Phenotype-genotype relationships of *Escherichia coli* O157 and O26 isolates from New Zealand

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

Tessy Thankachen George

2017

mEpiLab
Institute of Veterinary, Animal and Biomedical Sciences
Massey University

Abstract

Shiga toxin-producing *Escherichia coli* (STEC) is a zoonotic pathogen responsible for causing severe manifestations of gastroenteritis in humans worldwide. STEC is transmitted to humans through the consumption of compromised food and water, or direct animal contact. *Escherichia coli* (*E. coli*) O157 and *E. coli* O26 are considered to be among the important STEC serogroups worldwide due to their association with outbreaks and clinical cases of haemolytic uraemic syndrome and haemorrhagic colitis. This study is concerned with the application of phenotypic microarray technology and high-throughput sequencing technology to characterise *E. coli* O157 and *E. coli* O26 New Zealand isolates.

In this study, 190 phenotypes of carbon sources utilisation were studied in isolates belonging to *E. coli* O157 and O26 serogroups. The isolates were divided into groups based on the fermentation of sorbitol and rhamnose in *E. coli* O157 and O26, respectively. All the isolates respired on approximately 40% of the carbon sources. The sorbitol positive *E. coli* O157 isolates respired on more carbon sources compared to the sorbitol negative *E. coli* O157, rhamnose negative and rhamnose positive *E. coli* O26 isolates.

Genomic analysis showed that *E. coli* O157 isolates had shorter genomes compared to those of the *E. coli* O26 isolates. The core genome comparisons revealed differences between and within the *E. coli* O157 and O26 serogroups. Clustering of *E. coli* O157 and O26 isolates based on sorbitol and rhamnose fermentation, respectively, was observed.

The results obtained from this study illustrated that phenotypic and genotypic differences existed within the *E. coli* O157 isolates. The findings of the current study also demonstrated that while the *E. coli* O26 isolates had similar phenotypic characteristics, genotypic differences existed within the isolates.

Acknowledgements

I sincerely thank my supervisors Dr Anne Midwinter, Associate Professor Patrick Biggs, Dr Julie Collins-Emerson and Professor Nigel French for their valuable suggestions and guidance throughout my study.

Thank you to Massey University and the Institute of Veterinary, Animal and Biomedical Sciences for providing me financial support and funds towards my research. Thank you to Kathy Hamilton, Professor Kevin Stafford and Debbie Hill for their administrative support.

I sincerely thank Dr Ben Phiri for proof-reading my thesis. I would like to thank Dr Jonathan Marshall for his helpful tips with the software R and Simon Verschaffelt for his technical support. Thank you to the staff and postgraduate students at mEpiLab, Massey University, I enjoyed working alongside all of you. I would like to thank Sameer Siddiqui, Rima Shrestha, Saroj Yadav, Punja Prasertsri, Antoine Nohra, José Solis-Ramirez, Kruno Bojanić, Doris Adeyinka and Shirli Notcovich for your kindness and friendship.

Most importantly, I would like to thank my family for their love, support and encouragement.

Let all that I am praise the Lord;
May I never forget the good things He does for me.
Psalm 103:2

Abbreviations

%	Percentage
°C	Degree Celsius
IIA^{glc}	Glucose-specific IIA protein
Acetyl-CoA	Acetyl coenzyme A
ADJ	Adjusted
AE	Attaching and effacing lesion
AFLP	Amplified fragment length polymorphism
argW	Stx-encoding bacteriophage insertion site
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BD	Becton Dickinson and Company
BGI	Beijing Genomics Institute
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BWA	Burrows-Wheeler Alignment
cAMP	Cyclic adenosine monophosphate
CDC	Centers for Disease Control and Prevention
CO_2	Carbon dioxide
COGs	Clusters of Orthologous Groups of proteins
CRP	cAMP receptor protein
CTA	Cystine trypticase agar
CT-RMAC	Rhamnose MacConkey medium containing cefixime and tellurite
CT-SMAC	Sorbitol MacConkey medium containing cefixime and tellurite
DAEC	Diffusely adhering *E. coli*
DNA	Deoxyribonucleic acid
dsdA	D-serine deaminase or dehydratase
dsdC	LysR-type transcriptional regulator
dsdCXA	D-serine tolerance locus
dsdX	D-serine transporter
EII	Enzymes II
eae	*E. coli* attaching and effacing

EAEC	Enteroaggregative *E. coli*
EAHEC	Enteroaggregative haemorrhagic *E. coli*
EDP	Entner-Doudoroff pathway
EHEC	Enterohaemorrhagic *E. coli*
ehxA	Haemolysin
EIA	Enzyme immunoassay
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ERL	Enteric Reference Laboratory
espA	*E. coli* secreted protein
espB	*E. coli* secreted protein
espD	*E. coli* secreted protein
espP	Protease
ESR	Institute of Environment Science and Research Limited
ETEC	Enterotoxigenic *E. coli*
E. coli	*Escherichia coli*
E. fergusonii	*Escherichia fergusonii*
*fli*C	Flagellin
fliC_{H11}	Flagellin H11
fyuA	High-pathogenicity island gene
g	Grams
Gb	Gigabases
GUD	beta-glucuronidase
H	Flagellar antigen
h	Hours
HC	Haemorrhagic colitis
HGT	Horizontal gene transfer
HPI	High-pathogenicity island
HPr	Heat-stable protein
HUS	Haemolytic uraemic syndrome
irp2	High-pathogenicity island gene
K	Capsular antigen
KEGG	Kyoto Encyclopedia of Genes and Genomes

LEE	Locus of enterocyte effacement
LGT	Lateral gene transfer
M	Molar
Mb	Megabases
MDa	Megadalton
MEE	Multilocus enzyme electrophoresis
µl	Microlitres
ml	Millilitres
Min	Minutes
MLST	Multilocus sequence typing
MPI	Ministry of Primary Industries
NADH	Nicotinamide adenine dinucleotide
NCBI	National Centre for Biotechnology Information
ng	Nanograms
NI	North Island
nle	Non-LEE effector
NMD	National Microbiological Database
NZ	New Zealand
NZGL	New Zealand Genomics Limited
O	Somatic antigen
PAI	Pathogenicity island
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PFGE	Pulsed field gel electrophoresis
pH	Potential of hydrogen
PM	Phenotypic microarray
PPP	Pentose phosphate pathway
PTS	Phosphotransferase system
QC	Quality control
RFLP	Restriction fragment length polymorphism
RMAC	Rhamnose MacConkey medium
RN	Rhamnose non-fermenting or negative

RNA	Ribonucleic acid
RP	Rhamnose fermenting or positive
SBI	Shiga toxin-encoding bacteriophage insertion
sbcB	Stx-encoding bacteriophage insertion site
SBS	Synthesis by synthesis
SI	South Island
SMAC	Sorbitol MacConkey medium
SN	Sorbitol non-fermenting or negative
SP	Sorbitol fermenting or positive
STEC	Shiga toxin-producing *E. coli*
Stx, *stx*	Shiga toxin
stx1	Shiga toxin 1
stx2	Shiga toxin 2
stx2c	Subtype Shiga toxin 2
T3SS	Type III secretion system
TTC	2,3,5-triphenyl tetrazolium chloride
USA	United States of America
UTI	Urinary tract infection
v	Version
VTEC	Verotoxin producing *E. coli*
wrbA	Stx-encoding bacteriophage insertion site
yehV	Stx-encoding bacteriophage insertion site
yhaJ	LysR-type transcriptional regulator
yhaO	Inner membrane transporter
yhaOMKJ	D-serine sensory locus

Contents

List of Tables

List of Figures

ONE

General introduction

1.1 Background

The first outbreak of human disease associated with Shiga toxin-producing *Escherichia coli* (STEC) was reported in 1982 (Riley *et al.*, 1983). STEC infections can cause life-threatening illnesses such as haemorrhagic colitis (HC) and haemolytic uraemic syndrome (HUS) (Griffin *et al.*, 1988). The most important reservoir for STEC is considered to be ruminants, in particular, cattle (Bettelheim, 2007, Armstrong *et al.*, 1996). Humans can acquire STEC infection when they have direct contact with animals or through the consumption of faecally contaminated food and water (Bettelheim, 2007, Armstrong *et al.*, 1996). STEC serogroups frequently associated with causing disease in humans are O157, O26, O111, O145 and O103 (Beutin, 2006). Isolates belonging to the *Escherichia coli* (*E. coli*) O157 serogroup are usually sorbitol non-fermenters while those of *E. coli* O26 serogroup are rhamnose non-fermenters (March & Ratnam, 1986, Hiramatsu *et al.*, 2002).

1.2 Shiga toxin-producing *Escherichia coli* in New Zealand

In New Zealand (NZ), sporadic cases of Shiga toxin-producing *Escherichia coli* infection have been reported. No widespread outbreaks of STEC have been identified (ESR, 2016, Gadiel & Abelson, 2010). The sources of the sporadic cases are reported to be contact with farm animals, the consumption of dairy products and poultry products (ESR, 2016). However, a study of human cases of STEC in NZ found that the sources of infection were environmental and animal contact, but not food (Jaros *et al.*, 2013). In 2009, the cost for treating 210 incident cases of STEC was estimated to be NZ$ 380,000 (Gadiel & Abelson, 2010).

The dairy and meat industries make a significant contribution to the New Zealand economy (MacPherson, 2014, Cavanagh, 2003). Twenty percent of the primary sector export earnings in New Zealand comes from the export of meat and wool (MPI, 2015). The export value of meat and wool was recorded at NZ$ 8757 million in 2015, with 35.7% attributed to beef and veal exports and 38% from the export of lamb and mutton. While many countries receive varied

types of meat from New Zealand, the United States of America (USA) is the highest consumer of beef and veal (MPI, 2015). The National Microbiological Database (NMD) programme which is controlled by the Ministry of Primary Industries (MPI), New Zealand ensures that STEC-testing of beef is carried out by every export associated slaughter plant as required for overseas market access (MPI, no date). Compromised meats detected by New Zealand's trading partners and increased incidence of foodborne infections caused by *E. coli* would have serious implications for the New Zealand economy.

1.3 Thesis objectives and structure

In New Zealand, cattle and sheep are considered to be reservoirs of STEC (Cookson *et al.*, 2006a, Cookson *et al.*, 2006b). Studies have reported on the epidemiology of STEC in cattle in New Zealand (Irshad *et al.*, 2011, Jaros *et al.*, 2016). Bovine and human *E. coli* O157 and O26 isolates were obtained and identified from the North Island (NI) and South Island (SI) of New Zealand (Jaros, 2014). The isolates in the current study were chosen randomly to represent both the North and South Islands. The bovine isolates were obtained from faecal swabs collected at four large slaughter plants in New Zealand during the years 2009 to 2011 and the human isolates were obtained from confirmed STEC cases in New Zealand by the Enteric Reference Laboratory (ERL) (The Institute of Environment Science and Research Limited (ESR), New Zealand) during the years 2011 to 2012 (Jaros, 2014).

Four *E. coli* O157 and *six* O26 isolates were used in this study. The phenotypic profiles of these isolates were obtained using phenotypic microarray (PM) technology and the draft genomes of these isolates were sequenced by New Zealand Genomics Limited (NZGL), Palmerston North on the Illumina MiSeq analyser. These 10 isolates were compared to another 13 *E. coli* O157 isolates from a previous study carried out in 2012 (George, Massey University). The objectives of this study were to investigate:

- the phenotypic characteristics of sorbitol-fermenting atypical *E. coli* O157 and rhamnose-fermenting atypical *E. coli* O26 New Zealand isolates,
- the phylogeny of these sorbitol-fermenting *E. coli* O157 and rhamnose-fermenting O26 isolates.

For this study, it was hypothesised that the differences in the phenotypic characteristics of sorbitol-fermenting *E. coli* O157 and rhamnose-fermenting *E. coli* O26 isolates compared to those of sorbitol non-fermenting *E. coli* O157 and rhamnose non-fermenting *E. coli* O26 isolates, respectively, would be correlated with differences in the genetic characteristics of these isolates. This study will help our understanding of the behaviour and genetic differences that contribute to the survival and virulence of *E. coli* in the bovine and human host. The phenotypic profiles will offer an insight into their metabolic capabilities and the phylogenetic analysis will assist in understanding the evolutionary relationships of the New Zealand *E. coli* isolates. This study will also give an insight into whether the detection methods currently used by the meat industry and diagnostic laboratories would be able to readily and consistently detect *E. coli* O157 and O26 strains in meat and in *E. coli* related infections in humans respectively.

TWO

Literature review

2.1 Introduction

Escherichia coli (*E. coli)* was discovered by Theodor Escherich in 1885 (Sussman, 1985). It has become one of the best understood and characterised microbial organisms (Kuhnert *et al.*, 2000). Despite the vast knowledge accumulated, Shiga toxin-producing *E. coli* (STEC) is the most serious causative agent of bacterial gastroenteritis in humans worldwide. STEC, also known as verotoxin producing *E. coli* (VTEC), produce verotoxins which are genetically and immunologically similar to Shiga toxins (Jackson *et al.*, 1987, Bhunia, 2008). One of the most virulent STEC is *E. coli* O157:H7 while *E. coli* O26:H11 has emerged as the most important non-O157 STEC.

E. coli is commonly found in the intestinal tract of animals and humans as well as in the environment (Bhunia, 2008). However, the principal animal reservoir is cattle which are known to be asymptomatic carriers (Caprioli *et al.*, 2005). Consumption of undercooked or contaminated food or water, direct contact with farm animals and environmental contamination of pasture and waterways with animal faeces are among the most significant routes for acquiring an *E. coli* infection (Cookson *et al.*, 2006b, Riley *et al.*, 1983, Spickler *et al.*, 2010, Jaros *et al.*, 2013).

This review looks at published literature regarding STEC to provide a foundation to the objectives of the current research and for the other chapters. In this thesis, *E. coli* O157 refers to either/both *E. coli* O157:H7 and *E. coli* O157:H-, and *E. coli* O26 refers to either/both *E. coli* O26:H11 and *E. coli* O26:H-, unless specified otherwise. In the studies or reports cited for non-O157 *E. coli*/STEC strains/isolates, with exception to the isolates used in the current study, the O26 serogroup may or may not be included unless reported in the respective studies. Also, *E. coli* O157 and O157 STEC and *E. coli* O26 and O26 STEC are interchangeable respectively.

2.2 *Escherichia coli*

Members of the family Enterobacteriaceae, *E. coli* bacteria taxonomically belong to the gamma subdivision of the Proteobacteria phylum (Lukjancenko *et al.*, 2010). They are rod-shaped in appearance (Figure 2.1) (Anonymous, no date-b), facultatively anaerobic and are classified as Gram-negative (Oeggerli, 2009). The bacterial organism is often found in the intestinal flora and faeces of both humans and animals. Furthermore, *E. coli* is well adapted to survive in its ecological niche, displaying a wide variety of metabolic versatility which has made it a very successful organism (Lukjancenko *et al.*, 2010).

Within hours of birth, non-pathogenic variants exist commensally in mammals after colonising an infant gut. In healthy humans, *E. coli* can provide some limited form of protection to their hosts as their sheer abundance is an obstacle for invading pathogens trying to establish colonisation for infection (Kaper *et al.*, 2004). However, other variants of *E. coli* can be opportunistic pathogens causing a variety of human diseases that are gastrointestinal in nature or extraintestinal such as urinary tract infections (UTI), bacteraemia and meningitis (Brooks *et al.*, 1997).

There are various well described categories of intestinal *E. coli* which cause enteric diseases through their unique pathogenic mechanisms: STEC, enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and more recently enteroaggregative haemorrhagic *E. coli* (EAHEC) (Russo & Johnson, 2000, Zaw *et al.*, 2013, Brzuszkiewicz *et al.*, 2011). These categories of *E. coli* are defined by their virulence properties, the diseases they cause and from serologically differentiating them based on their O (somatic, those located in the cell wall of *E. coli*), H (flagellar) and K (capsular) antigens (Bhunia, 2008).

2.2.1 Shiga toxin-producing *Escherichia coli*

Shiga toxin-producing *Escherichia coli* strains are considered pathogenic and are characterised by the presence of a 60 megadalton (MDa) plasmid (pO157) encoding an enterohaemolysin, one or more phage encoded Shiga toxins, and the ability to cause attaching and effacing (AE) lesions on epithelial cells (Nataro & Kaper, 1998, Mead & Griffin, 1998, Rahal *et al.*, 2012).

STEC produce Shiga-like toxins which is a trait of this group of *E. coli*. The Shiga toxin, also known as verotoxin, destroys Vero cells (a lineage of kidney epithelial cells extracted from an African green monkey). Thus, STEC group is also referred to as VTEC (verotoxin-producing *E. coli*) (Bhunia, 2008).

E. coli O157:H7 and the non-motile *E. coli* O157:H- are the most studied and recognised serogroup of STEC. STEC was first observed as a human pathogen in 1982 when two outbreaks of haemorrhagic colitis (HC) occurred from consuming undercooked meat at a fast-food restaurant chain in USA. The pathogen associated with the outbreak was reported to be *E. coli* O157:H7 (Riley *et al.*, 1983).

Non-O157 serogroups such as O8, O91, O113, O103, O26, O111, O121 and O145, among others, have also been found to be associated with STEC infection (Werber *et al.*, 2008, Bettelheim, 1998). Infections caused by non-O157 STEC are often overlooked in diagnostic testing as methods are more tailored for the detection of *E. coli* O157 (Alfredo *et al.*, 1997). The first outbreak of *E. coli* O26:H11 in rural northwest Ireland occurred in a crèche. The cause of the infection was unknown and some of the symptoms presented by the children included non-bloody diarrhoea and abdominal cramps (McMaster *et al.*, 2001).

In the current research, isolates belonging to STEC serogroups O157 and O26 were examined.

2.2.2 Clinical spectrum of STEC

The clinical symptoms of *E. coli* infections are characterised by abdominal cramps and watery diarrhoea which progresses rapidly to bloody diarrhoea or haemolytic uraemic syndrome (HUS). A low infectious dose, fewer than 50 organisms, is sufficient to cause infection (Baker *et al.*, 1999). The infectious dose estimated for four patients from an outbreak caused by consuming one brand of salami, was found to be in the range of two to 45 bacteria (Tilden Jr *et al.*, 1996). Ingestion of the bacterium is followed typically by a three to four day incubation period during which colonisation of the large bowel occurs (Besser *et al.*, 1999). Illness begins with non-bloody diarrhoea and severe abdominal cramps for one to two days which progresses in the second or third day of illness to bloody diarrhoea that lasts for four to ten days (Tarr & Neill, 1996).

Children, the elderly and immunocompromised individuals are more susceptible to an STEC infection because of their underdeveloped or weakened immune system (Smith, 1998). HUS largely affects children, for whom it is the leading cause of acute renal failure. Some people may be infected but asymptomatic, that is *E. coli* O157 neither cause nor exhibit symptoms of a disease. The carriage is limited to three weeks in these asymptomatic cases (Griffin *et al.*, 1988).

Clinical manifestations of non-O157 STEC are similar to that of O157 STEC and the infection range from mild, watery diarrhoea to HC, HUS and death (Jelacic *et al.*, 2003, Werber *et al.*, 2004, Hughes *et al.*, 2006, Brooks *et al.*, 2005). Non-O157 serogroups appear to cause watery diarrhoea more often than bloody diarrhoea (Werber *et al.*, 2004).

2.2.3 Virulence characteristics of STEC

STEC possesses various virulence factors that assist in its pathogenicity. One of the features of STEC is that they contain at least one Shiga toxin gene (*stx*) and may cause life-threatening complications such as HUS (Karmali *et al.*, 1983). In some clinical cases, it is difficult to distinguish between STEC and non-STEC. However, not all non-O157 STEC can cause human infections or complications such as HUS.

The Shiga toxin 1 (*stx1*) and Shiga toxin 2 (*stx2*) genes are the most important virulence factors in STEC. *E. coli* O157 isolates possess both *stx1* and *stx2*, although the majority produce *stx2* (Boerlin *et al.*, 1999), and *E. coli* O26 isolates possess *stx1* (Scotland *et al.*, 1990). These toxins are encoded on lysogenic bacteriophages (Shiga toxin (Stx)-encoding bacteriophage) (Perna *et al.*, 2001). They have similar chemical structures and modes of action. They consist of A and B subunits. The binding of the toxin to the target cells is carried out by the B subunit. Once internalised, the A subunit is activated, and purine residues are removed from the host cell ribosomes causing the inhibition of protein synthesis in eukaryotic cells (Meng & Schroeder, 2007). However, *stx2* is more cytotoxic than *stx1* and targets the human renal microvascular endothelial cells which results in HUS (Louise & Obrig, 1995).

EHEC or STEC can adhere to epithelial cells and produce characteristic attaching and effacing lesions (Nataro & Kaper, 1998). An adhesion molecule called intimin allows the bacterium to attach tightly to epithelial cells. The gene *eae* (*E. coli* attaching and effacing) encodes intimin

and is clustered in a pathogenicity island (PAI), also known as locus of enterocyte effacement (LEE) (McDaniel *et al.*, 1995). Barrett *et al*. (1992) tested for the presence of the *eae* gene in 30 *E. coli* O157 isolates, 19 non-O157 STEC human isolates and 26 non-O157 STEC animal isolates. The authors reported that all the O157 isolates, 16 of 19 non-O157 human isolates and 12 of 26 non-O157 animal isolates possessed the *eae* gene (Barrett *et al.*, 1992). One PAI known as the high-pathogenicity island (HPI) was found among enterobacteria (Schubert *et al.*, 1998). *E. coli* O157:H7 do not possess the HPI element (Schubert *et al.*, 1998) while O26 STEC strains were shown to possess the HPI-specific genes, *irp2* and *fyuA* (Karch *et al.*, 1999).

The LEE consists of three functional domains: *eae* and *tir* genes in the central region, type III secretion system (T3SS) in the second, and genes for other secreted proteins (*espA*, *espB*, *espD*) in the third (Kaper & O'Brien, 1998). The T3SS enables STEC to attach to intestinal epithelia on infecting the host. It contributes to the colonisation, transmission and persistence of both O157:H7 and non-O157 STEC in cattle (van Diemen *et al.*, 2005, Wickham *et al.*, 2007, Coombes *et al.*, 2011). Genomic islands and other mobile genetic elements such as prophages, plasmids and transposons contribute to the virulence potential of STEC (Lawrence, 2005). Genomic islands encoding non-LEE effector (*nle*) genes were reported to be present in non-O157 isolates and absent in O157 isolates, adding to the pathogenic potential of non-O157 STEC (Coombes *et al.*, 2008).

