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

List of Figures

List of Tables

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