Metabolic characteristics and genomic epidemiology of *Escherichia coli* serogroup O145

A thesis presented in partial fulfilment of the
requirements for the degree of
Master of Science
in Microbiology
at Massey University,
Palmerston North, New Zealand

Rose Collis

2018

Abstract

Shiga toxin-producing *Escherichia coli* (STEC) are a global public health concern, and can cause severe human disease. Ruminants are asymptomatic reservoirs of STEC, shedding this pathogen via their faeces. There is 'zero tolerance' for the Top 7 STEC serogroups (O26, O45, O103, O111, O121, O145 and O157) in ground beef products exported to the USA. STEC may contaminate carcasses during processing and therefore are a major regulatory concern for New Zealand's meat industry. A previous study investigating the prevalence of STEC in young calves (n=1508) throughout New Zealand identified STEC O145 as the most prevalent serogroup (43%) at the dairy farm level compared to the other Top 7 serogroups. This high prevalence underlines STEC O145 as a public health concern and an issue for the meat industry.

Current culture-based methods for STEC detection are not fully discriminatory due to the lack of consistent differential characteristics between STEC and non-pathogenic *E. coli*. This study aims to (i) investigate metabolic characteristics of *E. coli* O145 to facilitate the differential culture of this serogroup and (ii) understand the genomic epidemiology of *E. coli* O145 using whole genome sequencing (WGS).

E. coli O145 strains examined in this study were genetically and metabolically diverse, according to carbon utilisation. The metabolic and genomic analyses were unable to differentiate between *stx*-positive and *stx*-negative O145 strains and there was no association with isolation source. However, clustering of O145 strains was observed according to multi-locus sequence type and at the level of *eae* subtype, a gene encoding the protein intimin which is involved in bacterial attachment to intestinal epithelial cells. Carbon substrates such as D-serine and D-malic acid were identified as candidate metabolites to differentiate defined O145 sequence types and may assist with identification in conjunction with currently available molecular methods.

This research has demonstrated the genetic heterogeneity of serogroup O145 and has made significant progress in the identification of metabolites that may prove beneficial in the development of a differential media for certain subsets of serogroup O145. Such a medium would prove a valuable tool for maintaining and monitoring public health and providing food quality and safety assurances that New Zealand meat for export is free of this pathogen.

Acknowledgements

Firstly, I would like to thank my supervisors Dr Adrian Cookson, Dr Anne Midwinter, A/Prof Patrick Biggs and Springer Browne for their guidance and encouragement throughout my study. Their help and support has been invaluable: from answering my many questions, listening to my ideas, reading numerous thesis drafts and providing constructive feedback to improve my lab, writing and genomic analysis skills. I really appreciate it!

I would also like to thank all members of the AgResearch Food Assurance and Meat Quality team and the Massey University *m*EpiLab for their support and encouragement throughout my study. I am very grateful to have been able to work with a group of people who are very encouraging and always happy to share their knowledge- and a few laughs of course! A special thank you to Dr David Wilkinson for his guidance and help with the library preparations for WGS and genomic analysis; Dr Samuel Bloomfield for his help with the genomic evolutionary analysis; and Dr Sara Burgess for helping me with the formatting of this thesis.

Thank you to Dr Colleen Ross and Dr Delphine Rapp (AgResearch Ltd) and Hugo Strydom and Naveena Karki (The Institute of Environmental Science and Research) for generously providing serogroup O145 isolates for use in this study.

I would like to acknowledge the financial support of the Palmerston North Medical Research Foundation, the IVABS post-graduate research fund and the AgResearch Food Provenance and Assurance Strategic Science Investment (SSI) Fund programme for generously funding components of this research project; and to the AgResearch SSI Fund and Massey University for awarding me a Masterate Scholarship in my first and second year of study, respectively.

Finally, thank-you to my parents for encouraging me to follow my passion and continue studying microbiology, and inspiring me every day with your hard work and determination. A special thanks to my parents, sisters, Louis, Ellie, family and friends for your continued love and support throughout my study.

Declaration

The virulence factor tree (section 2.12.3) and perl scripts for genomic analyses (Appendices C and D) were provided by A/Prof Patrick Biggs. The remainder of the work in this thesis was conducted by the candidate with guidance from supervisors.

Abbreviations

°C	Degrees Celsius
µg	Microgram
µL	Microlitre
A/E lesions	Attaching and effacing lesions
BHI	Brain heart infusion
bp	Base pairs
CDS	Coding sequences
CFU	Colony forming units
CGE	Center for Genomic Epidemiology
COGs	Clusters of Orthologous Groups
C_t	Cycle threshold
CT-SMAC	Cefixime and tellurite sorbitol MacConkey agar
DAEC	Diffuse-adherent *E. coli*
DEC	Diarrheagenic *E. coli*
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
EAEC	Enteroaggregative *E. coli*
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ESS	Effective sample size
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extraintestinal *E. coli*
FAE	Follicle associated duodenum
GC	Guanine-cytosine
HGT	Horizontal gene transfer
HKY substitution model	Hasegawa-Kishino-Yano substitution model
HPD	Highest posterior density
IMBs	Immunomagnetic beads
IMS	Immunomagnetic separation
Indels	Insertions/deletions
iTOL	Interactive Tree of Life
kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
LAA pathogenicity island	Locus of adhesion and autoaggregation pathogenicity island
LEE pathogenicity island	Locus of enterocyte effacement pathogenicity island
LOD	Limit of detection

MCL	Markov cluster
MCMC	Markov Chain Monte Carlo
min	Minute
mL	Millilitres
MLST	Multi-locus sequence typing
mPCR	Multiplex polymerase chain reaction
mTSB	Modified tryptone soya broth
ng	Nanogram
nM	Nanomolar
PCR	Polymerase chain reaction
pm	Picomolar
PMA	Propidium monoazide
RAMS	Recto-anal mucosal swabs
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rpm	Revolutions per minute
RT-PCR	Real-time polymerase chain reaction
sec	Seconds
SNP	Single nucleotide polymorphism
ST	Sequence type
STEC	Shiga toxin-producing *Escherichia coli*
“Super six” STEC serogroups	O26, O45, O103, O111, O121, O145
T3SS	Type three secretion system
TBE buffer	Tris-borate-EDTA buffer
TMRCA	Time of most recent common ancestor
Top 7 STEC serogroups	O26, O45, O103, O111, O121, O145 and O157
tRNA	Transfer RNA
UPEC	Uropathogenic *E. coli*
USDA-FSIS	United States Department of Agriculture Food Safety Inspection Services
V	Volt
v/v	Volume per volume
w/v	Weight per volume
WGS	Whole genome sequencing

Table of contents

List of figures

List of tables