Many *E. coli* O157 isolates express H7 flagella type which assist the bacterial organism to bind to surfaces in the hosts and the H antigen is encoded by the *fli*C gene (Stevens *et al.*, 2002, Fields *et al.*, 1997, Erdem *et al.*, 2007). Durso *et al*. (2005) reported that the somatic O- antigen and flagellar H11 (*fliC*$_{H11}$ gene) antigens were detected by multiplex polymerase chain reaction (PCR) in 45 out of 46 STEC and non-STEC O26:NM (non-motile) isolates (Durso *et al.*, 2005).

Additional virulence traits and the O157 antigen are believed to have been acquired through horizontal gene transfer and recombination. Other studies have reported a haemolysin (*ehxA*) (Schmidt *et al.*, 1995) and a protease (*espP*) encoded on the same plasmid as additional virulence factors in O157 STEC (Brunder *et al.*, 1997). *EhxA* and *espP* have also been reported to be present in *E. coli* O26 (Madic *et al.*, 2011, Brunder *et al.*, 1997).

2.2.4 Shiga toxin-encoding bacteriophage insertion typing

Shiga toxin genes that are encoded in bacteriophages associated with specific *stx* variants, are inserted at one or more chromosomal locations. Shiga toxin-encoding bacteriophage insertion (SBI) typing is determined by PCR mediated detection of *stx* genes, Stx-encoding bacteriophage chromosome junctions, and intact bacteriophage insertion site regions in the isolates (Shaikh & Tarr, 2003).

Initially, SBI genotyping was designated as cluster 1, 2 and 3 based on the presence of the intact or occupied insertion sites, variable *yehV* right junction and *wrbA (*Besser *et al*., 2007*).* Further studies have detected the presence of other phage insertion sites *argW* and *sbcB* (Jung *et al.*, 2013, Mellor *et al.*, 2013, Kotewicz *et al.*, 2008). The current method of SBI genotype nomenclature includes 'the concatenation of the upper case letters of the bacteriophage insertion loci (A = *argW*, S = *sbcB*, W = *wrbA* and Y = *yehV*) followed by numerical addition of the Stx variants detected' (Shringi *et al.*, 2012, pp. 4).

In a study by Besser *et al*. (2007) 282 human and 80 bovine *E. coli* O157 isolates were genotyped by SBI typing. The three principal genotypes were clusters 1 to 3, each one corresponding to the location of the insertion of Stx2-encoding bacteriophage at *wrbA* and *yehV* or at either one and not the other. The findings suggested that the bovine isolates showed more diversity compared to that of human isolates. Out of the bovine isolates, 41 out of 80, and among the human isolates, 268 out of 282, belonged to clusters 1, 2 or 3 (Besser *et al.*, 2007).

However, little is known and fewer studies have investigated the SBI typing of *E. coli* O26. A study by Irshad (2013) first reported on the SBI typing of *E. coli* O26 isolates. The findings suggested that the isolates belonged to two genotypes under the previous method of SBI genotype nomenclature, 9 and 20. Isolates belonging to genotype 9 possessed the *eae* and *ehxA* genes and isolates belonging to genotype 20 possessed *stx1*, *eae* and *ehxA* genes (Irshad, 2013).

2.2.5 Detection, isolation and identification of STEC

E. coli O157 was initially described as being sorbitol non-fermenting, oxidase negative, indole positive and glucose fermenting (March & Ratnam, 1986). However, some *E. coli* O157 were found to ferment sorbitol (Gunzer *et al.*, 1992, Aleksić *et al.*, 1992). MacConkey agar is a

selective and differential culture medium which is used to isolate Gram-negative bacteria such as *E. coli* O157:H7, where lactose fermenting, pink colonies are produced (Zimbro *et al.*, 2009). The agar base contains peptones, lactose, bile salts no. 3, crystal violet, neutral red, sodium chloride and agar. The growth of Gram-positive bacteria is inhibited by bile salts no. 3 and crystal violet (Zimbro *et al.*, 2009).

Selective and differential media such as sorbitol MacConkey (SMAC) or SMAC containing cefixime and tellurite (CT-SMAC) are also used to identify and isolate *E. coli* O157 (March & Ratnam, 1986, Gould *et al.*, 2009). After overnight incubation at 37 degree Celsius (°C), sorbitol non-fermenting *E. coli* O157 produces colourless colonies on SMAC and CT-SMAC media. *E. coli* O157 isolates that ferment sorbitol produce pink colonies on SMAC but might not grow on CT-SMAC due to its susceptibility to tellurite. On SMAC, the pink colonies may also be surrounded by a zone of precipitated bile (Gould *et al.*, 2009, Becton Dickinson and Company, no date).

The identity of the colonies is confirmed by agglutination with O157 specific antiserum or O157 latex reagents. Latex reagents are composed of blue latex particles that are sensitised with reactive (positive) and non-reactive (negative) antibodies (Anonymous, no date-a). H7 latex agglutination tests are also performed on confirmed *E. coli* O157 colonies to investigate motility. Enzyme immunoassay (EIA) or PCR methods are used to detect virulence genes such as *stx1* and *stx2* (Gould *et al*., 2009).

In order to isolate non-O157 *E. coli*, initially a Shiga toxin-positive broth inoculated with the sample is streaked onto less selective media such as MacConkey, SMAC or horse blood media. The *E. coli* O157 colonies are isolated and identified, following which the other *E. coli*-like morphology colonies are tested for sorbitol and lactose fermentation. The presence of virulence genes is then tested for using EIA or PCR. O-specific antisera are used to identify the STEC associated serogroup (Gould *et al.*, 2009).

As the sorbitol and lactose fermentation tests are not specific to isolating non-O157, Hiramatsu *et al*. (2002) developed rhamnose MacConkey (RMAC) and RMAC containing cefixime and tellurite (CT-RMAC) media (Hiramatsu *et al.*, 2002, Bielaszewska *et al.*, 2005). For RMAC, the lactose in the MacConkey medium was replaced by rhamnose, and for CT-RMAC, potassium tellurite and cefixime were added to RMAC. The *E. coli* O26 strains did not ferment

rhamnose and produced colourless colonies, proving both media to be successful in isolating *E. coli* O26. The *E. coli* O157, *E. coli* O111 and other non-O157 strains fermented rhamnose on RMAC producing red colonies and only a small number of these strains grew on CT-RMAC. These results suggested that CT-RMAC may be used as a primary medium for isolating *E. coli* O26 (Hiramatsu *et al.*, 2002). However, as seen with sorbitol fermenting *E. coli* O157 strains, some *E. coli* O26 strains were found to ferment rhamnose (Trevisani *et al.*, 2013).

2.2.6 Prevalence of STEC

Cattle, asymptomatic excretors of *E. coli* O157:H7, are considered to be the most important reservoir for *E. coli* O157:H7 (Caprioli *et al.*, 2005). Sheep and other animals such as horses are potential sources for *E. coli* O157 infections (Strachan *et al.*, 2001, Baljer, 1997). Healthy bobby calves were found to be asymptomatic reservoirs for *E. coli* O157 in New Zealand (NZ) (Irshad *et al.*, 2011).

Faecal samples from all cattle present at 10 Dutch dairy farms were collected and examined for the presence of *E. coli* O157 in late 1996. Seven out of 10 farms were tested positive for *E. coli* O157 and the excretion rate was highest in calves aged four to 12 months. Two of the *E. coli* O157 positive farms and two of the *E. coli* O157 negative farms were followed up five times in three months. On each of the farms, the cattle tested positive for *E. coli* O157 at least once (Heuvelink *et al.*, 1998).

Similarly, sources of non-O157 STEC responsible for causing *E. coli* infections have also been reported. In a study by Barlow & Mellor (2010), a total of 25 AUS-MEAT-accredited abattoirs across all of Australia participated. Three hundred beef cattle faecal samples were investigated for the prevalence of EHEC serotypes (a subgroup of STEC). The samples were characterised for the presence of virulence markers. Seventy eight samples contained the virulence markers *stx* and *eae*. Twenty seven different serotypes were isolated out of which one O26 and five O157 tested positive for any of the virulence markers. This showed that the Australian beef cattle had a low prevalence of EHEC (Barlow & Mellor, 2010).

In New Zealand, faecal swabs of 319 healthy animals which included sheep and cattle were tested for the presence of *stx1*, *stx2* and *eae* by Cookson *et al.* (2006a) using multiplex PCR.

The authors reported that 27.3% of cattle and 65.9% of sheep were positive for the *stx1* and/or *stx2* genes. They also found that 36.9% of cattle and 27.3% of sheep were positive for *eae*. The study demonstrated that STEC was prevalent in sheep and cattle (Cookson *et al.*, 2006a). In a follow-up study, some of the serogroups isolated from the same set of samples were O26:H11, O5:H-, O84:H-/H2, O91:H- and O128:H2, which have been found to cause disease in humans (Cookson *et al.*, 2006b).

2.2.7 Epidemiology of STEC outbreaks

STEC can be transmitted to humans through various routes. These are: (i) faecally-contaminated food, (ii) direct or indirect contact with STEC-shedding animals, or (iii) faecally-contaminated drinking or recreational water, as shown in Figure 2.2 (Reference laboratory for *Escherichia coli* (EcL), 2004). Inadequate attention to personal hygiene causes transmission of the pathogen from person to person, and is responsible for sporadic outbreaks in households (Parry & Salmon, 1998), day care centres (McMaster *et al.*, 2001) and nursing homes (Ryan *et al.*, 1986).

Several foodborne outbreaks of *E. coli* O157 have occurred from the consumption of contaminated beef, raw milk, municipal water, apple cider, raw vegetables, and sprouts (Rangel *et al.*, 2005). *E. coli* infections are transmitted by animals, both domestic and wild, such as sheep, goats, deer, dogs, horses, swine and cats (Beutin *et al.*, 1993). Water buffaloes have been shown to be reservoirs (Galiero *et al.*, 2005), and wild birds and rodents as carriers of *E. coli* O157 (Nielsen *et al.*, 2004).

Many outbreaks of *E. coli* O157 have been documented worldwide. However, non-O157 STEC have also proven to be a concern due to the number of outbreaks in the United States of America (USA), Australia and other countries where non-O157 STEC have been isolated from humans suffering from HUS and diarrhoea (Johnson *et al.*, 1996). A surveillance system was set up in France to monitor the incidence of HUS in children less than 15 years of age. For the years 1996 to 2006, 0.71 cases of HUS per 100,000 children less than 15 years of age and 1.87 cases of HUS per 100,000 children less than five years of age were recorded. Ninety six percent of these cases were sporadic, of which two were caused by *E. coli* O157 and a dual infection of *E. coli* O26 and O80 were detected (Espié *et al.*, 2008).

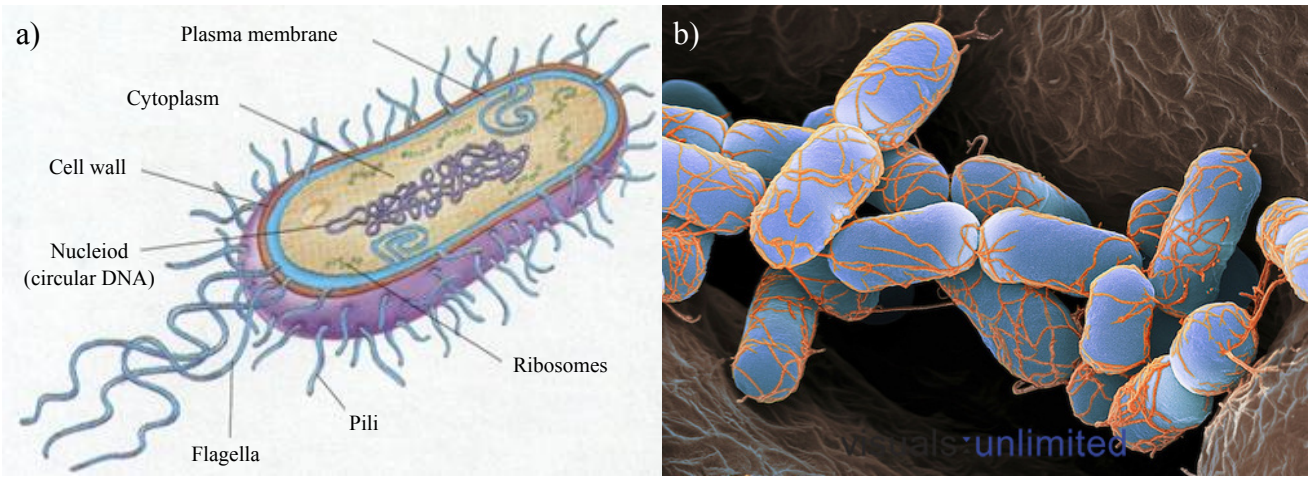

Figure 2.1. *Escherichia coli* (*E. coli*). a) Cell structure of the *E. coli* bacterium.
Source: (Anonymous, no date-b)
b) *E. coli* dividing by binary fission, captured by a scanning electron microscope (SEM), REM format.
Source: (Oeggerli, 2009)

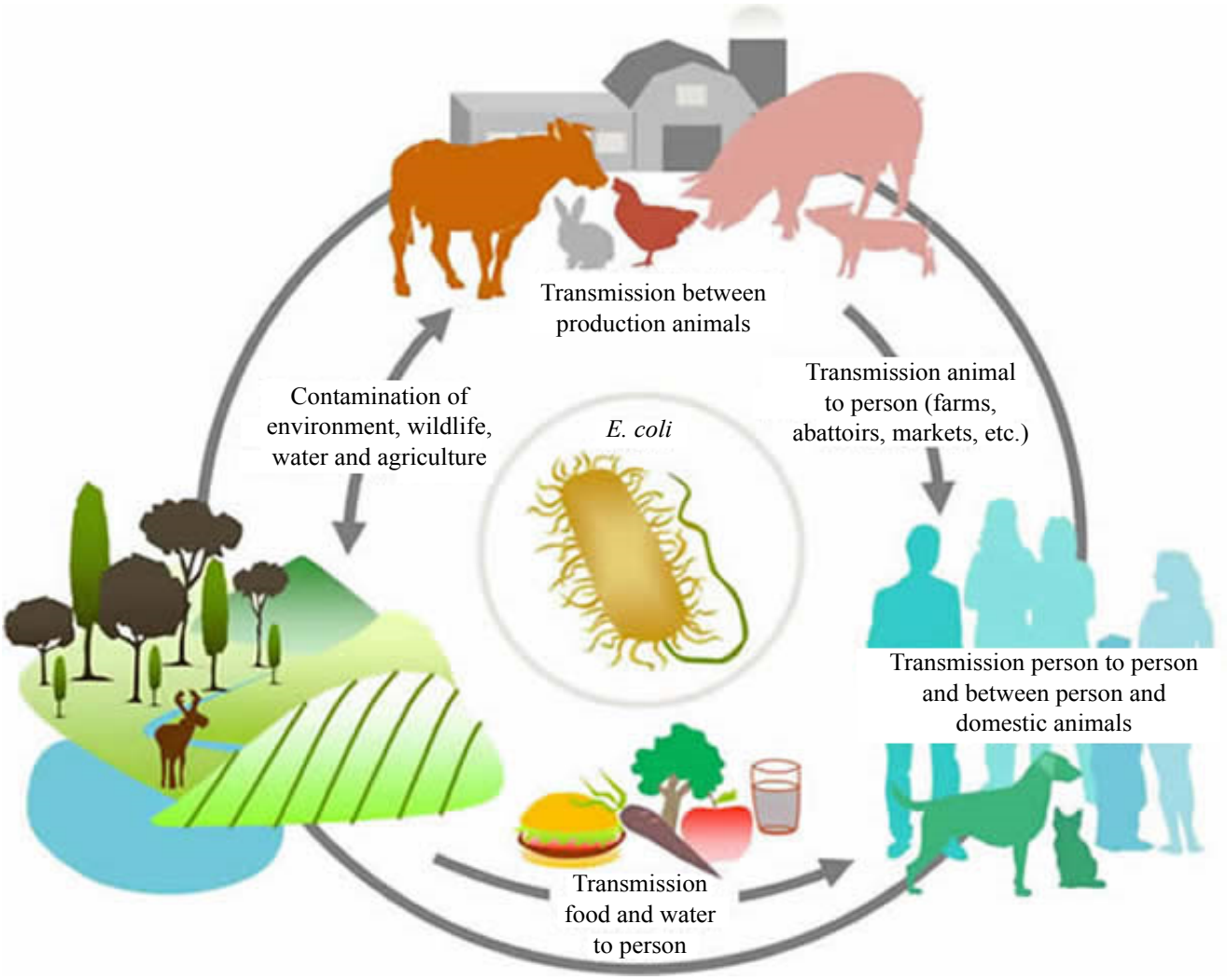

Figure 2.2. Routes of transmission. Possible routes of transmission from contamination of environment, animal products, food and produce to causing *E. coli* infection in humans.
Source: Permission obtained (Reference laboratory for *Escherichia coli* (EcL), 2004)

Similarly, a surveillance system was set up in Australia during the years 1994 to 1998 to monitor the incidence of HUS in children under the age of 15 years. A total of 98 cases were found. The incidence rates were 0.64 cases per 100,000 persons per year for five to 15 year olds and 1.35 cases per 100,000 for those younger than five years. These included 64 sporadic cases and 20 outbreak cases among other incidents associated with diarrhoea. The non-O157 serogroups associated with the sporadic cases included O26:H-, O113:H21 and O157:H- among others. However, no *E. coli* O157:H7 strains were isolated from any of the cases. The most common serogroups associated with both the sporadic cases and outbreaks belonged to O111:H- (Elliott *et al.*, 2001).

In the United States of America, during the years 1983 to 2002, 940 clinical cases of non-O157 STEC were tested by the Centers for Diseases Control and Prevention (CDC). Twenty two percent of these isolates belonged to *E. coli* O26 and the remaining belonged to O111, O103, O121, O45 and O145 serogroups (Brooks *et al.*, 2005). During 2000 to 2006, in the United States of America, a surveillance of infections recorded clinical cases of STEC infections in small cities and rural areas. Out of 206 identified STEC isolates, 108 (52%) were non-O157 serotypes and 98 (48%) were O157 serotypes. The O157 cases involved more cases of bloody diarrhoea, hospitalisation and HUS compared to the non-O157 cases (Hedican *et al.*, 2009). Further, the foodborne diseases active surveillance network (FoodNet) reported 2006 cases of non-O157 STEC infections and 5688 cases of O157 STEC infections during 2000 to 2010. Twenty six percent of the non-O157 STEC reported belonged to the O26 serogroup (Gould *et al.*, 2013).

2.2.8 Epidemiology of STEC outbreaks in New Zealand

In New Zealand, the first case of *E. coli* O157:H7 was recorded in October 1993 in an 11-month-old boy who developed HUS (Wright *et al.*, 1993). Although no large outbreaks cases of *E. coli* O157:H7 and non-O157 STEC involving consumption of contaminated meat have been reported in New Zealand, 17 small outbreaks of STEC infection involving 94 cases were reported in 2015 (ESR, 2016).

STEC emerged relatively recently, but the incidence has risen rapidly (Baker *et al.*, 1999, Crump *et al.*, 2001). The incidence of STEC infection notified in New Zealand was 1.3 cases per 100,000 population in 1999 (Baker *et al.*, 1999). The rate of notification in 2015 was 7.2

cases per 100,000 population (330 cases), which was significantly higher than the rates in 2014 that stood at 4.1 cases per 100,000 population (187 cases) (ESR, 2016). The change in notification rate is partly due to the introduction of a PCR method to detect STEC (ESR, 2016).

The highest notification rates were observed in the one to four years age group (40.5 per 100,000 population; 100 cases), followed by children aged less than one year (32.1 per 100,000; 19 cases). Eighty cases were hospitalised, among which 14 developed HUS and no deaths were reported. The most common risk factors reported were contact with pets, farm animals, manure and the consumption of dairy products, raw produce and poultry products (ESR, 2016, Jaros *et al.*, 2013).

2.3 Impact of compromised meat on the meat industry

The broad spectrum of foodborne infections has changed dramatically with new infections emerging. While the pathogenic spectrum has changed substantially, others have emerged through mutations or moved into a new niche in the food chain (Tauxe, 2002). Outbreaks of *E. coli*, *Listeria monocytogenes* and *Salmonella* have caused serious risk to the well-being and health of consumers (Mead *et al.*, 1999, Piggott & Marsh, 2004). Obtaining high quality foods consistently is a challenge that food industries face.

Contamination of carcasses during slaughter and processing in an abattoir (Figure 2.3) contaminated water sources and unhygienic plant personnel can significantly compromise the quality and safety of meat products and transmit the pathogen to the consumer (Chapman *et al.*, 1993, Rochan, 2013, Bower *et al.*, 1996, Gill, 1998). An *E. coli* O157 outbreak in Washington State, USA, in the winter of 1992 resulted in 501 reported cases including three deaths following the consumption of contaminated hamburgers. This led to 250,000 potentially contaminated hamburgers being recalled (Bell *et al.*, 1994).

In New Zealand, agriculture is the mainstay of the economy and export of agricultural products is approximately 40% of New Zealand's export earnings. Dairy and beef are leading export earners, generating approximately 33% of total export earnings (MacPherson, 2014). New Zealand's top five export markets include China, Australia, the United States of America, Japan and Korea (Statistics New Zealand, 2013). As in most developed countries, regulatory activities in New Zealand have been implemented, aimed at obtaining high quality meat and assuring the

hygienic adequacy of the meat supply. The Ministry of Primary Industries (MPI), New Zealand regulated National Microbiological Database (NMD) programme is a mandatory food safety and microbiological programme aimed at minimising foodborne pathogens (MPI, no date). It ensures that all primary processors of meat, poultry, game and ratites in New Zealand produce food suitable for human consumption and meet microbiological standards for food sold domestically and the export requirements set by New Zealand's trading partners (MPI, no date).

2.4 Evolution of *Escherichia coli*

Biological evolution is defined by the adaptation of organisms to their natural habitat and the divergence of populations and species from each other. Several mechanisms, such as natural selection, chance events, genetic differences caused by point mutations and horizontal gene transfer (HGT) contribute to the evolution and divergence of microorganisms (Lenski *et al.*, 1991, Wright, 1982, Ochman *et al.*, 2000, Gogarten & Townsend, 2005).

