems for the uptake and/or biosynthesis of biotin and nicotinamide adenine mononucleotide, as well as the enzymes required for utilisation of fumarate as the terminal electron acceptor during anaerobic respiration, none of which are found on the main chromosome. In addition, BPc2 contains two complete rRNA operons, a large number of enzymes involved in the metabolism of carbohydrates, nitrogen and fatty acids. In contrast to BPc2, both megaplasmsids appear largely cryptic, collectively encoding 421 genes not previously described in public databases. Nevertheless, only the 186-Kb, but not 361-Kb megaplasmsid, could be cured from *Butyrivibrio proteoclasticus* B316^T. The largest megaplasmsid has a copy number of 5, while all other replicons are present at a copy number of 1. %GC content and codon usage analyses strongly suggests that all three auxiliary replicons have co-resided with the major chromosome for a significant evolutionary period. Moreover, the replication machineries of these three replicons are conserved. Interestingly, a survey of a number of *Butyrivibrio* / *Pseudobutyrvibrio* species revealed that the megaplasmsids are widespread in this assemblage, however these other large plasmids do not show concordance with their 16S rRNA phylogeny and appear distinct to those of *B. proteoclasticus* B316^T.

A microarray analysis of gene expression in a co-culture experiment between *B. proteoclasticus* and the important ruminal methanogen, *Methanobrevibacter ruminantium* M1, revealed a potentially mutualistic interspecies interaction. In this relationship *M. ruminantium* appears to provide *B. proteoclasticus* with glutamate, essential to the final step of NAD⁺ biosynthesis, while *B. proteoclasticus* appears to provide *M. ruminantium* with formate, hydrogen and carbon dioxide, each important substrate for methanogenesis.

Acknowledgements

It is a privilege and an honour to acknowledge the many great people who have guided me throughout this project in one way or another. But more importantly have helped to shape me into the person that I am today.

To my supervisors Dr Graeme Attwood, Dr William (Bill) Kelly, and Dr Jasna Rakonjac, it is impossible to put into words the sincere gratitude I have for you all. I certainly wouldn't have lasted long without your influence and direction.

To the great people of AgResearch for the practical and financial support, without either this project would not have been possible.

To the kind people of Meat and Wool New Zealand, in particular Alan Frazer, who have funded, essentially, my life throughout not only this Ph.D. but the preceding MSc work.

To the Tertiary Education Commission for their financial support with this Ph.D.

To Dr Keith Joblin, I am eternally grateful to you and Robert Skipp who collectively played a role in bringing me into the AgResearch family and in doing so keeping me in the field of science.

To the Rumen microbiology laboratory, particularly, Dr Sinead Leahy for your help with the assembly, Dr Eric Altermann for your help with all things computational, Zhan-hou Kong and Sam Taylor for your help with the microarrays, Dong Lee for being my go to guy, Diana Pacheco for being my go to gal, Dr Adrian Cookson, Dr Ron Ronimus, Dr Peter Janssen, Graham Naylor, Dr Christina Moon, Carrie Sang, Rechelle Perry, Gemma Henderson, Catherine Tootill, and the ex-pats Dr Lucy Skillman, Dr Nicola Walker, Dr Mathew Nicholson, Dr Karen Olsen, Dr Hassan Husein, Paul Evans, Sam Noel and Nikki Kenters for your friendship.

To my parents Trish and Jim Yeoman thanks for all your support and encouragement.

To my beautiful wife Casey Norris-Yeoman thanks for sticking by me and supporting me while I pursue my dreams it has been a long and tough road but hopefully will be worth it in the long run. Love always!

Dedication

This thesis is dedicated to the most beautiful girls in the world my daughters
Summer Ashlee Pamela Yeoman and Sienna Caitlyn Estelle Yeoman.

Whatever you need, whenever you need it, I'll always be there for you both!!

Abbreviations

AWGS	Alan Wilson Centre Genome Sequencing
BAC	Bacterial Artificial chromosome
BER	BLAST-extend-repraze
BSA	Bovine Serum Albumin
bp	Base pair
CDS	Coding sequence
Contigs	Contiguous sequences
DR	Direct repeat
dso	Double-stranded origin
FDR	False discovery rate
g	Gravity
GH	Glycosyl hydrolase
HMM	Hidden Markov-model
IR	Inverted repeat
IVR	Inverse repeat
Kb	Kilobase pair
Mb	Megabase pair
Mpf	Mating pair formation complex
nt	Nucleotide(s)
OD	Optical Density
ORF	Open reading frame
PARP	Poly(ADP-ribose) polymerase
PCB	Polychlorinated biphenyl
PCR	Polymerase chain reaction

PFGE	Pulsed-field gel-electrophoresis
pI	Isoelectric point
polIII	DNA polymerase III
RC	Rolling circle
RE	Restriction endonuclease
rRNA	Ribosomal ribonucleic acid
sso	Single-stranded origin
TA	Toxin-Antitoxin
tRNA	Transfer ribonucleic acid
qPCR	Quantitative real-time PCR