Point mutations cause the modification, inactivation or differential regulation of existing genes in a genome (Ochman *et al.*, 2000, Papadopoulos *et al.*, 1999). Horizontal gene transfer (HGT), also known as lateral gene transfer (LGT), involves the movement of genetic information between more or less distantly related organisms (Keeling & Palmer, 2008). HGT contributes to the loss and gain of virulent and non-virulent features in the process of evolution or between lineages (Zhang *et al.*, 2007).

A study by Whittam *et al*. (1988) first looked to understand the evolution of *E. coli* O157 using the multilocus enzyme electrophoresis (MEE) method. The method detects the rate of protein migration, determined by its amino acid sequence, during electrophoresis by measuring the mobility variants of enzymes (Selander *et al.*, 1986). The analysis of various *E. coli* O157 and non-O157 strains showed that the O157 strains grouped together indicating close genetic association (Whittam *et al.*, 1988). In another study, Whittam *et al*. (1993) used MEE to study genetic relatedness between various *E. coli* isolates of 16 serotypes, including O26:H11, O55:H6, O55:H7, O111:H2, O128:H2 and O157:H7. The O157:H7 and O55:H7 clones were found to be closely genetically related. The authors concluded that the O157:H7 and O55:H7 isolates were derived from a common ancestral cell. They also proposed the idea that the O157:H7 isolate evolved from an O55:H7-like progenitor cell (Whittam *et al.*, 1993).

E. coli O157:H7 lack beta-glucuronidase (GUD) activity and do not ferment sorbitol, unlike most *E. coli* strains in which the enzyme glucuronidase is present (Farmer & Davis, 1985, Doyle & Schoeni, 1984, Kilian & Bülo, 1976). A model of the evolution of *E. coli* O157 proposed by Feng *et al*. (1998) postulated the evolution of O157:H7 from *E. coli* O55:H7- like progenitor strain, shown in Figure 2.4. The O157:H7 clone evolved by acquiring the *stx2* gene and the *rfbE* gene. The *rfbE* gene is important in the expression and biosynthesis of the O157 antigen (Bilge *et al.*, 1996, Bertrand & Roig, 2007). This O157:H7 clone which was both a sorbitol fermenter and GUD positive diverged into two clones O157:H7 (sorbitol non-fermenter, GUD positive, *stx1* gene, motile) and O157:H- (sorbitol fermenter, GUD positive, non-motile). The model proposed that *stx2* gene was acquired before other virulence factors such as *stx1* and the plasmid pO157 were acquired. The model also postulated the subsequent loss of GUD (also a sorbitol non-fermenter) which produced another clone (Feng *et al.*, 1998).

Unlike *E. coli* O157, the evolution of non-O157 *E. coli* is not well understood. However, studies have been carried out to investigate the characteristics and genetic composition of non-O157 *E. coli* strains. The genome sequences of three clinical isolates from Japan, one of each belonging to O26-, O111- and O103- serogroups, were compared to that of O157 Sakai isolate (Ogura *et al.*, 2009). The whole gene repertoire of the non-O157 isolates was found to be similar to that of the O157 isolate, including *stx* and other virulence genes. The genome analysis of the three non-O157 clinical isolates showed that multiple lambdoid prophages, integrative elements and virulence plasmids were independently acquired by HGT. The independent acquisition of these mobile genetic elements drove the parallel evolution of each of the serogroups (Ogura *et al.*, 2009).

Differences in evolution of non-O157 *E. coli* from *E. coli* O157 have also been reported. Insertion sites for the LEE (attaching and effacing of *E. coli*) vary in serogroups O26 and O111 compared to O157 serogroups (Wieler *et al.*, 1997). Nevertheless, further studies are required to understand the similarities and differences in the evolution of and between various non-O157 *E. coli* serogroups.

The evolutionary relationship of the STEC serogroups O157 and O26 was investigated as an objective of this piece of research to understand the similarities and differences in their genetic composition.

Figure 2.3. Meat carcasses in an abattoir.
Source: (Rochan, 2013)

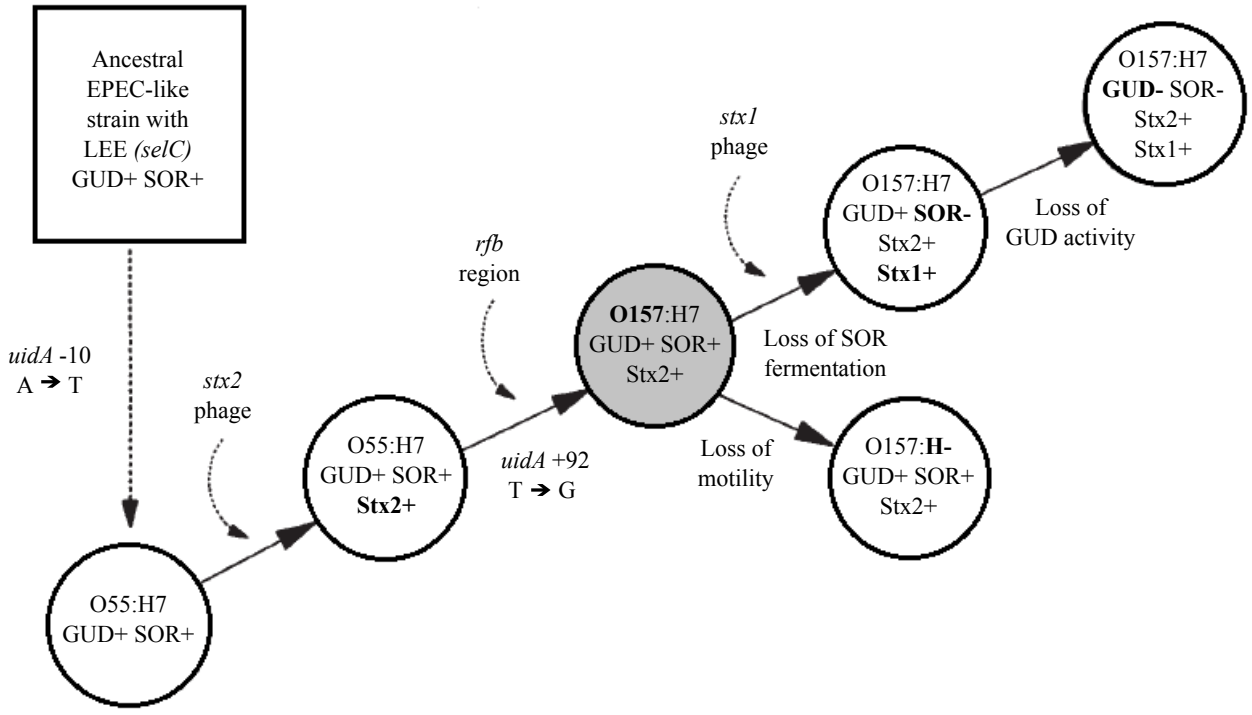

Figure 2.4. Proposed evolutionary model of *E. coli* O157:H7. The emergence of the O157:H7 complex based on mutations in *uidA*, Stx production, and SOR and GUD phenotypes. The changes predicted to have occurred at each evolutionary event are in bold.
selC: insertion loci of LEE, GUD: beta-glucuronidase, *uidA*: gene encoding GUD, SOR: sorbitol
Source: (Feng *et al*., 1998)

2.5 Carbohydrate metabolism in bacteria

Bacteria must adapt and co-ordinate the metabolism of nutrients to survive and replicate in host cells and under various physical conditions. Multiple uptake pathways allow the transport and use of nutrients and substrates made available to the bacteria (Muñoz-Elías & McKinney, 2006, Brown *et al.*, 2008). *E. coli* rely on phosphotransferase system (PTS) dependent carbon sources such as glucose, sucrose and non-PTS dependent carbon sources such as lactose, arabinose, xylose, sorbitol, ribose and rhamnose to enhance its grow and survival in different environments (Sabri *et al.*, 2013, Aidelberg *et al.*, 2014).

The transport and phosphorylation of a large number of carbohydrates in bacteria is regulated by the phosphoenolpyruvate (PEP): carbohydrate phosphotransferase system. The PEP: PTS consists of two proteins, enzyme I and heat-stable protein (HPr), and numerous sugar-specific permeases, called enzymes II (EII). These are composed of three to four domains, IIA, IIB, IIC and the others, which are membrane bound. In bacteria, when a substrate is transported through the cytoplasmic membrane, the phosphoryl group is transferred from PEP to HPr, through EI using PEP as a source of energy. The phosphoryl group is then transferred to EIIA, EIIB and EIIC. The carbohydrate is subsequently metabolised (Kotrba *et al.*, 2001, Postma & Lengeler, 1985, Postma *et al.*, 1993). Induction and carbon catabolite repression mechanism regulate the synthesis of enzymes to ensure important substrates are utilised (Stülke & Hillen, 1999).

Glycolysis, the pentose phosphate pathway (PPP) and the Entner-Doudoroff pathway (EDP) are among the best characterised pathways involved in the catabolism of carbohydrates in bacteria. Schematic diagrams of the glucose metabolism, fatty acid catabolism, citric acid cycle, gluconeogenesis and anaplerotic pathways are shown in Figures 2.5 and 2.6 as examples (Muñoz-Elías & McKinney, 2006). In general, β-oxidation cycle causes the oxidative degradation of fatty acids to acetyl coenzyme A (acetyl-CoA). Acetyl-CoA, in the citric acid cycle, is oxidised to carbon dioxide (CO_2) and produces reducing equivalents, adenosine triphosphate (ATP) and biosynthetic precursors. Glycolysis and gluconeogenesis also provides precursors for biosynthetic pathways. The citric acid cycle intermediates produced from the decarboxylation reactions are replenished by the process of anaplerosis (Entner & Doudoroff, 1952, Kunau *et al.*, 1995, Clark & Cronan, 2005, Textor *et al.*, 1997, Anderson & Wood, 1969, Muñoz-Elías & McKinney, 2006).

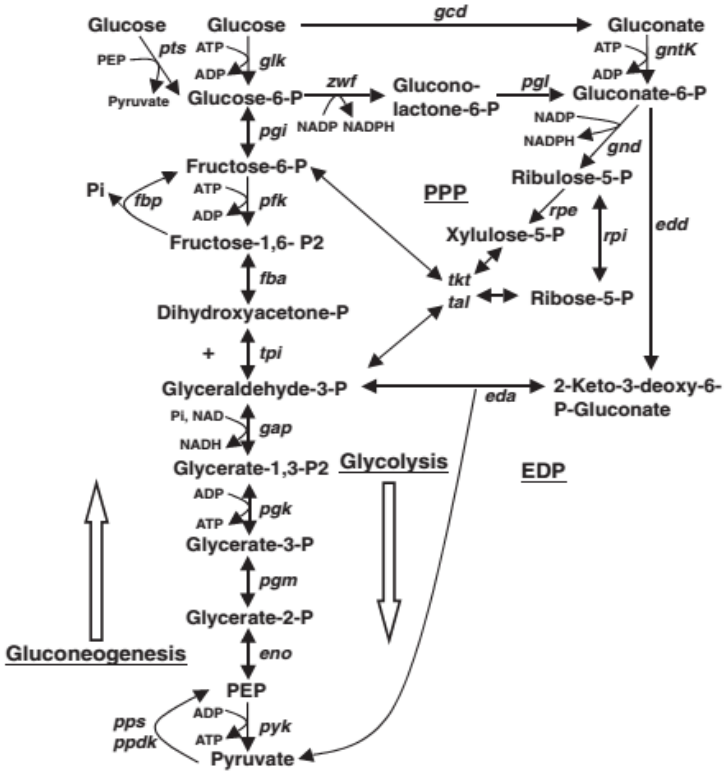

Figure 2.5. Major pathways of carbohydrate metabolism in prokaryotes.
The best characterised pathways of carbohydrate catabolism in bacteria are glycolysis, the pentose phosphate pathway (PPP) and the Entner–Doudoroff pathway (EDP). Glucose anabolism is accomplished via gluconeogenesis. Enzymes common to glycolysis and gluconeogenesis are indicated by double-headed arrows. Irreversible glycolytic reactions are catalysed by phosphofructokinase and pyruvate kinase. Irreversible gluconeogenic reactions are catalysed by fructose-1,6-bisphosphatase and PEP synthase (or, in some species, pyruvate phosphate dikinase). Gene names are italicised.
pts, PEP:glucose phosphotransferase system; *glk*, glucose kinase; *pgi*, phosphoglucose isomerase; *pfk*, phosphofructokinase; *fbp*, fructose-1,6-bisphosphatase; *fba*, fructose-1,6-bisphosphate aldolase; *tpi*, triosephosphate isomerase; *gap*, glyceraldehyde 3-phosphate dehydrogenase; *pgk*, phosphoglycerate kinase; *pgm*, phosphoglyceratemutase; *eno*, enolase; *pyk*, pyruvate kinase; *pps*, PEP synthase; *ppdk*, pyruvate phosphate dikinase; *gcd*, glucose dehydrogenase; *gntK*, gluconate kinase; *zwf*, glucose-6-phosphate dehydrogenase; *pgl*, 6-phosphogluconolactonase; *edd*, 6-phosphogluconate dehydratase; *eda*, 2-keto-3-deoxy-6-phosphogluconate aldolase; *gnd*, 6-phosphogluconate dehydrogenase; *rpe*, ribulose-5-phosphate epimerase; *tkt*, transketolase; *tal*, transaldolase; *rpi*, ribose-5-phosphate isomerase.
Source: (Muñoz-Elías & McKinney, 2006)

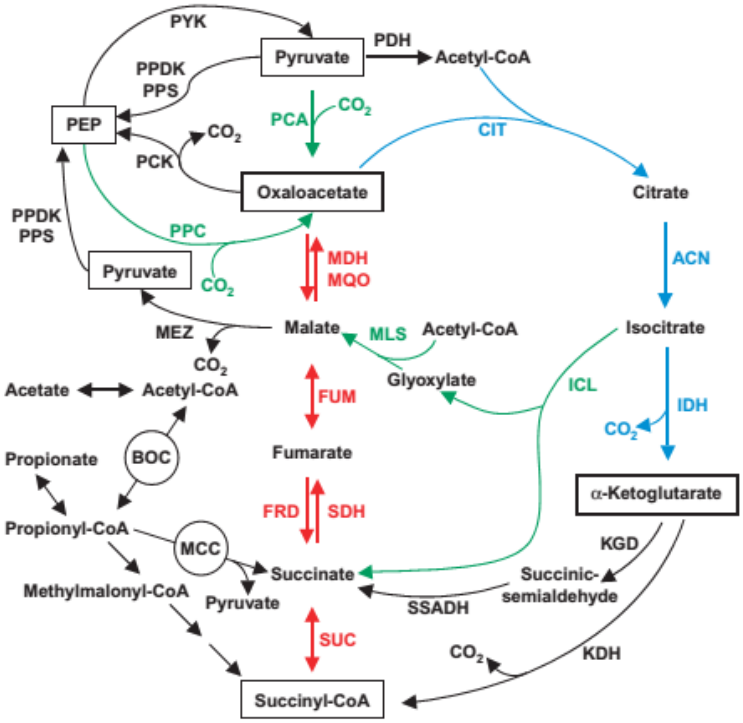

Figure 2.6. Citric acid cycle and related anaplerotic pathways.
The bifurcated biosynthetic mode of operation of the citric acid cycle, comprising oxidative (blue) and reductive (red) branches, is employed by bacteria growing under aerobic conditions in excess glucose or anaerobically on any carbon substrate, and by bacteria that lack KDH/KGD. Some intermediates of glycolysis and the citric acid cycle (boxed) are also diverted into biosynthetic pathways. Depleted intermediates are replenished via anaplerotic reactions (green), catalysed by PPC or PCA during growth on excess carbohydrate, or by ICL and MLS during growth on fatty acids or limiting carbohydrate. MEZ, malic enzyme; PPC, PEP carboxylase; PCK, PEP carboxykinase; PPS, PEP synthase; PPDK, pyruvate phosphate dikinase; PYK, pyruvate kinase; PCA, pyruvate carboxylase; PDH, pyruvate dehydrogenase; CIT, citrate synthase; ACN, aconitase; IDH, isocitrate dehydrogenase; KGD, α-ketoglutarate decarboxylase; SSADH, succinic semialdehyde dehydrogenase; KDH, α-ketoglutarate dehydrogenase; SUC, succinyl-CoA synthetase; SDH, succinate dehydrogenase; FRD, fumarate reductase; FUM, fumarase; MDH, malate dehydrogenase; MQO, malate:quinone oxidoreductase; ICL, isocitrate lyase; MLS, malate synthase; BOC, β-oxidation cycle; MCC, methylcitrate cycle.
Source: (Muñoz-Elías & McKinney, 2006)

2.6 Phenotype and genotype

Phenotype refers to the observable properties of an organism such as its morphology, physiology, behaviour and ecological aspects. The phenotypic information derived from proteins and different chemotaxonomic markers reveals what role the species plays in its environment, including what it requires or feeds on for survival and its interaction with other species (Griffiths *et al.*, 2000, Vandamme *et al.*, 1996, Fenchel & Finlay, 2006). Traditional phenotypic identification methods include observation of growth, pigmentation and colony morphology on various media and analysis of individual biochemical reactions (Cloud *et al.*, 2010, Stager & Davis, 1992).

Genotypic information is derived from the nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), present in a microbe. Restriction fragment length polymorphism (RFLP), pulsed field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), PCR for single locus and multilocus sequence typing (MLST) are among a few methods that are used to study genetic diversity in microbes (Virdi & Sachdeva, 2005).

Methods to analyse genetic diversity within and between microorgansms that exist and grow in the environment are important to understand the complexities of the microbial world, the evolution of microbial commensals and pathogens. Staple mechanisms such as point mutation, gene duplication, gene rearrangements, gene loss and HGT have been repeatedly shown to play an important role in generating genetic diversity among microbial pathogens (Virdi & Sachdeva, 2005).

Studies comparing genomes of related microorganisms with genetic differences demonstrated a correlation to phenotypic changes among microbes (McClelland *et al.*, 2000, Bentley & Parkhill, 2004, Mukherjee *et al.*, 2006). The 2011 outbreak of enteroaggregative *E. coli* O104:H4 in Germany illustrated the power of DNA sequencing and its vital role in public health. The rapid sequencing of the outbreak associated isolates and related strains and the rapid release of DNA sequence data by the Beijing Genomics Institute (BGI) allowed researchers to study the causative agent at a faster pace (Rasko *et al.*, 2011).

2.7 Phenotypic microarray technology

Bochner and Savageau, in 1976, determined a method to discern metabolic differences between *Salmonella typhimurium* and *E. coli*. The authors developed an indicator plate with diverse types of substrates. Other components required were agar, buffer, growth-supporting nutrients, a test substrate and 2,3,5-triphenyl tetrazolium chloride (TTC). The authors noted that when the microorganism or colonies catabolised a test substrate, it caused the reduction of TTC producing a deep red colour and that intermediate rates of catabolism affected the rate and/or extent of colour formation. When the microorganism or colonies failed to catabolise, the substrate remained colourless (Bochner & Savageau, 1977). This study served as a commercial test and birthed the development of phenotypic microarray (PM) technology by Biolog Inc (Hayward, USA) (Bochner, 2009, Bochner *et al.*, 2001).

A standardised and high throughput technology, PM technology (Biolog Inc, Hayward, USA) allows for the characterisation of around 2,000 different media conditions. Eight pre-formulated 96-well microplates (PM1-PM8) serve as sources of carbon, nitrogen, peptide nitrogen, sulphur and phosphorus. The remaining 17 (PM9-PM25) are used to determine sensitivity to environmental stresses such as osmolytes, potential of hydrogen (pH), and chemical agents including antibiotics. Seven different redox dyes (redox dye mix A-H) (Biolog Inc, Hayward, USA) are available and are chosen based on the microorganism being studied. The dyes detect the respiration of cells colourimetrically.

The physical flow of electrons from the carbon source to co-enzyme nicotinamide adenine dinucleotide (NADH), down the electron transport chain of the cell onto the dye produces the colour change, as illustrated in Figure 2.7. This reduction of the dye is called a positive reaction, where the colonies/cell catabolised a substrate allowing its growth. When one of these processes is not working at full capacity, the colour change is reduced. This is called a weak reaction. Likewise, a loss of function or gain through gene acquisition, results in no growth or colour change. This is called a negative reaction (Bochner *et al.*, 2001).

The OmniLog® instrument (Biolog Inc, Hayward, USA) can hold fifty 96 well plates and is developed to read and record the colour change in the PM assays. A monochrome CCD camera records the colour change every 15 minutes (min). The OmniLog® software (Biolog Inc, Hayward, USA) is used to plot the output generated from each run (Bochner *et al.*, 2001,

Mukherjee *et al.*, 2006). Figure 2.8 depicts the stages involved in an OmniLog® run, from the preparation of PM plates to obtaining PM data.

The PM technology was used in the current research to characterise the utilisation of carbon sources by *E. coli* O157 and O26 New Zealand isolates.

2.8 Whole genome sequencing

DNA sequencing is used in a number of scientific fields such as molecular sciences, genetics, biotechnology, archaeology and forensics among others, aiding in new discoveries in each of these fields. In 1975, Sanger and Coulson first developed a method called 'plus and minus' to sequence DNA which was inefficient (Sanger & Coulson, 1975). Following this, Sanger and his co-workers designed a method to sequence oligonucleotides through enzymatic polymerisation. This method, which revolutionised the field of genomics, was called the chain termination method or the dideoxynucleotide method. The catalysed enzymic reaction caused the DNA fragments to polymerise complementary to the template DNA of interest (Sanger *et al.*, 1977, França *et al.*, 2002).

Other methods that were used to sequence DNA include a chemical degradation method by Maxam and Gilbert (Maxam & Gilbert, 1977), shotgun sequencing (Rowen & Koop, 1994) and primer walking sequencing (Studier, 1989, Martin-Gallardo *et al.*, 1992). Methods such as pyrosequencing (Nyrén & Lundin, 1985, Ronaghi *et al.*, 1996) and single-molecule sequencing (Jett *et al.*, 1989) were also developed for DNA sequencing. Venter and co-authors, in 1996, introduced a variation to the shotgun sequencing method which was proven successful with the sequencing of *Haemophilus influenza* genome in 18 months (Fleischmann *et al.*, 1995, Venter *et al.*, 1996).

The human genome project was carried out from 1993 to 2003 using the slow process of bacterial artificial chromosome (BAC) based sequencing (Collins & Galas, 1993, Collins *et al.*, 1998). This project promised a profound contribution to biomedical research and medicine and revolutionised the approach to genome sequencing giving way to a rapid method called high-throughput sequencing. High-throughput sequencing technology was used to sequence the human genome for the first time in 2008 which was completed in a brief period of two months and was significantly less expensive to perform (Wheeler *et al.*, 2008, Mardis, 2008b).

High-throughput sequencing technologies produce enormous volumes of data. Illumina-based sequencing is now the dominant high-throughput sequencing platform (Fedurco *et al.*, 2006, Turcatti *et al.*, 2008).

2.8.1 Illumina technology

The Illumina technology was first launched commercially when the Illumina Genome Analyzer (Illumina Inc, USA) was introduced in 2006. Illumina Inc, USA more recently released the HiSeq 2000 (released in 2010), HiSeq 2500 (2012) and MiSeq (2011) instruments (Kulkarni & Pfeifer, 2014). The technology uses the synthesis by synthesis (SBS) approach (Bentley *et al.*, 2008). The steps involved in the Illumina workflow are illustrated in Figure 2.9. The steps described here was originally for the Genome Analyzer. However, the technology is broadly similar across Genome Analyzer, HiSeq and MiSeq instruments with subtle variations.

The library is denatured to form a mixture of single-stranded, adaptor oligoligated DNA fragments. These are attached to the surface of a glass flow cell. This step is performed using a microfluidic cluster station in HiSeq, MiSeq instruments. Each flow cell is divided into one (MiSeq), two (HiSeq 2500) and eight (HiSeq 2000, HiSeq 2500, Genome Analyzer) separate lanes. The interior surfaces have covalently attached oligonucleotides complementary to the specific adapters that are ligated onto the library fragments. Active heating and cooling is used to hybridise the DNA onto the oligonucleotides on the flow cell followed by incubation with reactants and an isothermal polymerase (bridge amplification). The flow cell is placed into a fluidics cassette within the sequencer (Genome Analyzer). Four fluorescent nucleotides of a different colour with a blocking group attached and polymerase are added to each cluster. Captured by a CCD camera, at each base incorporation cycle, the added nucleotides compete for binding sites on the template. A chemical step removes the unbound fluorescent group and deblocks the 3' end for the next base incorporation cycle. The Illumina data from each run is subjected to quality filtering to remove poor-quality sequences (Mardis, 2008a, Mardis, 2008b, Buermans & den Dunnen, 2014).

The Genome Analyzer instrument generates up to 95 gigabases (Gb) data with 640 million reads per run. The HiSeq 2000 and HiSeq 2500 instruments are used for high-throughput parallel sequencing and generates up to 600 Gb data with 6 billion reads per run. The MiSeq instrument is used for lower-throughput sequencing and generates up to 15 Gb data with 25

million reads in a single run (Kulkarni & Pfeifer, 2014, Caporaso *et al.*, 2012, Quail *et al.*, 2012).

The Illumina MiSeq analyser was used in the current research to generate entire genomes for the *E. coli* O157 and O26 New Zealand isolates.

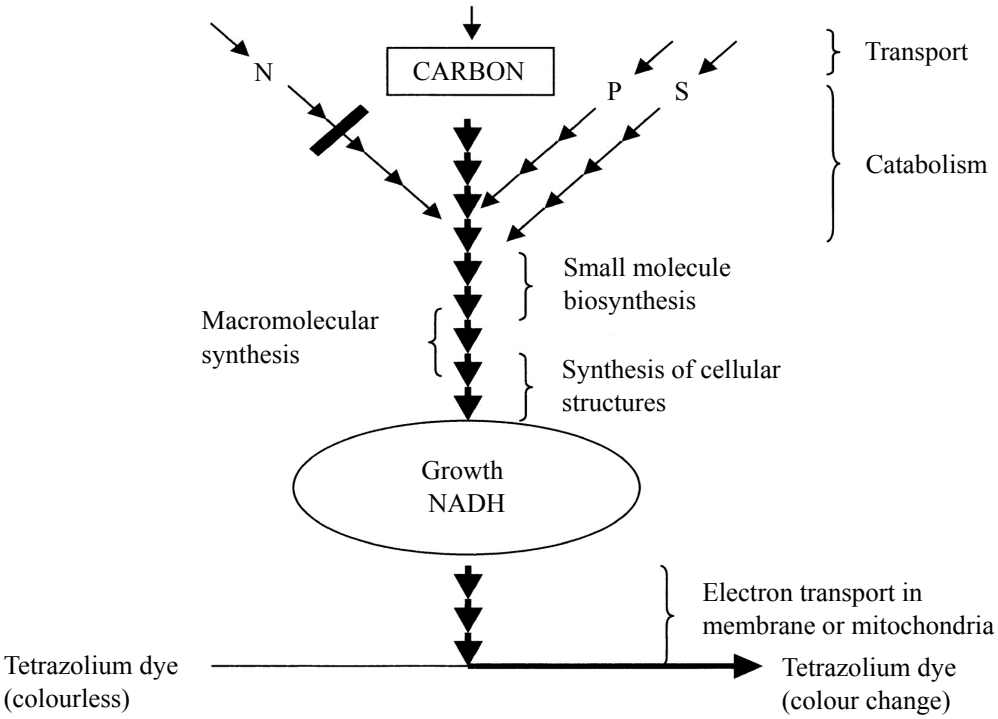

Figure 2.7. Respiration pathways coupled to cell physiology.
Source: (Bochner *et al*., 2001)

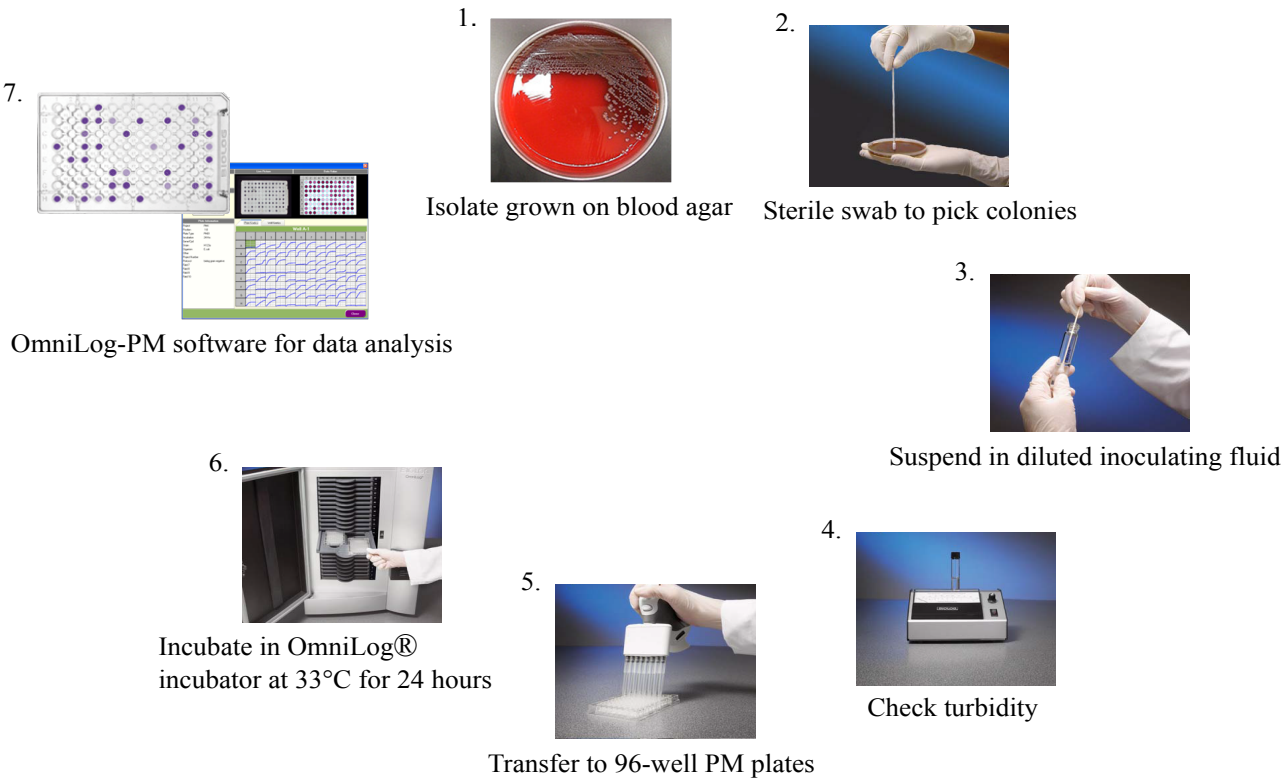

Figure 2.8. Phenotypic microarray (PM) technology. An overview of the steps involved in the preparation of PM plates and obtaining high-throughput results for the phenotypic study.
Source: (Biolog Inc., 2008)

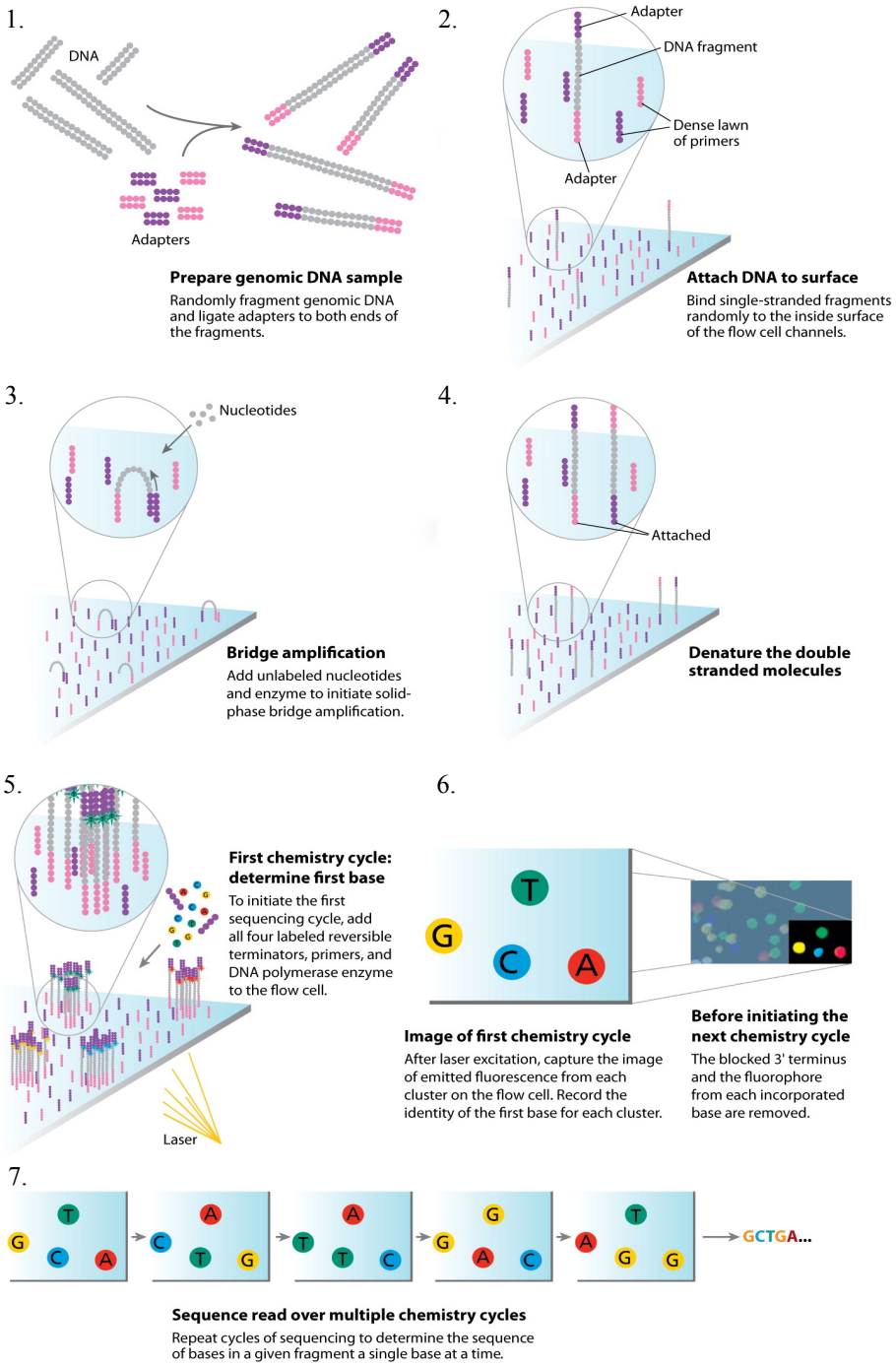

Figure 2.9. Illumina workflow. Beginning with fragmentation and adapter ligation steps, the library is added to a flow cell for bridge amplification (an isothermal process that amplifies each fragment into a cluster). The cluster fragments are denatured, annelaed with a sequencing primer and subjected to sequencing by synthesis using 3' blocked labelled nucleotides.
Source: (Mardis, 2008a, Mardis, 2008b)

THREE

Materials and methods

3.1 Bacterial isolates

Four *E. coli* O157 and six *E. coli* O26 New Zealand (NZ) isolates (Jaros, 2014) were grown on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ) and incubated overnight at 37°C (Table 3.1). Pure cultures were suspended in glycerol broths and stored at -80°C.

The 10 isolates were compared to 13 other *E. coli* O157 isolates (Table 3.2) from a previous study carried out in 2012 (George, Massey University). These 23 isolates were used for the phenotypic microarray (PM) studies, confirmatory tests for phenotypes, core genome comparison and Clusters of Orthologous Groups of proteins (COGs) function analysis.

All 23 *E. coli* isolates were previously screened for serotypes and the presence/absence of virulence genes by culture methods, latex tests and multiple typing methods including PCR (Jaros, 2014). The results are summarised in both Tables 3.1 and 3.2. For the purpose of analysis, the *E. coli* O157 isolates were divided into sorbitol non-fermenting or negative (SN) and sorbitol fermenting or positive (SP). The *E. coli* O26 isolates were also divided into rhamnose non-fermenting or negative (RN) and rhamnose fermenting or positive (RP) groups.

3.2 Shiga toxin-encoding bacteriophage insertion typing

Shiga toxin-encoding bacteriophage insertion typing is a method to detect the four known bacteriophage insertion sites which indicate the presence or absence of inserted *stx* genes (*stx1*, *stx2* and subtype *stx2c*) in *E. coli* O157 isolates (Besser *et al.*, 2007, Shaikh & Tarr, 2003). In order to type the genomes under the new scheme designation, in conjunction with Dr Biggs and Dr Jaros, Massey University, and using the PCR primers for the amplicons, a synthetic PCR scheme was defined to type the *E. coli* genomes.

For each of the insertion loci (*argW*, *sbcB*, *wrbA* and *yehV*), the sequences of DNA between the PCR primers were searched against a set of *E. coli* O157 and O26 genomes. The PCR primer sequences used for all four insertion loci are shown in Table 3.3. In all cases, these synthetic sequences covering the region with no insertion, and the insertion for the left and right flanking product were then used in genomic Basic Local Alignment Search Tool (BLAST) searches. There was a variant junction for *yehV*, which may or may not occur. Only a single PCR product was used for detecting the three *stx* variants: *stx1*, *stx2* and *stx2c*.

Through a set of training datasets, the parameters for each insertion site or *stx* gene were resolved empirically to detect the correct combination on genomes with independently verified SBI insertion information. Only at this point were new genomes tested. The naming of the insertion types is based on the presence of at least two products at any junction except for *yehV* which also includes a variant junction. This is shown in Table 3.4. Also, each of the junctions has a specific label which forms the basis for the naming of the insertion types, *argW*: A, *sbcB*: S, *wrbA*: W, *yehV*: Y, *stx1*: 1, *stx2*: 2a and *stx2c*: 2c. For instance, SBI type ASY2c, as indicated by the orange shaded area in Table 3.4, includes two *argW* products (A), two *sbcB* products (S), one *wrbA* product (therefore no insertion), two *yehV* products (Y) at each of the respective junctions and also includes the *stx2c* gene (2c).

Table 3.1. Features of *E. coli* O157 and O26 New Zealand isolates.

Isolate name	Species	Island	Serotype	Plate type	Colony colour	PCR					
						***stx* 1**	***stx* 2**	***stx* 2c**	***eae***	***hlyA***	**H7**
EcCa21a	Adult beef	NI	O157	CT-SMAC	Pink	Negative	Negative	Negative	Negative	Negative	Negative
EcCa24a	Calf	SI	O157	CT-SMAC	Pink	Negative	Negative	Negative	Negative	Negative	Negative
EcCa519a	Adult beef	NI	O157	CT-SMAC	Pink	Negative	Negative	Negative	Negative	Negative	Negative
EcCa583a	Adult beef	NI	O157	CT-SMAC	Pink	Negative	Negative	Negative	Negative	Negative	Negative
EcCa5a	Calf	SI	O26	CT-RMAC	Pink	Negative	Negative	Negative	Positive	Negative	NA
EcCa8a	Calf	SI	O26	CT-RMAC	Grey	Negative	Negative	Negative	Positive	Positive	NA
EcCa513a	Calf	NI	O26	CT-RMAC	Grey	Positive	Negative	Negative	Positive	Positive	NA
EcCa514a	Calf	NI	O26	CT-RMAC	Pink	Negative	Negative	Negative	Positive	Positive	NA
EcCa569a	Adult beef	SI	O26	CT-RMAC	Pink	Negative	Negative	Negative	Positive	Negative	NA
EcCa1a	Calf	SI	O26	CT-RMAC	Grey	Positive	Negative	Negative	Positive	Positive	NA

NI: North Island, SI: South Island, CT-SMAC: MacConkey agar with sorbitol, cefixime, and tellurite, CT-RMAC: MacConkey agar with rhamnose, cefixime, and tellurite, PCR: Polymerase chain reaction, *stx1, stx2, eae, hlyA* : Virulence/toxin genes, H7: Flagellar antigen, NA: Not applicable

Table 3.2. Features of *E. coli* O157 New Zealand isolates from previous study. These isolates were used to compare the findings of isolates from Table 3.1.

Isolate name	Species	Island	Serotype	Serotype ERL	Plate type	Colony colour	PCR					
							***stx* 1**	***stx* 2**	***stx* 2c**	***eae***	***hlyA***	**H7**
EcCa452a	Calf	SI	O157		CT-SMAC	Grey	Negative	Positive	Positive	Positive	Positive	Positive
EcCa63b	Adult beef	NI	O157		CT-SMAC	Grey	Negative	Positive	Negative	Positive	Positive	Positive
EcCa491a	Calf	NI	O157		CT-SMAC	Grey	Negative	Positive	Negative	Positive	Positive	Positive
EcCa32a	Adult beef	NI	O157		CT-SMAC	Grey	Positive	Positive	Negative	Positive	Positive	Positive
EcCa461b	Calf	NI	O157		CT-SMAC	Grey	Positive	Positive	Negative	Positive	Positive	Positive
EcCa635a	Adult beef	SI	O157		CT-SMAC	Grey	Negative	Positive	Positive	Positive	Positive	Positive
EcCa26a	Calf	NI	O157		CT-SMAC	Pink	Negative	Positive	Negative	Positive	Positive	Negative
EcHu44a	Human	NI	O157	O157:H7	CT-SMAC	Grey	Negative	Positive	Negative	Positive	Positive	
EcHu93a	Human	SI	O157	O157:H7	CT-SMAC	Grey	Negative	Positive	Positive	Positive	Positive	
EcHu56a	Human	SI	O157	O157:H7	CT-SMAC	Grey	Positive	Positive	Negative	Positive	Positive	
EcHu76a	Human	NI	O157	O157:H7	CT-SMAC	Grey	Positive	Positive	Negative	Positive	Positive	
EcHu68a	Human	NI	O157	O157:[H7]	CT-SMAC	Grey	Negative	Positive	Positive	Positive	Positive	
EcHu114a	Human	SI	O157	O157:[H7]	CT-SMAC	Grey	Negative	Positive	Positive	Positive	Positive	

NI: North Island, SI: South Island, Serotype ERL: Serotype information obtained from the Enteric Reference Laboratory (ERL) (The Institute of Environment Science and Research Limited (ESR), New Zealand), CT-SMAC: MacConkey agar with sorbitol, cefixime, and tellurite, PCR: Polymerase chain reaction, *stx1, stx2, eae, hlyA* : Virulence/toxin genes, H7: Flagellar antigen, [H7]: Flagella present but not motile

Table 3.3. Primers used for PCR in SBI typing.

Target	Sequences (5'-3')
stx1	CAGTTAATGTGGTGGCGAAG
	CACAGACTGCTGCAGTGAGG
stx2	AGGATGACACATTTACAGTGAAGGTT
	CACAGGTACTGGATTTGATTGTGAC
stx2c	CGACAGGCCCGTTATAAAAA
	GGCCACTTTTACTGTGAATGTATC
argW insertion site region	CCGTAACGACATGAGCAACAAG
	AATTAGCCCTTAGGAGGGGC
Bacteriophage *argW* left junction	GCATCTCACCGACGATAACA
	AATTAGCCCTTAGGAGGGGC
Bacteriophage *argW* right junction	CCGTAACGACATGAGCAACAAG
	GCAGTATCACGCAGAGCTGAAG
sbcB insertion site region	GACAGCAGAAACAACGGATTTAAC
	TCCAGGCGTAAGGATCGTAG
Bacteriophage *sbcB* left junction	GTGCTTGGGTCTTTTCTCTG
	TCCAGGCGTAAGGATCGTAG
Bacteriophage *sbcB* right junction	GACAGCAGAAACAACGGATTTAAC
	CCAGCGTGGGATAAAAGAGAATC
wrbA insertion site region	AGGAAGGTACGCATTTGACC
	CGAATCGCTACGGAATAGAGA
Bacteriophage *wrbA* left junction	CCGACCTTTGTACGGATGTAA
	CGAATCGCTACGGAATAGAGA
Bacteriophage *wrbA* right junction	AGGAAGGTACGCATTTGACC
	ATCGTTCGCAAGAATCACAA
yehV insertion site region	AAGTGGCGTTGCTTTGTGAT
	AACAGATGTGTGGTGAGTGTCTG
Bacteriophage *yehV* left junction	CACCGGAAGGACAATTCATC
	AACAGATGTGTGGTGAGTGTCTG
Bacteriophage *yehV* right junction	AAGTGGCGTTGCTTTGTGAT
	GATGCACAATAGGCACTACGC
Variant bacteriophage *yehV* right junction	AAGTGGCGTTGCTTTGTGAT
	AGCGATACAGATCTCAACAC

Table 3.4. SBI typing summary of *E. coli* isolates.

SBI type	Target region	***argW***			***sbcB***			***stx1***	***stx2***	***stx2c***	***wrbA***			***yehV***			
		argW	*argW* _lj	*argW* _rj	*sbcB*	*sbcB* _lj	*sbcB* _rj	*stx1*	*stx2*	*stx2c*	*wrbA*	*wrbA* _lj	*wrbA* _rj	*yehV*	*yehV* _lj	*yehV* _rj	*yehV* _vrj
	Product size	216	462	583	406	915	730	478	126	243	314	537	592	340	824	702	295
AY2a		No	Yes	Yes	Yes	No	No	No	Yes	No	Yes	No	No	No	Yes	Yes	Variable
WY12a		Yes	No	No	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	Variable
ASY2c		No	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	No	No	No	Yes	Yes	Variable
SY2c		Yes	No	No	No	Yes	Yes	No	No	Yes	Yes	No	No	No	Yes	Yes	Variable

SBI: Shiga toxin-encoding bacteriophage insertion, lj: Left junction, rj: Right junction, Yes: Region contains a product, No: Region does not contain a product, Variable: Region may or may not contain a product, Orange shaded area indicates the target region that defines each SBI type

3.3 Phenotypic microarray technology

In the current study, phenotypic microarray technology was used to test the utilisation of various carbon sources by four *E. coli* O157 and six O26 isolates. The PM1 plate comprises of 95 carbon compounds of alcohol, amine, amino acid, carbohydrate, carboxylic acid, ester and fatty acid. The PM2A plate comprises of 95 carbon compounds of alcohol, amide, amine, amino acid, carbohydrate, carboxylic acid, ester and polymer (Kalai Chelvam *et al.*, 2015).

3.3.1 Preparation of phenotypic microarray plates

Phenotypic microarray plates PM1, PM2A, inoculating fluid-0a (IF-0a GN/GP Base (1.2x) and redox dye mix A (tetrazolium dye) (100x) (Biolog Inc, Hayward, USA) were used to test for utilisation of 190 different carbon sources. For each assay, a suspension of *E. coli* cells was prepared by inoculating a tube of 16 millilitres (ml) IF-0a with pure culture colonies. A turbidimeter (Biolog Inc, Hayward, USA) was used to measure a density of 42 percent (%). From this suspension, 3 ml was transferred to a mixture containing 12.5 ml of undiluted IF-0a, 2.32 ml MilliQ water and 0.18 ml dye mix A. The final cell density of 85% was reached. This mixture was used to inoculate the PM plate including the negative control with 100 microlitres (µl)/well. The first well in each PM plate was the negative control which contained no substrate. The plate was labelled and incubated in the OmniLog® incubator (Biolog Inc, Hayward, USA) at 33°C for 24 hours (h). This was performed for all isolates using PM1 and PM2A plates.

All assays were repeated on different days to obtain replicates. The number of replicate assays was not set a priori and some isolates had up to four replicates. Negative plates were prepared for both PM1 and PM2A plates. The inoculating fluid mixture was prepared as mentioned above but without the addition of the suspension of *E. coli* cells.

For the previously tested 13 *E. coli* O157 isolates in PM1 and PM2A, duplicate assays were only performed on seven *E. coli* O157 isolates at random (5 in PM1 and 2 in PM2A) (George, Massey University, 2012). Replicate assays were performed on all isolates. The number of replicate assays was not set a priori.

3.3.2 Data recording and analysis of kinetic data

A charged coupled device (CCD) camera recorded the respiration in both PM1 and PM2A assays every 15 min over 24 h in the OmniLog®. In both PM1 and PM2A 96-well plates, each well contained a unique metabolite. In each PM plate, the negative control would result in no signal. When the bacteria respired by utilising the metabolite, a purple colour was visible from the reduction of the redox dye. This colour intensity and digital images of the plates were captured by the camera for 24 h at 15 min intervals and stored in electronic files (Bochner *et al.*, 2001). These files were accessed and analysed using OmniLog-PM software (version (v) 1.20.02) (Lim *et al.*, 2010a). One of the outputs of the OmniLog-PM analysis included a kinetic response curve. The area under the curve provided a quantitative measure for the kinetic response.

The raw data obtained from the OmniLog-PM software were then used in software R (v 3.1.2) (Ihaka & Gentleman, 1996) to visualise the PM assays as a heatmap. The PM data of previously tested 13 *E. coli* O157 isolates including the replicate and duplicate assays were combined with the data obtained from the four *E. coli* O157 isolates. The R package *opm* was used to convert the raw data into opm files (Vaas *et al.*, 2012). Packages including *lattice* (Sarkar, 2008) with the function *levelplot()* along with *gplots* (Warnes *et al*., 2013) and *RColorBrewer* (Neuwirth, 2011) were used to generate a heatmap. On the heatmap, the purple shaded areas indicated greater respiration; the green areas indicated minimal respiration whereas grey indicated no respiration of the isolates. The function *xyplot()* was used to plot respiration curves for a more detailed view of all the repeats per isolate in a grid-like structure.

The *opm* package was used in software R (v 3.1.2) to generate a table where the '+' indicated positive reaction/respiration, '-' indicated negative reaction/no respiration and 'w' indicated poor reaction/respiration of the substrates. A table was generated for both PM1 and PM2A across all assays for each isolate including the previously tested 13 *E. coli* O157 isolates, its replicate and duplicate assays. The 'COUNTIF' function in Microsoft Office Excel was used to calculate the percentage of positive, negative and weak respiration/reactions in each plate. The negative wells from each assay and the negative assays were excluded.

The percentage of utilisation of substrates was calculated as follows:

$$\text{Percentage of utilisation of substrates} = \left(\frac{\text{Positive/negative/weak reactions}}{\text{Total number of substrates tested across multiple assays}} \right) \times 100$$

This method was used to determine the difference in carbon utilisation within the *E. coli* O157 (SN compared to SP) and O26 (RN compared to RP) isolates.

For the table mentioned above, a cut-off margin was determined for the purpose of comparative analysis with the findings of the confirmatory tests. For an isolate, when the replicate and/or duplicate assays had 50% weak and 50% negative respiration in a substrate, the result was recorded as negative respiration. For those with more than 50% negative respiration across replicate and/or duplicate assays, the result was recorded as negative. Likewise, for isolates with more than 50% weak respiration across replicate and/or duplicate assays, the result was recorded as weak.

3.4 Confirmatory tests for phenotypes

The confirmatory tests were performed by conventional testing methods using MacConkey agar plate and prepared tubed cystine trypticase agar (CTA) medium. Ten out of 17 *E. coli* O157 isolates were chosen to perform the tests. These included the SN isolates EcCa491a, EcHu93a, EcCa461b, EcCa452a and EcHu114a and the SP isolates EcCa26a, EcCa583a, EcCa21a, EcCa519a and EcCa24a.

3.4.1 Cystine trypticase agar medium with sorbitol

Cystine trypticase agar medium with added sorbitol (Fort Richard Laboratories Ltd, Auckland, NZ) was used to determine fermentation reactions. An acidic reaction in the semisolid medium indicates the utilisation of sorbitol and changes the colour of the medium from red to yellow. When sorbitol is not fermented, an alkaline shift causes the medium to appear red (Becton Dickinson and Company, 2012). The CTA medium with sorbitol was inoculated by stabbing the medium with a wire loop containing a fresh colony picked from a pure culture. This was

performed for all the *E. coli* O157 isolates being tested. The tubes were incubated at 37°C with loosened caps.

3.4.2 MacConkey agar with added carbon sources

To validate findings of the PM assay, MacConkey media were prepared with the carbon sources; D-serine, p-hydroxyphenylacetic acid, glycolic acid, glyoxylic acid and propionic acid.

3.4.2.1 MacConkey agar with D-serine or p-hydroxyphenylacetic acid or glycolic acid or glyoxylic acid

The agar was prepared in two ways. For the first mixture, 10 grams (g) of Difco MacConkey agar base (Becton Dickinson and Company (BD), Maryland, USA) and 2.5 g D-serine or p-hydroxyphenylacetic acid or glycolic acid or glyoxylic acid (Sigma-Aldrich Inc, Missouri, USA) were added to 250 ml MilliQ water in a 500 ml Schott bottle. For MacConkey agar with glycolic acid or glyoxylic acid, 6.6 g of Gibco BRL bacteriological agar (Life Technologies Ltd, Scotland) was also added to the mixture. The mixture was boiled in a microwave oven to dissolve the agar. The agar mixture was cooled to approximately 45°C in a water bath. The pH of the mixture was measured using a PHM220 pH meter (Radiometer Analytical SAS, France) and recorded. The medium was autoclaved at 121°C for 15 min and cooled to approximately 45°C in a water bath. A small amount of the medium was poured into a plastic sample cup to remeasure and record the pH. In a laminar flow, the medium was poured immediately into approximately 20 sterile petri dishes.

Isolates were first grown on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ) and were then streaked on the different MacConkey media except for MacConkey medium supplemented with p-hydroxyphenylacetic acid. The inoculated plates were incubated overnight at 37°C. To inoculate MacConkey medium supplemented with p-hydroxyphenylacetic acid, a 2 ml phosphate buffered saline (PBS) (0.01 Molar (M), pH 7.3) was inoculated with a few colonies of an isolate from the growth on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ) to a MacFarland standard of 2.0. From this suspension, 100 µl was pipetted onto the centre of the MacConkey medium with added p-hydroxyphenylacetic acid. The inoculated plates were incubated overnight at 37°C.

For the second mixture, 10 g of Difco MacConkey agar base (BD, Maryland, USA) and 2.5 g D-serine or p-hydroxyphenylacetic acid or glycolic acid or glyoxylic acid (Sigma-Aldrich Inc, Missouri, USA) were added to 250 ml MilliQ water in a 500 ml Schott bottle. The mixture was boiled in a microwave to dissolve the agar. The agar mixture was cooled to approximately 45°C in a water bath. The pH of the mixture was measured and was adjusted to 6.5±0.3 using 10 M sodium hydroxide (NaOH) and corrected using concentrated hydrogen chloride (HCl). The medium was then autoclaved at 121°C for 15 min and cooled back to approximately 45°C in a water bath. A small amount of the medium was poured into a plastic sample cup and the pH was measured to check if it was within the 6.5±0.3 range. In a laminar flow, the medium was poured immediately into approximately 20 sterile petri dishes. Isolates, first grown on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ), were streaked on the different MacConkey media except for MacConkey medium supplemented with glycolic acid (adjusted pH (pH ADJ)). The inoculated plates were incubated overnight at 37°C.

To inoculate MacConkey medium with added glycolic acid (pH ADJ), a 2 ml PBS (0.01 M, pH 7.3) was inoculated with a few colonies of an isolate from the growth on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ) to a MacFarland standard of 2.0. From this suspension, 100 µl was pipetted onto the centre of the MacConkey medium with added glycolic acid (pH ADJ). The inoculated plates were incubated overnight at 37°C.

In all cases, the growth of colourless colonies on MacConkey agar with added carbon sources indicated the growth of the *E. coli* isolates and pink colonies on the medium, sometimes surrounded by a zone of precipitated bile, indicated carbohydrate fermentation (Zimbro *et al.*, 2009).

3.4.2.2 MacConkey agar with propionic acid

For this medium, two mixtures were prepared by adding 10 g of Difco MacConkey agar base (BD, Maryland, USA) and 2.5 g propionic acid (Sigma-Aldrich, Missouri, USA) to 248.8 ml MilliQ water in 500 ml Schott bottles to make a total volume of 250 ml. For the unadjusted pH mixture, 6.6 g of Gibco BRL bacteriological agar (Life Technologies Ltd, Scotland) was also added to the mixture. The medium mixture was subject to the methods (pH unadjusted and pH ADJ) previously described for MacConkey agar with D-serine or p-hydroxyphenylacetic acid or glycolic acid or glyoxylic acid.

3.4.2.3 MacConkey control agar

The control was prepared by adding 10 g of Difco MacConkey agar base (BD, Maryland, USA) to 250 ml MilliQ water in a 500 ml Schott bottle. The medium mixture was subject to the same method described for MacConkey agar with D-serine or p-hydroxyphenylacetic acid or glycolic acid or glyoxylic acid. Only one mixture was prepared and the pH was measured before and after autoclaving, as mentioned earlier, when the mixture was cooled to approximately 45°C. The pH was not adjusted at any point.

3.5 Genomic DNA preparation

Four *E. coli* O157 and six O26 isolates, listed in Table 3.1, were recovered from glycerol broths stored at -80°C and grown on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ) overnight at 37°C. DNA was extracted using the Qiagen DNA Mini and Blood Mini Kit (Bio-Strategy, Auckland, NZ). The manufacturer's instructions for the isolation of genomic DNA from bacterial plate cultures were followed. The elution step was modified to use 200 µl sterile MilliQ water instead of 150 µl Buffer AE according to the manufacturer's instruction. The concentration of the purified genomic DNA for each sample had a concentration of 20 nanograms (ng)/µl or greater. Qubit dsDNA HS/RNA/Protein assay kits (Life Technologies, Auckland, NZ) were used to check the quality and quantity of the DNA. The samples were also run on a 1% agarose, electrophoresis gel to check the quality of the DNA. The DNA samples were then stored at -20°C. The samples along with the Qubit results and the gel photograph were submitted for sequencing to New Zealand Genomics Limited (NZGL), Massey Genome Service, Palmerston North, New Zealand.

3.6 Genome sequencing

The following methods were performed by NZGL, Massey Genome Service, Palmerston North.

3.6.1 Library preparation method

A library was prepared for each of the 10 samples (Table 2.1) using the Illumina TruSeq™ DNA PCR-Free method (Illumina Inc, Australia). In this method, the DNA samples were

randomly fragmented by mechanical shearing. Fragments with an average size of 550 base pairs (bp) were selected by magnetic beads and then barcoded Illumina adapters were ligated onto each end of the fragments. The libraries were pooled by equal molarity before loading onto the Illumina MiSeq analyser.

3.6.2 Illumina MiSeq runs

The sequencing reactions using the 10 libraries were performed on an Illumina MiSeq analyser with the MiSeq Reagent Kit v2 (500 cycle) (Illumina Inc, Australia). The kit manufacturer's instruction using the 2 x 250 base paired end run, v 2 chemistry was followed. For each genome, approximately 1.2 to 1.5 million reads, 0.6 to 0.75 Gb of total data output and approximately 120 to 150 x genomic coverage with a standard *E. coli* genome size of 5.5 megabases (Mb) were generated. Three types of data were obtained automatically at the end of the run which were:

- Unprocessed data: "Raw" sequence files after expansion from the compressed sequence files generated from the MiSeq.
- Processed data: The raw sequences were mapped against the PhiX genome using Burrows-Wheeler Alignment (BWA). The PhiX genome is a high-quality, well-characterised genome used in Illumina sequencing as a quality control (Minoche *et al.*, 2011, Illumina Inc., 2014). Any PhiX reads were removed from the resultant SAM files (Li *et al.*, 2009).
- Processed trimmed data: Each of the sequences in the 'processed' folder were also trimmed to their longest contiguous segment for which quality scores were less than a quality cutoff, set at 0.01, using the software DynamicTrim from the SolexaQA package (Cox *et al.*, 2010) (http://solexaqa.sourceforge.net/).

The data output was provided in fastq format in all cases. Quality control (QC) checks were performed on all the sequence data generated using the following processes:

- FastQC
- FastQscreen
- Solexa QA (Perl-based software package)

3.7 Genome assembly and contig selection

The genome assembly and analysis was performed by Dr Biggs, Massey University. The algorithm package Velvet (v 1.2.10) (Zerbino & Birney, 2008) was used for *de novo* genome assembly. For each genome, the short reads sequences of 250 bp were first broken into small pieces, known as k-mers. A de Bruijn graph was constructed from those short sequences. The sequences were assembled across a range of k-mer lengths in decrements of 10 between 245 and 55, inclusive. Four subsets of random number of sequences were generated with an average base coverage of approximately 150x to obtain 80 contiguous assemblies. The resulting Velvet contiguous sequences (contigs) were stored in a MySQL database (v 5.5). For each genome, the best assembly was chosen using a Perl-based in-house ranking system based on the number of contigs, size of contigs, genome length and N_{50} value. The concatenated contigs from the chosen assembly were annotated using another Perl script incorporating Prokka (v 1.10) (Seemann, 2014).

3.8 Genome clustering and reference genomes

The annotated genomes were clustered in OrthoMCL (v 2.0.9) (Li *et al.*, 2003) and were analysed in-house. Also, a selection of *E. coli* whole genome sequences, available on the National Centre for Biotechnology Information (NCBI) database, was downloaded (Table 3.5). These included 37 *E. coli* O157 genomes and 20 *E. coli* O26 genomes which were used as reference genomes. In addition to this, a set of three *Escherichia fergusonii* (*E. fergusonii*) genomes were also downloaded from NCBI. The reference genomes were compared to the 17 *E. coli* O157 and six *E. coli* O26 New Zealand isolates. These reference genomes were annotated using Prokka as described in section 3.7.

3.8.1 Clustering and core genome tree construction

The pan-genome consists of gene sets present in the genomes of the species under investigation at one time. It comprises of core and accessory genes which contribute towards describing a bacterial species. The genes associated with the biological aspects and phenotypic traits of a species constitute the core genome. The accessory genes are associated with biochemical pathways essential for adaptation, antibiotic resistance and colonisation of a species among others (Medini *et al.*, 2005, Lapierre & Gogarten, 2009).

Core genome comparison was performed on four different combinations of data:

- *E. coli* O157 and O26 New Zealand isolates
- *E. coli* O157 New Zealand isolates and O157 reference isolates
- *E. coli* O26 New Zealand isolates and O26 reference isolates
- *E. coli* O157 and O26 New Zealand isolates and *E. fergusonii* reference isolates

The above combinations were used to create and visualise a NeighborNet tree in SplitsTree (v 4.1.3) (Bryant & Moulton, 2004, Huson, 1998, Huson & Bryant, 2006) where the genes were exactly the same length.

3.9 Phenotypic-genotypic analysis

The whole genome sequences of the *E. coli* O157 and O26 isolates were analysed further to investigate any possible associations with the phenotypic findings from this study.

3.9.1 Clusters of Orthologous Groups of proteins analysis

The database of Clusters of Orthologous Groups of proteins (COGs) was designed to classify proteins encoded in complete genomes for phylogenetic analysis and also to allow for functional annotations of newly sequenced genomes (Tatusov *et al*., 2001). The database currently consists of 4631 COGs and includes proteins from 711 complete genomes. The COG genes were derived as a result of the analysis of 1.875 million genes (Galperin *et al*., 2015).

For this study, the highly curated set of COG genes was downloaded from http://www.ncbi.nlm.nih.gov/COG/, accessed July 2015, which was abbreviated by 26 letters for their functional categories. The classified functions of the COG genes are listed in Table 3.6. The regulators, pathways and proteins associated with carbon utilisation were classified under the following COG categories:

G: Carbohydrate transport and metabolism

H: Coenzyme transport and metabolism

O: Post-translational modification, protein turnover, chaperone functions

Q: Secondary metabolites biosynthesis, transport and catabolism

In order to obtain an output in a short amount of time and provide an insight into the *E. coli* O157 and O26 New Zealand isolates, 12 isolates were chosen for the COG analysis. These included four SN *E. coli* O157 isolates (EcCa63b, EcCa32a, EcHu93a, EcHu114a), three SP *E. coli* O157 isolates (EcCa26a, EcCa24a, EcCa583a), two RN *E. coli* O26 isolates (EcCa8a, EcCa513a) and three RP *E. coli* O26 isolates (EcCa5a, EcCa514a, EcCa569a). Using a BLAST program, each of these *E. coli* genomes were compared to the COG reference set using the COGNITOR software (v 2012.04). A table was generated consisting of COG identifications for each gene in the genome. The rows in the table where there were differences were identified. Using a script written by Dr Marshall, Massey University, a distance matrix was generated in R (v 3.1.2). The matrix was viewed in SplitsTree (v 4.1.3) as a NeighborNet tree.

An output derived from the COGNITOR software, which contained the presence and absence of proteins or protein domains against their respective COG ID and COG category for each of the isolates tested, was transferred to Microsoft Office Excel (Microsoft Corporation). This was performed to explore the number of differences in proteins or protein domains associated with carbon metabolism using the 'STDEV' function between and within the *E. coli* O157 and O26 New Zealand isolates. A standard deviation value of zero showed that either genes were absent or no differences were seen in the number of genes present for the respective protein or protein domain. A standard deviation value greater than zero showed that either there were genes present or there were differences in the number of genes present for the respective protein or protein domain. The 'STDEV' function was carried out for *E. coli* O157 isolates and *E. coli* O26 isolates separately until the 'no differences/ STDEV = 0' in protein or protein domains were removed and only the 'differences/STDEV > 0' were remaining. The proteins or protein domains belonging to COG categories G, H, O and Q (associated with carbon metabolism) were then searched for in both *E. coli* O157 and O26 isolates.

Table 3.5. Summary of genome characteristics of *E. coli* O157, *E. coli* O26 and *E. fergusonii* reference isolates obtained from NCBI database.

Reference isolate names	NCBI Accession numbers	Organism	Serotype	Genome length*	Number of contigs	GC content	rRNA	Repeat region	Signal peptide	CDS	tRNA	tmRNA
DEC3A	NZ_AIFE00000000	*E. coli*	O157	5445263	3	50.22	34	1	464	5244	114	1
DEC3B	NZ_AIFF00000000	*E. coli*	O157	5502782	1	50.22	32	1	464	5341	104	1
DEC3C	NZ_AIFG00000000	*E. coli*	O157	5481558	1	50.22	38	1	452	5365	114	1
DEC3D	NZ_AIFH00000000	*E. coli*	O157	5437121	1	50.22	33	1	456	5278	110	1
DEC3E	NZ_AIFI00000000	*E. coli*	O157	5524029	2	50.22	35	1	460	5391	106	1
DEC4A	NZ_AIFK00000000	*E. coli*	O157	5378775	2	50.22	38	1	440	5231	103	1
DEC4B	NZ_AIFL00000000	*E. coli*	O157	5634084	1	50.22	33	1	468	5565	112	1
DEC4C	NZ_AIFM00000000	*E. coli*	O157	5534167	1	50.22	42	0	459	5410	106	1
DEC4D	NZ_AIFN00000000	*E. coli*	O157	5388008	2	50.22	34	1	452	5216	108	1
DEC4E	NZ_AIFO00000000	*E. coli*	O157	5410763	2	50.22	35	1	453	5215	97	1
Ec06_4039	NZ_JHNG00000000	*E. coli*	O157	5266208	3	50.22	4	1	454	5004	83	1
Ec08_4540	NZ_JHHH00000000	*E. coli*	O157	5362509	8	45.56	4	0	457	5147	86	1
Ec2009C_4258	NZ_JHGQ00000000	*E. coli*	O157	5353946	7	45.56	4	1	457	5138	81	1
Ec2009EL2109	NZ_JHGC00000000	*E. coli*	O157	5384883	1	45.64	8	1	467	5184	90	1
Ec2010EL_2045	NZ_JHMD00000000	*E. coli*	O157	5319432	4	47.64	6	1	460	5091	90	1
Ec2011EL_1107	NZ_JHLK00000000	*E. coli*	O157	5414139	5	47.64	5	1	474	5240	93	2
Ec2011EL_2313	NZ_JHKA00000000	*E. coli*	O157	5340384	3	47.64	3	1	458	5109	86	1
F6749	NZ_JHJQ00000000	*E. coli*	O157	5330186	3	47.64	4	1	455	5070	84	1
K6687	NZ_JHHV00000000	*E. coli*	O157	5570585	5	43.86	13	1	478	5499	97	1
EC4042	NZ_ABHM00000000	*E. coli*	O157:H7	5617728	1	50.46	22	1	473	5473	108	1
EC4045	NZ_ABHL00000000	*E. coli*	O157:H7	5634850	1	50.47	22	1	474	5453	108	1
EC4076	NZ_ABHQ00000000	*E. coli*	O157:H7	5705645	1	50.56	27	1	478	5558	110	1
EC4113	NZ_ABHP00000000	*E. coli*	O157:H7	5655847	1	50.6	20	1	464	5603	89	1
EC4115	NC_011353	*E. coli*	O157:H7	5572075	1	50.52	22	1	473	5401	108	1
EC4196	NZ_ABHO00000000	*E. coli*	O157:H7	5620606	1	50.55	24	1	468	5492	96	1
EC4206	NZ_ABHK00000000	*E. coli*	O157:H7	5629932	1	50.47	22	1	469	5537	107	1
EC4401	NZ_ABHR00000000	*E. coli*	O157:H7	5733133	1	50.46	23	1	485	5630	97	1
EC4486	NZ_ABHS00000000	*E. coli*	O157:H7	5933166	1	50.42	23	1	502	5818	106	1
EC4501	NZ_ABHT00000000	*E. coli*	O157:H7	5677181	1	50.58	19	1	469	5625	77	1
EC508	NZ_ABHW00000000	*E. coli*	O157:H7	5656666	1	50.65	21	1	470	5535	83	1

bp: Base pairs, GC content: Guanine-cytosine content in percentage, rRNA: Mature ribosomal RNA, CDS: Coding sequence, tRNA: Mature transfer RNA, tmRNA: Transfer messenger RNA, *: Genome length according to NCBI database

Table 3.5 continued. Summary of genome characteristics of *E. coli* O157, *E. coli* O26 and *E. fergusonii* reference isolates obtained from NCBI database.

Reference isolate names	NCBI Accession numbers	Organism	Serotype	Genome length*	Number of contigs	GC content	rRNA	Repeat region	Signal peptide	CDS	tRNA	tmRNA
EC869	NZ_ABHU00000000	*E. coli*	O157:H7	5731065	1	50.54	23	1	465	5588	95	1
EDL933	AE005174	*E. coli*	O157:H7	5528445	1	50.38	22	1	468	5321	102	1
FRIK2000	NZ_ACXO00000000	*E. coli*	O157:H7	5408690	1	50.18	4	1	464	5253	79	1
FRIK966	NZ_ACXN00000000	*E. coli*	O157:H7	5376914	1	50.25	3	1	454	5236	79	1
Sakai	NC_002695	*E. coli*	O157:H7	5498450	1	50.54	22	1	470	5295	105	1
TW14359	NC_013008	*E. coli*	O157:H7	5528136	1	50.51	22	1	472	5344	108	1
TW14588	NZ_ABKY00000000	*E. coli*	O157:H7	5670297	1	50.51	22	1	475	5416	103	1
DEC9A	NZ_AIGK00000000	*E. coli*	O26	5408446	2	50.22	42	2	445	5230	112	1
DEC9B	NZ_AIGL00000000	*E. coli*	O26	5361604	1	50.22	36	2	442	5169	104	1
DEC9C	NZ_AIGM00000000	*E. coli*	O26	5194722	2	50.22	34	2	433	5000	97	1
DEC9D	NZ_AIGN00000000	*E. coli*	O26	5485621	1	50.22	38	2	458	5307	104	1
DEC9E	NZ_AIGO00000000	*E. coli*	O26	5430771	3	50.22	29	2	449	5319	103	1
DEC10A	NZ_AIGP00000000	*E. coli*	O26	5368171	6	50.22	34	2	441	5360	102	1
DEC10B	NZ_AIGQ00000000	*E. coli*	O26	5578446	1	50.22	36	2	445	5591	104	1
DEC10C	NZ_AIGR00000000	*E. coli*	O26	5531102	4	50.22	35	2	438	5552	108	1
DEC10D	NZ_AIGS00000000	*E. coli*	O26	5404073	1	50.22	37	2	432	5405	108	1
Ec03_3500	NZ_JHNT00000000	*E. coli*	O26	5343090	1	50.22	3	2	440	5205	89	1
Ec05_3646	NZ_JHOE00000000	*E. coli*	O26	5446986	2	50.22	5	2	441	5348	90	1
Ec06_3464	NZ_JHNO00000000	*E. coli*	O26	5336400	7	50.22	3	2	442	5185	85	1
Ec2009C_3612	NZ_JHGZ00000000	*E. coli*	O26	5222538	7	45.56	2	2	442	5045	82	1
Ec2009C_4826	NZ_JHGI00000000	*E. coli*	O26	5359440	8	45.64	3	2	444	5243	83	1
Ec2010C_3051	NZ_JHGA00000000	*E. coli*	O26	5354218	5	45.64	2	2	441	5235	84	1
Ec2010C_4347	NZ_JHFB00000000	*E. coli*	O26	5298885	7	47.64	3	2	440	5164	82	1
Ec2010C_4819	NZ_JHMP00000000	*E. coli*	O26	5401165	5	47.64	2	2	451	5280	81	1
Ec2011C_3270	NZ_JHLY00000000	*E. coli*	O26	5453558	9	47.64	3	2	438	5349	93	1
Ec2011C_3655	NZ_JHLN00000000	*E. coli*	O26	5297926	10	47.64	3	2	441	5160	91	1
11368	AP010953	*E. coli*	O26:H11	5697240	1	50.68	22	2	458	5578	105	1
B253	NZ_AEIA00000000	*E. fergusonii*	-	4768896	20	49.71	11	2	396	4499	88	1
ATCC 35469	NC_011740	*E. fergusonii*	-	4588711	2	49.92	22	2	380	4303	89	1
ECD227	NZ_AEVY00000000	*E. fergusonii*	-	4857906	6	49.88	3	3	411	4589	62	1

bp: Base pairs, GC content: Guanine-cytosine content in percentage, rRNA: Mature ribosomal RNA, CDS: Coding sequence, tRNA: Mature transfer RNA, tmRNA: Transfer messenger RNA, *: Genome length according to NCBI database

Table 3.6. Classified functions of COGs. Source: http://www.ncbi.nlm.nih.gov/COG/, accessed July, 2015

COGs	Functions
A	RNA processing and modification
B	Chromatin structure and dynamics
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/envelop biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, chaperone functions
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism
R	General functional prediction only
S	Function unknown
T	Signal transduction mechanisms
U	Intracellular trafficking, secretion, and vesicular transport
V	Defense mechanisms
W	Extracellular structures
X	Mobilome: prophages, transposons
Y	Nuclear structure
Z	Cytoskeleton

COGs: Clusters of Orthologous Groups of proteins

FOUR

Results

4.1 Bacterial isolates

The *E. coli* O157 and O26 isolates, after overnight incubation at 37°C, grew smooth, round and grey colonies on Columbia horse blood agar (Fort Richard Laboratories Ltd, Auckland, NZ).

4.2 Shiga toxin-encoding bacteriophage insertion typing

All 17 *E. coli* O157 isolates were classified by their Shiga toxin-encoding bacteriophage insertion sites. The SBI types determined for these isolates are shown in Table 4.1. The presence of at least two products at any of the bacteriophage insertion sites except for the variant *yehV* junction determined the naming of the insertion types. Twelve sorbitol negative (SN) *E. coli* O157 isolates (EcCa452a, EcCa63b, EcCa491a, EcCa32a, EcCa461b, EcCa635a, EcHu44a, EcHu93a, EcHu56a, EcHu76a, EcHu68a and EcHu114a) were found to contain at least two products in more than one insertion loci (*argW*, *sbcB*, *wrbA* and *yehV*) and also included at least one of the three *stx* genes. The remaining sorbitol positive (SP) *E. coli* O157 isolates (EcCa26a, EcCa21a, EcCa24a, EcCa519a and EcCa583a) contained fewer than two products at any of the bacteriophage insertion sites.

In the case of the *E. coli* O26 isolates, rhamnose negative (RN) isolates EcCa8a and EcCa1a were found to contain two products at junction *sbcB* and also possessed the *stx1* gene. In isolate EcCa514a, which is rhamnose positive (RP), two products were present at junctions *wrbA* and *yehV*. The isolate, however, did not include any of the *stx* genes. The RN isolate EcCa513a and RP isolates EcCa5a and EcCa569a, contained fewer than two products at any of the bacteriophage insertion sites.

Table 4.1. SBI types and sorbitol and rhamnose characterisation of *E. coli* O157 and O26 New Zealand isolates.

Isolate name	Serotype	Sorbitol/ Rhamnose characterisation	SBI type
EcCa452a	O157	SN	SY2c
EcCa63b	O157	SN	AY2a
EcCa491a	O157	SN	AY2a
EcCa32a	O157	SN	WY12a
EcCa461b	O157	SN	WY12a
EcCa635a	O157	SN	ASY2c
EcCa26a	O157	SP	NIL
EcCa21a	O157	SP	NIL
EcCa24a	O157	SP	NIL
EcCa519a	O157	SP	NIL
EcCa583a	O157	SP	NIL
EcHu44a	O157	SN	AY2a
EcHu93a	O157	SN	Y2c
EcHu56a	O157	SN	Y1
EcHu76a	O157	SN	WY12a
EcHu68a	O157	SN	SY2c
EcHu114a	O157	SN	ASY2c
EcCa8a	O26	RN	S1
EcCa513a	O26	RN	NIL
EcCa1a	O26	RN	S1
EcCa5a	O26	RP	NIL
EcCa514a	O26	RP	WY
EcCa569a	O26	RP	NIL

SBI: Shiga toxin-encoding bacteriophage insertion, NIL: No SBI classification, SN: Sorbitol negative, SP: Sorbitol positive, RN: Rhamnose negative, RP: Rhamnose positive

4.3 Phenotypic microarray technology

Four *E. coli* O157 and six *E. coli* O26 isolates were examined for their ability to utilise carbon sources by phenotypic microarray technology using OmniLog® (Biolog Inc., Hayward, USA) (Bochner *et al.*, 2001). The tetrazolium dye colour changed from colourless to purple ('positive' reaction) after 24 h of incubation where the substrate had been utilised. An example of a PM assay after incubation is shown in Figure 4.1. The figure showed a light purple coloured well or 'weak reaction' where poor utilisation of the substrate occurred, and a colourless well or 'negative' reaction where the substrate was not utilised.

The coloured heatmaps generated from PM analysis using software R (v 3.1.2) (Ihaka & Gentleman, 1996) provided a better way to visualise and analyse the data obtained from the PM assays. The differences between the *E. coli* O157 isolates were explored by combining the PM data obtained from the four *E. coli* O157 isolates with those from the previously tested 13 *E. coli* O157 isolates. The heatmap demonstrated the variation as well as the reproducibility of the replicated experiments.

4.3.1 Carbon utilisation profiles of *E. coli* O157 isolates

The PM plates, PM1 and PM2A, were used to study the carbon source utilisation profiles of *E. coli* O157 isolates. Together, the utilisation of 190 carbon sources was tested. Following the completion of PM1 and PM2A runs, the analysis showed that fewer carbon sources in PM2A had been utilised compared to PM1. The negative assays and their replicates for both PM1 and PM2A did not result in any respiration. The negative assays demonstrated that in the absence of a bacterial isolate, the inoculating fluid or other external causes did not cause respiration in the wells.

A heatmap depicting the utilisation of carbon sources by the *E. coli* O157 isolates is shown in Figure 4.2. A more comprehensive heatmap including the duplicate and replicate assays for each *E. coli* O157 isolate is shown in Figure A1. α-D-glucose-1-phosphate (well E03), D-glucuronic acid (B05), D-gluconic acid (B06), D-fructose-6-phosphate (E04), D-glucose-6-phosphate (C01), glycerol (B03), L-lactic acid (B09) and 2'-deoxyadenosine (E11) were utilised by all the isolates. After 24 h of incubation, isolates EcCa452a, EcCa63b, EcCa491a, EcCa32a, EcCa461b, EcCa635a, EcHu44a, EcHu93a, EcHu56a, EcHu76a, EcHu68a and

EcHu114a did not respire on D-sorbitol (B02) compared to isolates EcCa26a, EcCa21a, EcCa24a, EcCa519a and EcCa583a which respired on D-sorbitol.

Isolates EcHu56a, EcHu93a and EcHu68a utilised additional substrates compared to all other isolates. These isolates respired on L-threonine (G04), L-glutamine (E01) and lactulose (D10). Also, isolate EcCa635a showed a different phenotypic pattern compared to the other *E. coli* O157 isolates. It respired on fewer substrates in the PM1 assay, 13 positive reactions out of 95 substrates.

All the isolates that did not possess a SBI insertion and respired on sorbitol (labelled pink in Figure 4.2) showed higher utilisation of carbon substrates in PM1 compared to other *E. coli* O157 isolates of different SBI types. Isolates EcCa26a, EcCa583a, EcCa21a, EcCa519a and EcCa24a respired on D-serine, glycolic acid, glyoxylic acid and propionic acid. These isolates, except for EcCa21a, respired on p-hydroxyphenylacetic acid. The isolates that possessed a SBI type did not respire on D-serine, glycolic acid and p-hydroxyphenylacetic acid but respired weakly on glyoxylic acid and propionic acid.

The *E. coli* O157 isolates displayed fewer positive reactions in PM2A as compared to PM1 (Figure A2). The carbon sources in PM2A that were utilised by all the isolates were N-acetyl-neuraminic acid (B02) and D-glucosamine (E05). Of the isolates, EcCa26a utilised the most number of substrates (11 positive reactions out of 95 substrates).

As shown in Tables 4.2 and 4.3, across both PM1 and PM2A plates, the positive wells indicated that SN isolates utilised 32.5% (1205 positives out of 3705 substrates) and 7.8% (274/3515) respectively, compared to SP isolates which utilised 45.2% (601/1330) and 7.9% (112/1425) respectively.

In addition, human isolates were calculated to have respired on 37.5% (641/1710) of the substrates in PM1 compared to 28.3% (564/1995) by the animal isolates.

4.3.2 Carbon utilisation profiles of *E. coli* O26 isolates

Phenotypic microarray technology was also used to investigate the utilisation of carbon sources by six *E. coli* O26 isolates. Again, 190 carbon sources were tested when PM1 and PM2A plates were used to create a carbon profile for the *E. coli* O26 isolates.

The utilisation of carbon sources in PM1 by the *E. coli* O26 isolates is shown in Figure 4.3. In PM1, substrates D, L-α-glycerol-phosphate (B07), L-galatonic acid-g-lactone (H09), D-galacturonic acid (H10), α-D-glucose-1-phosphate (E03), D-glucuronic acid (B05), D-gluconic acid (B06), D-fructose (C07), D-glucose (C09), pyruvic acid (H08), thymidine (C12), methylpyruvate (G10), inosine (F12), D-fructose-6-phosphate (E04), D-glucose-6-phosphate (C01), L-lactic acid (B09), L-serine (G03) and 2'-deoxyadenosine (E11) were utilised by all the isolates. All the isolates did not respire on D-serine (B01). Isolates EcCa1a, EcCa513a and EcCa8a did not respire on L-rhamnose (C06) compared to isolates EcCa5a, EcCa514a and EcCa569a which respired on L-rhamnose. After 24 h of incubation, isolate EcCa8a respired on fewer substrates compared to the other *E. coli* O26 isolates (38 positive reactions out of 95 substrates). Isolates EcCa569a and EcCa5a respired on D-alanine (A09) and dulcitol (A12) whereas the other isolates did not. Isolates EcCa1a, EcCa513a, EcCa514a and EcCa8a utilised L-fucose (B04) whereas the other isolates did not.

The *E. coli* O26 isolates utilised fewer carbon substrates in PM2A as compared to PM1 (Figure A3). All the isolates respired on D-raffinose (D01), melibionic acid (F03), dextrin (A06) and D-lactic acid methyl ester (F01). Isolates EcCa513a, EcCa1a, EcCa8a and EcCa514a did not utilise D-tagatose (D06) whereas isolates EcCa5a and EcCa569a utilised D-tagatose.

Tables 4.4 and 4.5 show that, for both PM1 and PM2A plates, the RP isolates utilised 41% (351 positives out of 855 substrates) and 7.1% (61/855) respectively, compared to RN isolates which utilised 40% (342/855) and 5.7% (54/950) respectively.

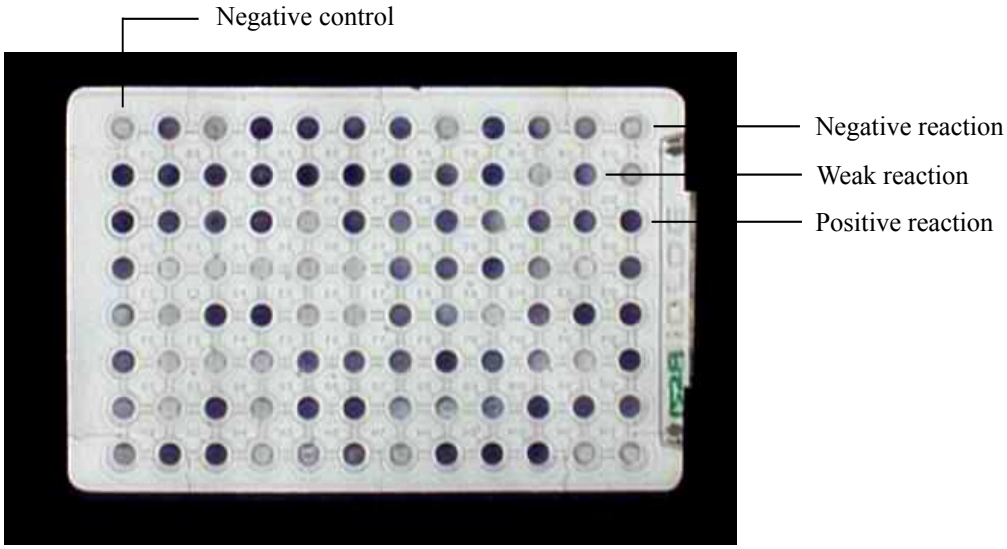

Figure 4.1. Isolate EcCa519a in PM1. This image is of a PM1 plate inoculated with *E. coli* isolate EcCa519a. The image was captured at the end of a 24 hour incubation by a monochrome CCD camera in the OmniLog® incubator. The well on the top left on the plate is the negative control. The dark purple indicates a 'positive reaction', light purple indicates a 'weak reaction' and a colourless well indicates a 'negative reaction'.

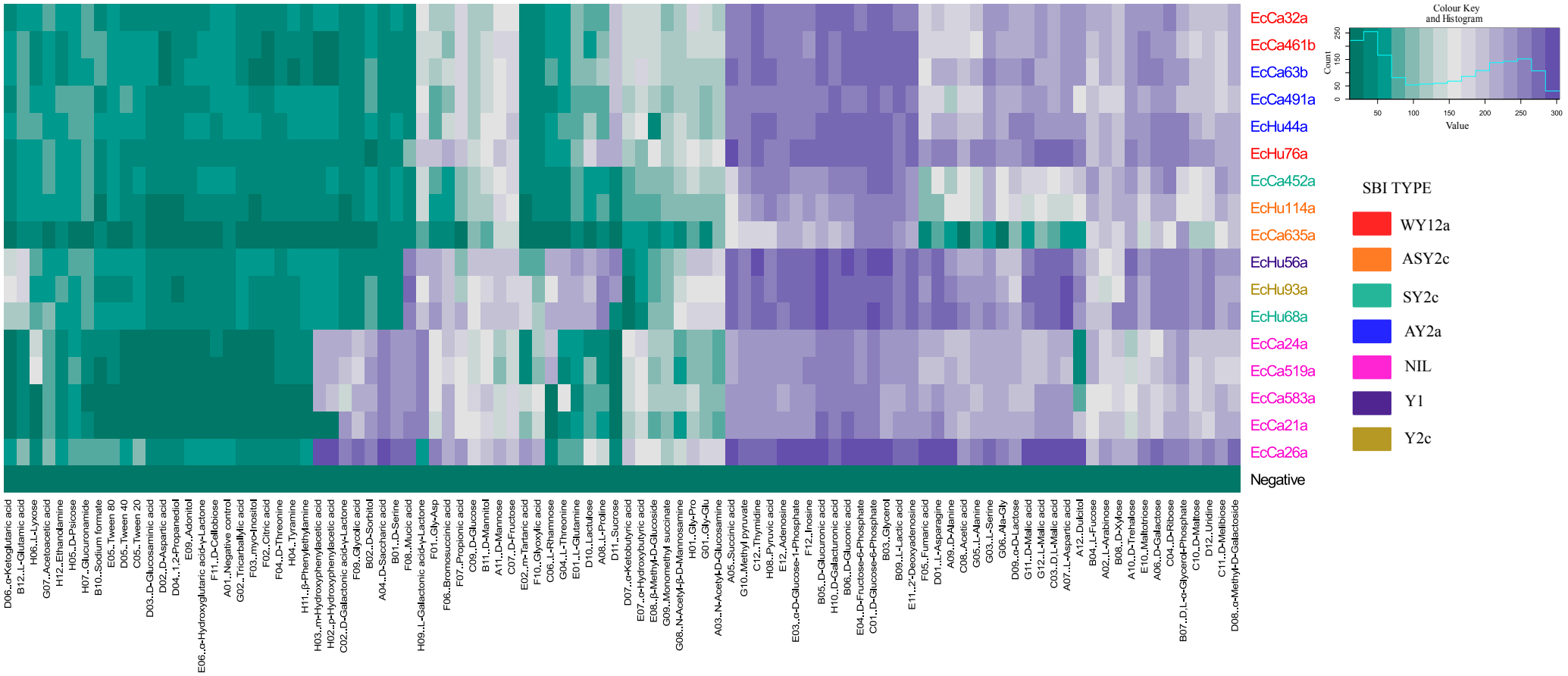

Figure 4.2. Heatmap of *E. coli* O157 respiration changes on PM1. The heatmap illustrates the cell respiration of 96 phenotypes by 17 *E. coli* O157 isolates. The purple areas indicate better respiration, the grey areas indicate minimal respiration and the green areas indicate no respiration of the isolates. The column on the right represents the isolates tested for respiration and the row at the bottom represents the 96 wells of PM1 (A01 to H12). The graph on the right includes the colour key and histogram of the heatmap. The key on the right indicates the SBI types of the *E. coli* O157 isolates. This heatmap does not include the replicate and duplicate assays.

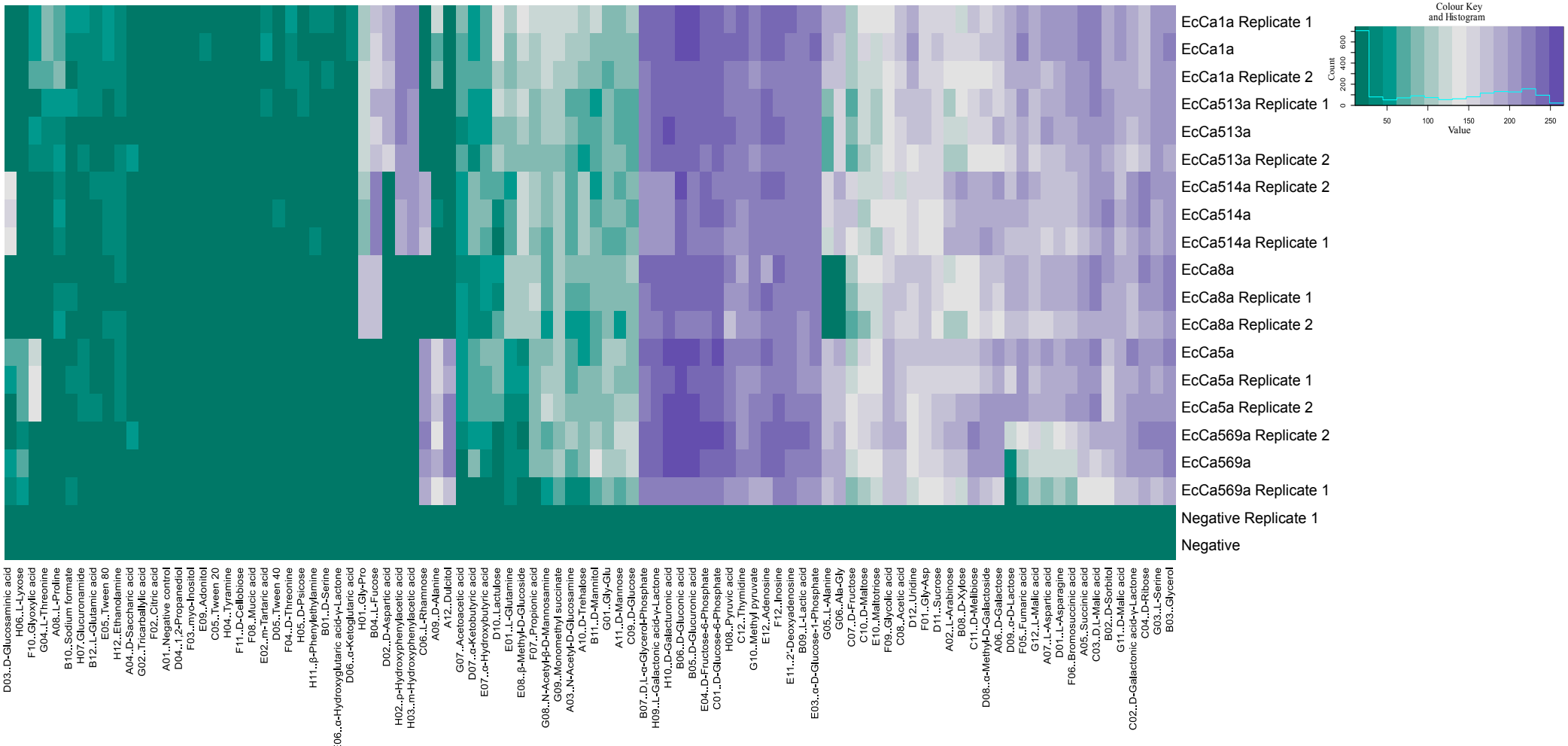

Figure 4.3. Heatmap of *E. coli* O26 respiration changes on PM1. The heatmap illustrates the cell respiration of 96 phenotypes by six *E. coli* O26 isolates. The purple areas indicate better respiration, the grey areas indicate minimal respiration and the green areas indicate no respiration of the isolates. The column on the right represents the isolates tested for respiration and the row at the bottom represents the 96 wells of PM1 (A01 to H12). The graph on the right includes the colour key and histogram of the heatmap.

Table 4.2. Percentage of substrate utilisation by sorbitol negative *E. coli* O157 isolates in PM1 and PM2A.

PM plate	Positive reactions	Negative reactions	Weak reactions	Total no. of substrates tested across multiple assays
PM1	1205	1618	882	3705
Percentage	32.5	43.7	23.8	
PM2A	274	2369	872	3515
Percentage	7.8	67.4	24.8	

Table 4.3. Percentage of substrate utilisation by sorbitol positive *E. coli* O157 isolates in PM1 and PM2A.

PM plate	Positive reactions	Negative reactions	Weak reactions	Total no. of substrates tested across multiple assays
PM1	601	423	306	1330
Percentage	45.2	31.8	23	
PM2A	112	1111	202	1425
Percentage	7.9	78	14.2	

Table 4.4. Percentage of substrate utilisation by rhamnose negative *E. coli* O26 isolates in PM1 and PM2A.

PM plate	Positive reactions	Negative reactions	Weak reactions	Total no. of substrates tested across multiple tests
PM1	342	309	204	855
Percentage	40	36.2	23.9	
PM2A	54	819	77	950
Percentage	5.7	86.2	8.1	

Table 4.5. Percentage of substrate utilisation by rhamnose positive *E. coli* O26 isolates in PM1 and PM2A.

PM plate	Positive reactions	Negative reactions	Weak reactions	Total no. of substrates tested across multiple assays
PM1	351	320	184	855
Percentage	41	37.4	21.5	
PM2A	61	721	73	855
Percentage	7.1	84.3	8.4	

Percentage of utilisation of substrates = (Positive/negative/weak reactions ÷ Total number of substrates tested across multiple assays) × 100, PM: Phenotypic microarray

4.3.3 Reproducibility of phenotypic microarray tests

Phenotypic microarray tests were performed in replicates for both PM1 and PM2A plates for all *E. coli* O157 and O26 New Zealand isolates. Duplicate assays were also performed for both plates for a few *E. coli* O157 isolates. The heatmaps obtained from both the PM1 and PM2A assays showed that there was general concordance between the duplicate assays and between assays repeated on different days (replicates), although these were not obtained for all tests. The reproducibility did vary between some wells/substrates between the replicates and duplicates. Figure 4.4 shows the respiration curves of *E. coli* O157 isolate EcCa519a and its replicates generated as an x-y plot, during a 24 h incubation period. The respiration curves of the replicates overlap each other for various substrates demonstrating that the PM tests conducted in the current study are reproducible. However, in the case of a few substrates such as L-fucose (B04), L-lactic acid (B09) and glucuronamide (H07), variations in the respiration curves are visible among the three assays.

4.4 Confirmatory tests for phenotypes

To verify the findings of the phenotypic microarray analysis, tests were performed using 10 of the 17 *E. coli* O157 isolates. SN isolates EcCa491a, EcHu93a, EcCa461b, EcCa452a, EcHu114a and SP isolates EcCa26a, EcCa583a, EcCa21a, EcCa519a, EcCa24a were selected to perform the confirmatory tests.

4.4.1 Cystine trypticase agar medium with sorbitol

Cystine trypticase agar medium is recommended for the differentiation of microorganisms by means of fermentation reactions. CTA medium with added sorbitol was used to test the *E. coli* O157 isolates for the fermentation of sorbitol. The uninoculated medium, which was the control for the experiment, is red in colour. The isolates EcCa491a, EcHu93a, EcCa461b, EcCa452a and EcHu114a did not ferment sorbitol. An alkaline shift caused the medium to remain red. The isolates EcCa26a, EcCa583a, EcCa21a, EcCa519a and EcCa24a fermented sorbitol turning the medium from red to yellow as shown in Figure 4.5. The results, summarised in Table 4.6, showed that 100% (all 10) of the isolates tested for fermentation reactions of the CTA medium with sorbitol were in agreement with the results of the PM1 assay.

4.4.2 MacConkey agar with added carbon sources

Ten types of MacConkey media were prepared for the confirmatory tests, including the plain MacConkey medium. Four out of 10 media set properly and the isolates were cultured on these media by streaking. These included MacConkey agar (control), MacConkey agar with propionic acid (pH ADJ), MacConkey agar with p-hydroxyphenylacetic acid (pH ADJ) and MacConkey agar with D-serine. Although the MacConkey agar with glycolic acid (pH ADJ) and MacConkey with p-hydroxyphenylacetic acid media set, they were semi-solid and difficult to streak on. Here, bacterial suspensions of the isolates were pipetted onto the media and incubated. The addition of Gibco BRL bacteriological agar (Life Technologies Ltd, Scotland) to MacConkey agar with glycolic acid or glyoxylic acid or propionic acid did not assist in setting the media. The three media mixtures had pH values recorded as less than four before and after autoclaving.

The pH value of the MacConkey control agar after autoclaving was recorded at 7.0 and the pH values of the MacConkey agar with added carbon sources were at 6.5±0.3. A summary of the findings of the pH values and MacConkey agar with added carbon sources compared to their respective PM1 results are listed in Table 4.6. Also, the findings of the SN isolates EcCa461b, EcHu114a and SP isolates EcCa26a, EcCa519a are shown as examples of the confirmatory tests in Figure 4.6.

All 10 *E. coli* O157 isolates grew on MacConkey control agar and produced colourless colonies. The colour of the medium changed from dark pink to yellow. On MacConkey agar with added D-serine, the SN isolates EcCa491a, EcHu93a, EcCa461b, EcCa452a and EcHu114a produced light growth of pink colonies. The pink colonies indicated that D-serine was fermented. The colour of the medium remained unchanged and no precipitation of bile was observed. However, the SP isolates EcCa26a, EcCa583a, EcCa21a, EcCa519a and EcCa24a produced heavy growth of colourless colonies. All 10 isolates, 100%, tested on MacConkey agar with added D-serine were not in agreement with the results of the PM1 assay (Table 4.6). The PM1 assay showed that the SN isolates had not respired on D-serine while SP isolates were shown to have respired on the same.

On MacConkey with added glycolic acid (pH ADJ), the 100µl of bacterial suspension of the isolates produced moderate colourless growth. This was observed for all 10 isolates. Fifty

percent of the isolates (5 SN) tested on MacConkey agar with added glycolic acid (pH ADJ) were in agreement with the results of the PM1 assay while the remaining 50% of the isolates (5 SP) were not in agreement with the PM1 results (Table 4.6). The PM1 assay showed that the SN isolates had not respired on glycolic acid. However, the PM1 assay also showed that SP isolates respired on glycolic acid except for EcCa519a which demonstrated weak respiration.

On MacConkey agar with added p-hydroxyphenylacetic acid, no growth was observed for all the isolates where the 100µl of bacterial suspension were pipetted onto the medium. However, all 10 isolates grew on MacConkey agar with added p-hydroxylphenylacetic acid (pH ADJ) producing moderate to heavy growth. It was also observed that, for isolates EcCa26a and EcCa583a, the media turned brownish black in colour where there was growth. Six out of 10 isolates, 60%, tested on MacConkey agar with added p-hydroxyphenylacetic acid (pH ADJ) confirmed the results of the PM1 assay whereas 40% of isolates failed to confirm the PM1 results. SN isolates EcCa491a, EcHu93a, EcCa461b, EcCa452a, EcHu114a and SP isolate EcCa519a did not respire on p-hydroxyphenylacetic acid while SP isolates EcCa26a, EcCa583a, EcCa21a and EcCa24a respired on p-hydroxyphenylacetic acid in the PM1 assay.

Lastly, all isolates except EcCa583a produced colourless colonies with growth varying from light to heavy on MacConkey agar with added propionic acid (pH ADJ). Isolate EcCa583a did not grow on MacConkey agar with added propionic acid (pH ADJ). Three out of 10 isolates, 30%, tested on MacConkey agar with added propionic acid (pH ADJ) confirmed the PM1 results while the remaining 70% of the isolates were not in agreement with the findings of the PM1 assay. SN isolates EcCa491a, EcCa461b and SP isolates EcCa26a, EcCa583a, EcCa21a, EcCa519a and EcCa24a respired on propionic acid while SN isolates EcHu93a, EcCa452a and EcHu114a did not respire on propionic acid in the PM1 assay.

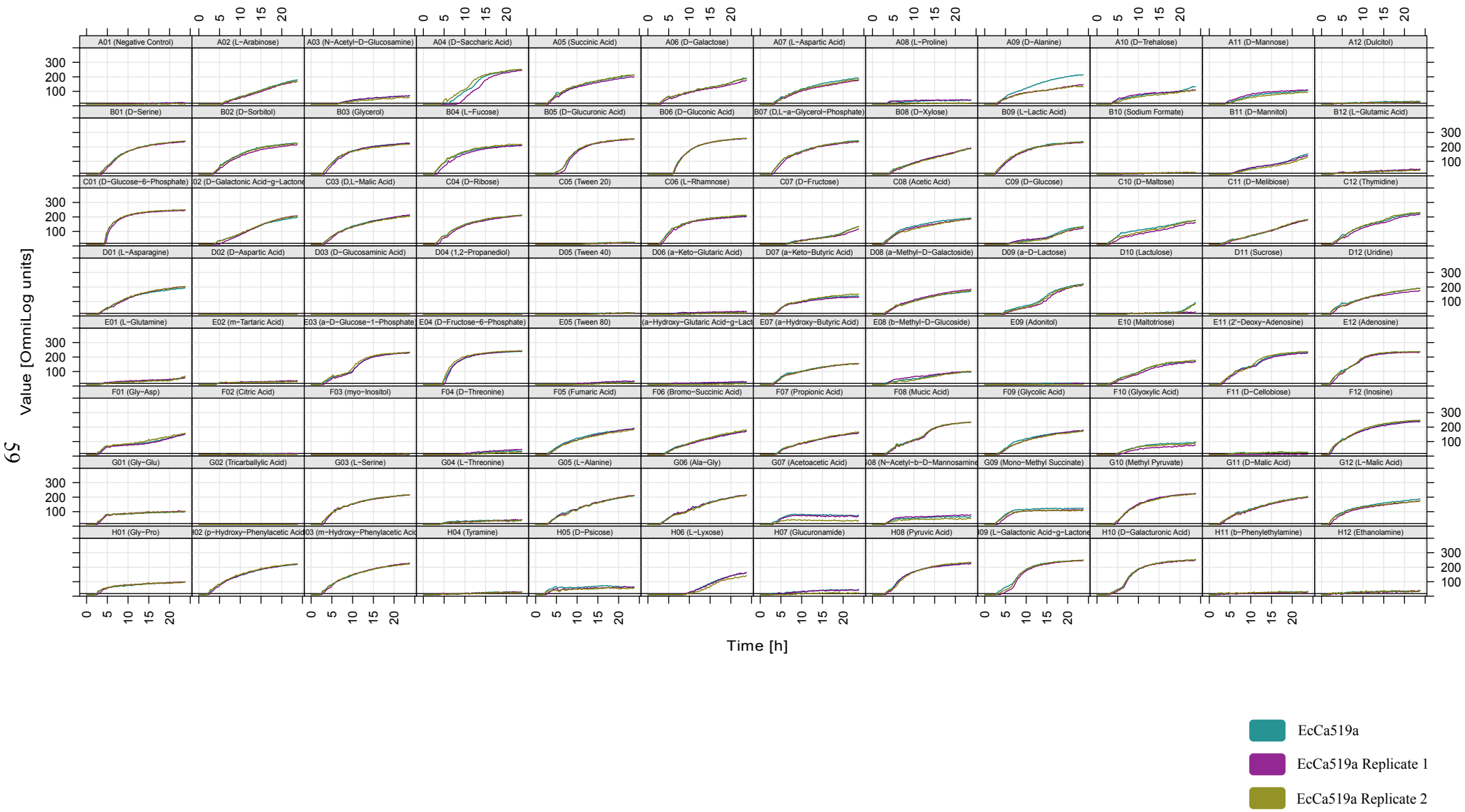

Figure 4.4. Reproducibility of phenotype microarray tests. The x-y plot depicts the cell respiration rates of *E. coli* O157 isolate EcCa519a and its replicates in PM1 during a 24 hour incubation period. The data obtained from the OmniLog-PM software was used in R (v 3.1.2) to generate the x-y plot. The x axis indicates the time from 0 to 24 hours for each well and the y axis indicates the value in OmniLog units of the respiration curve for each substrate.

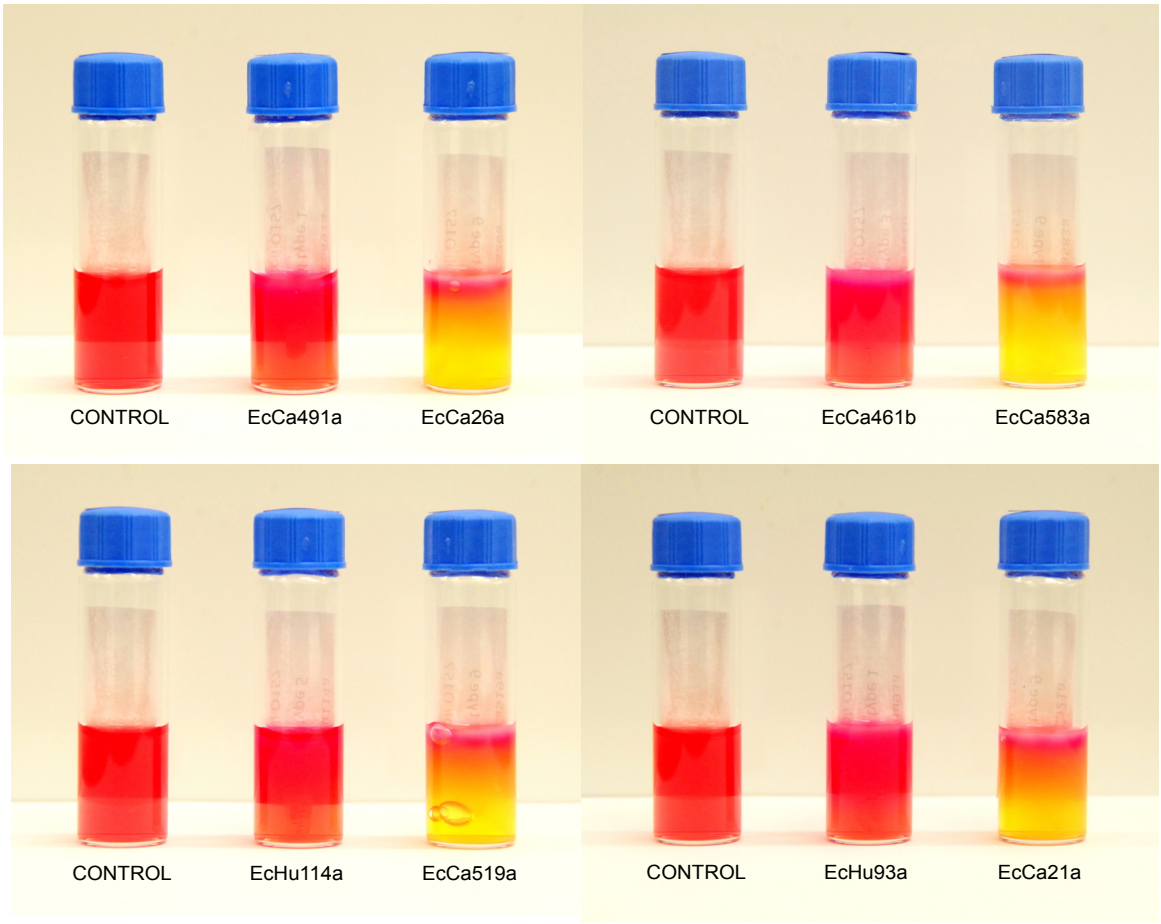

Figure 4.5. *E. coli* O157 New Zealand isolates in CTA medium with sorbitol. The uninoculated tube (control) is red in colour. When the medium is fermented, the colour changes from red to yellow and if no fermentable carbohydrate is present, the medium remains red.

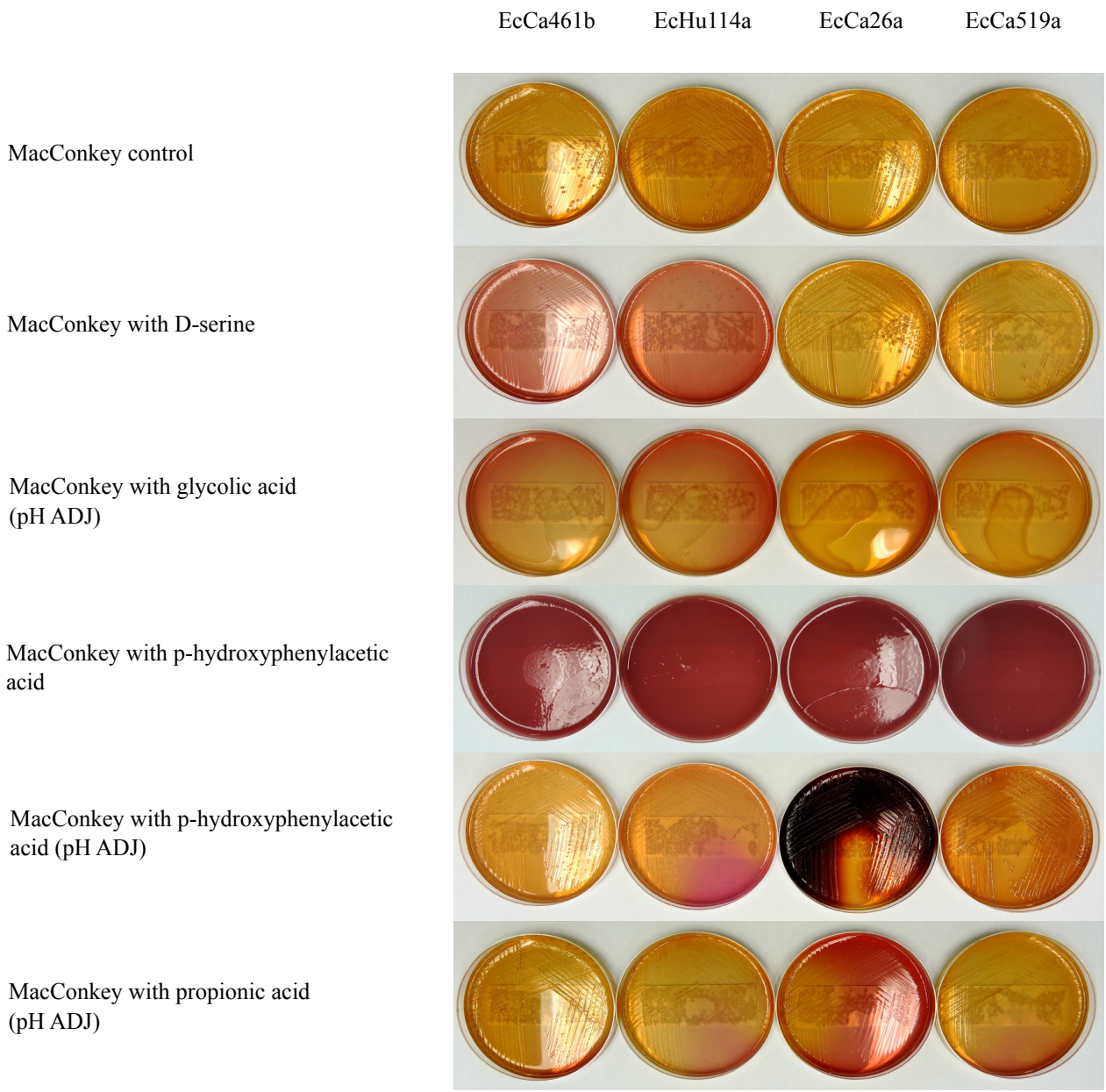

Figure 4.6 *E. coli* O157 New Zealand isolates on MacConkey agar with added carbon sources.
The *E. coli* O157 isolates produced colourless colonies on MacConkey control agar. Isolates EcCa461b and EcHu114a grew few pink colonies while isolates EcCa26a and EcCa519a grew colourless colonies on MacConkey agar with added D-serine. All the isolates did not grow on MacConkey agar with added p-hydroxyphenylacetic acid. MacConkey agar with added glycolic acid, p-hydroxyphenylacetic acid and propionic acid (all pH ADJ) produced colourless colonies.
pH ADJ: pH adjusted to 6.5±0.3

Table 4.6. Summary of the confirmatory test results and their respective PM1 results.

Isolate name	PM sorbitol	CTA medium with sorbitol	MacConkey control	PM D-serine	MacConkey with D-serine	PM glycolic acid	MacConkey with glycolic acid (pH ADJ)	PM p-hydroxyphenylacetic acid	MacConkey with p-hydroxyphenylacetic acid	MacConkey with p-hydroxyphenylacetic acid (pH ADJ)	PM propionic acid	MacConkey with propionic acid (pH ADJ)
			pH^1: 6.8		pH^1: 6.7		pH^1: 6.4		pH^1: 4.1	pH^1: 6.5		pH^1: 6.5
			pH^2: 7.0		pH^2: 6.9		pH^2: 6.8		pH^2: 4.1	pH^2: 6.5		pH^2: 6.8
EcCa491a	-	Red	H. growth (colourless)	-	L. growth (pink)	-	M. growth (colourless)	-	No growth	H. growth (colourless)	W	H. growth (colourless)
EcHu93a	-	Red	H. growth (colourless)	-	L. growth (pink)	-	M. growth (colourless)	-	No growth	M. growth (colourless)	-	L. growth (colourless)
EcCa461b	-	Red	H. growth (colourless)	-	L. growth (pink)	-	M. growth (colourless)	-	No growth	H. growth (colourless)	W	H. growth (colourless)
EcCa452a	-	Red	H. growth (colourless)	-	L. growth (pink)	-	M. growth (colourless)	-	No growth	H. growth (colourless)	-	H. growth (colourless)
EcHu114a	-	Red	H. growth (colourless)	-	L. growth (pink)	-	M. growth (colourless)	-	No growth	H. growth (colourless)	-	H. growth (colourless)
EcCa26a	+	Yellow	H. growth (colourless)	+	H. growth (colourless)	+	M. growth (colourless)	+	No growth	H. growth (colourless, brownish black agar)	+	L. growth (colourless)
EcCa583a	+	Yellow	H. growth (colourless)	+	H. growth (colourless)	+	M. growth (colourless)	+	No growth	H. growth (colourless, brownish black agar)	W	No growth
EcCa21a	+	Yellow	H. growth (colourless)	+	H. growth (colourless)	+	M. growth (colourless)	-	No growth	M. growth (colourless)	W	L. growth (colourless)
EcCa519a	+	Yellow	H. growth (colourless)	+	H. growth (colourless)	W	M. growth (colourless)	+	No growth	H. growth (colourless)	W	H. growth (colourless)
EcCa24a	+	Yellow	H. growth (colourless)	+	H. growth (colourless)	+	M. growth (colourless)	+	No growth	M. growth (colourless)	W	H. growth (colourless)

PM: phenotypic microarray, CTA: Cystine trypticase agar, pH ADJ: pH adjusted, pH^1: pH value before autoclaving medium, pH^2: pH value after autoclaving medium, L: light, M: moderate, H: heavy, -: No respiration of *E. coli* isolates in sorbitol/D-serine/glycolic acid/p-hydroxyphenylacetic acid/propionic acid in PM1 , +: Positive respiration of *E. coli* isolates in sorbitol/D-serine/glycolic acid/p-hydroxyphenylacetic acid/propionic acid in PM1 , W: Weak respiration of *E. coli* isolates in glycolic acid/propionic acid in PM1

4.5 Genome sequencing

The four *E. coli* O157 and six *E. coli* O26 isolates were sequenced on a MiSeq platform (Illumina Inc, Australia), by NZGL, Massey University, Palmerston North. Figure 4.7 shows a graph of the cumulative frequency of trimmed read lengths for isolate EcCa24a. The figure showed that the trimmed read lengths were of good quality. The previously sequenced 13 *E. coli* O157 (George, Massey University, 2012), four *E. coli* O157, six *E. coli* O26 New Zealand isolates and the reference genomes were used in the genome analysis. Table 4.7 shows a summary of some of the key characteristics of the 17 *E. coli* O157 and six O26 NZ isolates. The average genome length of the five SP *E. coli* O157 isolates is 4.9 Mb, unlike the remaining 12 SN *E. coli* O157 isolates whose average genome length is 5.4 Mb. In the case of the *E. coli* O26 isolates, the average genome length of all the isolates (RP and RN) is 5.5 Mb. The standard deviation of the genome lengths for both serogroups is 0.2 Mb. Table 3.5 shows a summary of some of the key characteristics of the reference *E. coli* O157 and O26 isolates that were downloaded from the NCBI database. The average genome lengths of the *E. coli* O157 and O26 reference genomes are 5.5 Mb and 5.4 Mb, respectively. The genomes lengths of the SN *E. coli* O157 and RP and RN *E. coli* O26 New Zealand isolates have similar average genome lengths compared to that of the reference genomes, showing that they share the common genomic characteristics as the reference genomes along with acquired genomic characteristics or genes.

4.5.1 Core genome comparisons

The genetic associations of SN *E. coli* O157, SP *E. coli* O157, RN *E. coli* O26 and RP *E. coli* O26 isolates were studied using a phylogenetic approach.

4.5.1.1 *E. coli* O157 and *E. coli* O26 New Zealand isolates

The core genomes of 17 *E. coli* O157 and six *E. coli* O26 isolates were used to determine the genomic diversity between the New Zealand isolates. The NeighborNet tree was based on 2,037 genes all of the same length with a concatenated length of 665,021 amino acids (Figure 4.8). The tree illustrated the formation of three separate clusters. A large difference was evident between clusters 1, 2 and 3. Cluster 1 included the SN *E. coli* O157 isolates. Cluster 2, which included SP *E. coli* O157 isolates, was further divided into two groups. This indicated that there

were variations within the SP *E. coli* O157 isolates. However, in Cluster 3, less variation was seen between RP and RN *E. coli* O26 isolates in comparison to the variation seen between Cluster 1 and 2.

4.5.1.2 *E. coli* O157 New Zealand isolates and *E. coli* O157 reference isolates

Seventeen *E. coli* O157 isolates were compared to 37 reference *E. coli* O157 isolates. Twelve of the SN *E. coli* O157 isolates, including human and bovine isolates, formed a distinct cluster (Cluster 1) along with the other 37 reference isolates in the NeighborNet tree. The tree was built based on 1,513 genes all of the same length with a concatenated length of 475,774 amino acids (Figure 4.9). A large difference was observed between Cluster 1 and Clusters 2 and 3. Clusters 2 and 3 included the SP *E. coli* O157 isolates. A magnified image of Cluster 1 is inserted in the top right of the figure. On magnifying Cluster 1, the various groups formed by the SN *E. coli* O157 isolates was shown clearly. These groups also branched out further indicating that there was variation in the core genome used in the calculation of the figure.

4.5.1.3 *E. coli* O26 New Zealand isolates and *E. coli* O26 reference isolates

Six *E. coli* O26 isolates were compared to 20 reference *E. coli* O26 isolates. In Figure 4.10, the NeighborNet tree was based on 1,012 genes all of the same length with a concatenated length of 364,626 amino acids. The tree showed multiple branches with 11 of the reference isolates grouped together at the top of the tree. The RN *E. coli* O26 isolates (Cluster 2) were grouped close to Cluster 1. However, the RP *E. coli* O26 isolates branched further away from Cluster 1, especially isolates EcCa569a and EcCa5a in Cluster 3, illustrating a genomic diversity between the RP, RN and reference *E. coli* O26 isolates.

4.5.1.4 *E. coli* O157 and O26 New Zealand isolates and *E. fergusonii* reference isolates

E. fergusonii is a Gram-negative member of the Enterobacteriaceae family. It is rod-shaped and does not ferment lactose. It is also sorbitol negative and does not produce beta-glucuronidase, which are also some of the traits shared with the *E. coli* O157 serogroup (Farmer *et al.*, 1985, Rice *et al.*, 1991, Gaastra *et al.*, 2014). The core genomes of 17 *E. coli* O157 isolates and six *E. coli* O26 isolates were compared to three *E. fergusonii* reference isolates to determine if the fermenting isolates were closely related to the *E. fergusonii* isolates. The

NeighborNet tree, constructed on 1,860 genes all of the same length with a concatenated length of 583,909 amino acids, showed the isolates grouped into five clusters (Figure 4.11). Cluster 1 included SN *E. coli* O157 isolates, Cluster 2 and 3 included SP *E. coli* O157 isolates and Cluster 4 included both RN and RP *E. coli* O26 isolates. Clusters 1, 2, 3 and 4 showed small differences between them. A large difference was observed between all four clusters and Cluster 5 which contained the *E. fergusonii* reference isolates. This large difference indicated a genomic diversity between the *E. coli* O157 and O26 isolates and the *E. fergusonii* reference isolates.

4.5.2 Clusters of Orthologous Groups of proteins analysis

The Clusters of Orthologous Groups of proteins (COGs) system allows for the phylogenetic and functional annotations of genomes using the COGNITOR program and is used for evolutionary analyses (Natale *et al.*, 2000). The COG system was applied to annotate *E. coli* O157 (4 SN, 3 SP) and *E. coli* O26 (2 RN, 3 RP) isolates. These isolates were compared to each other to explore the similarities and differences in COG functions between and within the isolates. The NeighborNet tree was based on 1,293 COGs that showed variable counts across the genomes under investigation (Figure 4.12). The isolates formed three clusters caused by discordant signals. Cluster 1, which consisted of the SN *E. coli* O157 isolates, was separated from Cluster 2 which included the SP *E. coli* O157 isolates. In Cluster 2, isolate EcCa24a branched away from isolates EcCa26a and EcCa583a. However, in the case of the *E. coli* O26 isolates, multiple branches were formed in Cluster 3 and fewer conflicting signals were observed between the isolates. Cluster 3 included both RN and RP *E. coli* O26 isolates.

The matrix file derived from the COGNITOR software was processed in Microsoft Office Excel (Microsoft Corporation) using the 'STDEV' function until only the protein or protein domains that had 'differences/STDEV>0' in the *E. coli* O157 and O26 isolates were remaining. An example of the output from the 'STDEV' analysis is shown in Table 4.8. The COG categories of protein or protein domains in relation to carbon metabolism, G, H, O and Q, were searched for in all NZ isolates. In Figure 4.13, 47 differences were seen in G, H, O and Q categories between the SN *E. coli* O157 isolates compared to 89 differences between the SP *E. coli* O157 isolates. In addition to the *E. coli* O157 findings, 41 differences were present in G, H, O and Q categories between RN *E. coli* O26 isolates compared to 65 differences between RP *E. coli* O26 isolates. In the figure, the blue shaded areas indicated the endpoints at each stage of the flowchart.

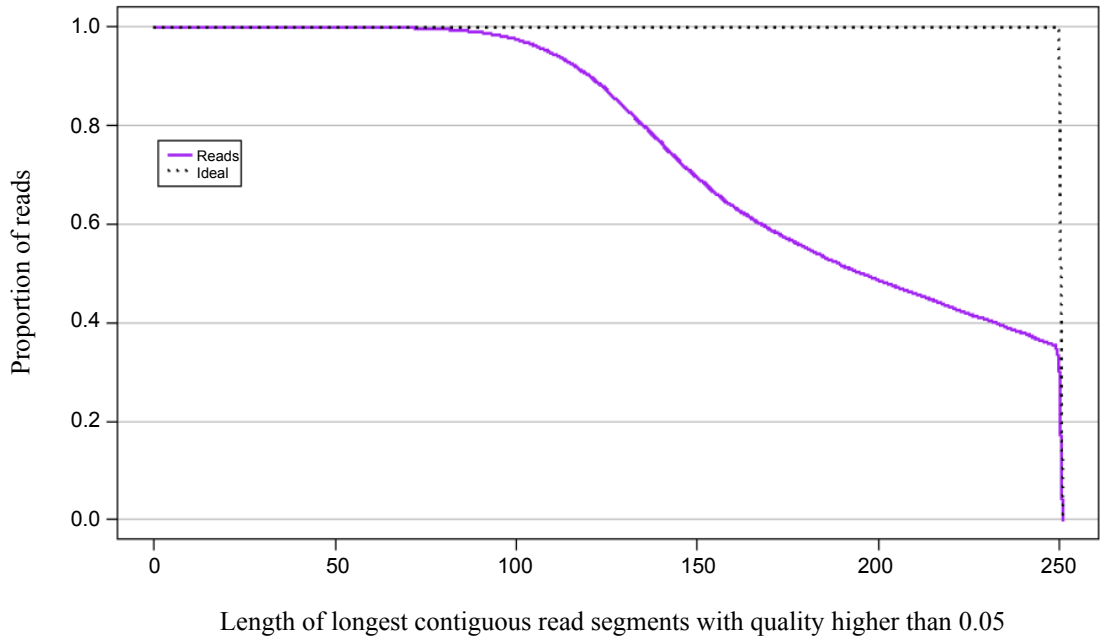

Figure 4.7. Cumulative frequency of trimmed read lengths for isolate EcCa24a . The line graph shows the cumulative frequency of trimmed read lengths of *E. coli* O157 isolate EcCa24a (purple line), with the perfect result shown as a dotted line. The quality control cutoff value was 0.05.
Source: Sequencing report by NZGL, Massey University

Table 4.7. Summary of genome characteristics of *E. coli* O157 and O26 New Zealand isolates.

Isolate name	Serotype	Genome length	Number of contigs	GC content	rRNA	Repeat region	Signal peptide	CDS	tRNA	tmRNA
EcCa452a	O157	5421117	187	50.22	4	1	463	5196	104	1
EcCa63b	O157	5457683	181	50.23	4	1	465	5233	98	1
EcCa491a	O157	5446753	176	50.3	4	0	465	5236	98	1
EcCa32a	O157	5389648	177	50.31	3	1	459	5162	102	1
EcCa461b	O157	5383438	182	50.25	4	1	459	5145	103	1
EcCa635a	O157	5439452	194	50.29	4	1	461	5220	103	1
EcCa26a	O157	5063509	117	50.55	6	2	447	4850	82	1
EcCa21a	O157	4938528	272	50.68	6	2	423	4762	72	1
EcCa24a	O157	4770969	230	50.72	2	2	408	4509	78	1
EcCa519a	O157	4720390	205	50.76	5	2	410	4442	60	1
EcCa583a	O157	5036580	242	50.48	4	2	446	4815	64	0
EcHu44a	O157	5394390	154	50.26	4	1	470	5154	94	1
EcHu93a	O157	5441529	235	50.36	4	1	469	5183	98	1
EcHu56a	O157	5266700	149	50.34	4	1	453	4985	96	1
EcHu76a	O157	5358378	151	50.34	4	1	457	5106	98	1
EcHu68a	O157	5441086	193	50.31	4	1	468	5224	101	1
EcHu114a	O157	5439010	199	50.28	3	1	460	5219	103	1
EcCa8a	O26	5633045	657	50.26	5	2	453	5579	74	1
EcCa513a	O26	5439437	647	50.37	4	2	429	5330	73	1
EcCa1a	O26	5574795	645	50.28	5	2	446	5523	75	1
EcCa5a	O26	5469836	467	50.43	4	2	450	5376	82	0
EcCa514a	O26	5467387	605	50.34	4	2	444	5397	72	0
EcCa569a	O26	5404868	439	50.42	4	2	448	5309	70	1

bp: base pairs, GC content: guanine-cytosine content in percentage, rRNA: mature ribosomal RNA, CDS: coding sequence, tRNA: mature transfer RNA, tmRNA: transfer messenger RNA

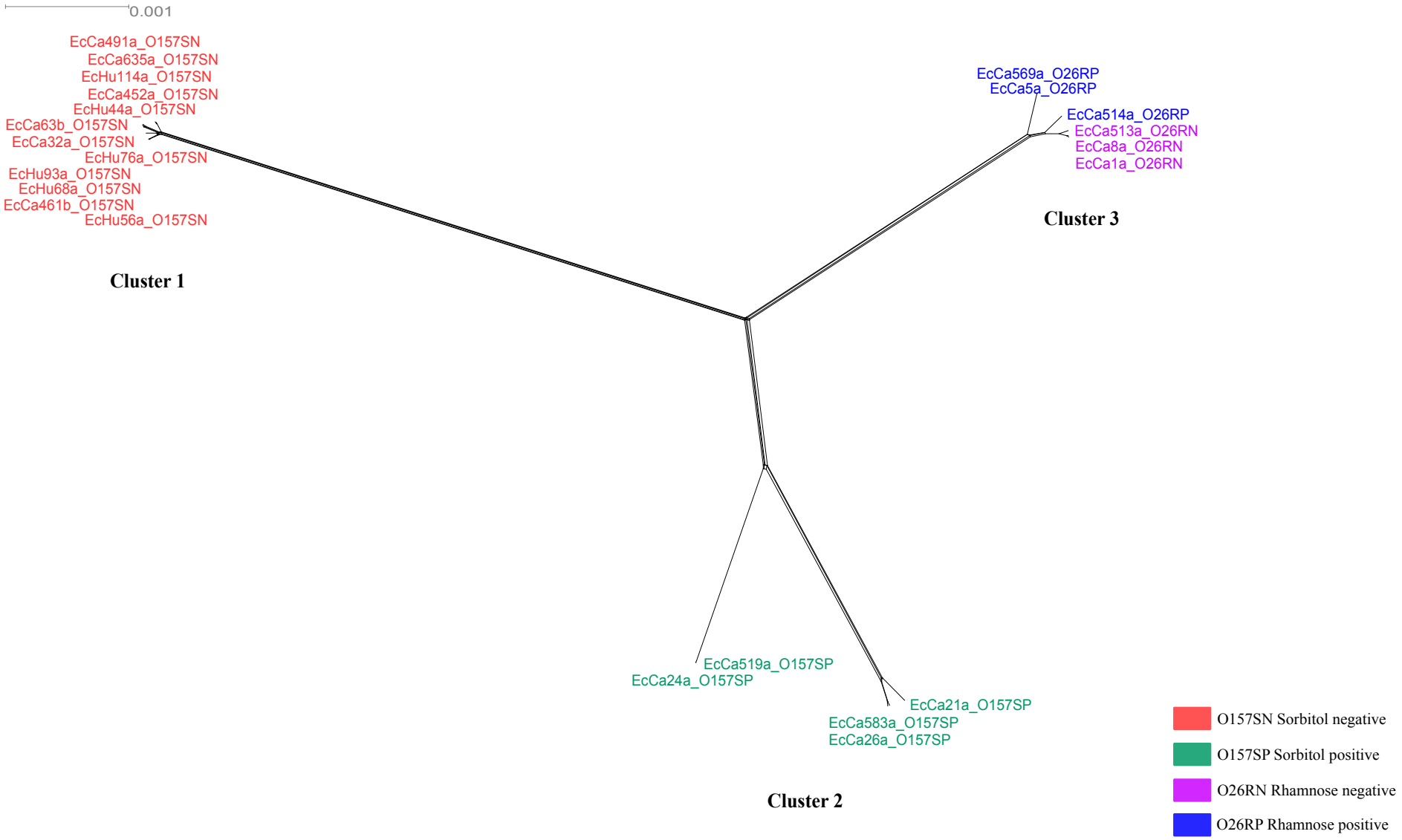

Figure 4.8. NeighborNet tree of *E. coli* O157 and *E. coli* O26 New Zealand isolates. The tree was based on 2,037 genes all of the same length with a concatenated length of 665,021 amino acids.
SN: Sorbitol negative or non-fermenter, SP: Sorbitol positive or fermenter, RN: Rhamnose negative or non-fermenter, RP: Rhamnose positive or fermenter

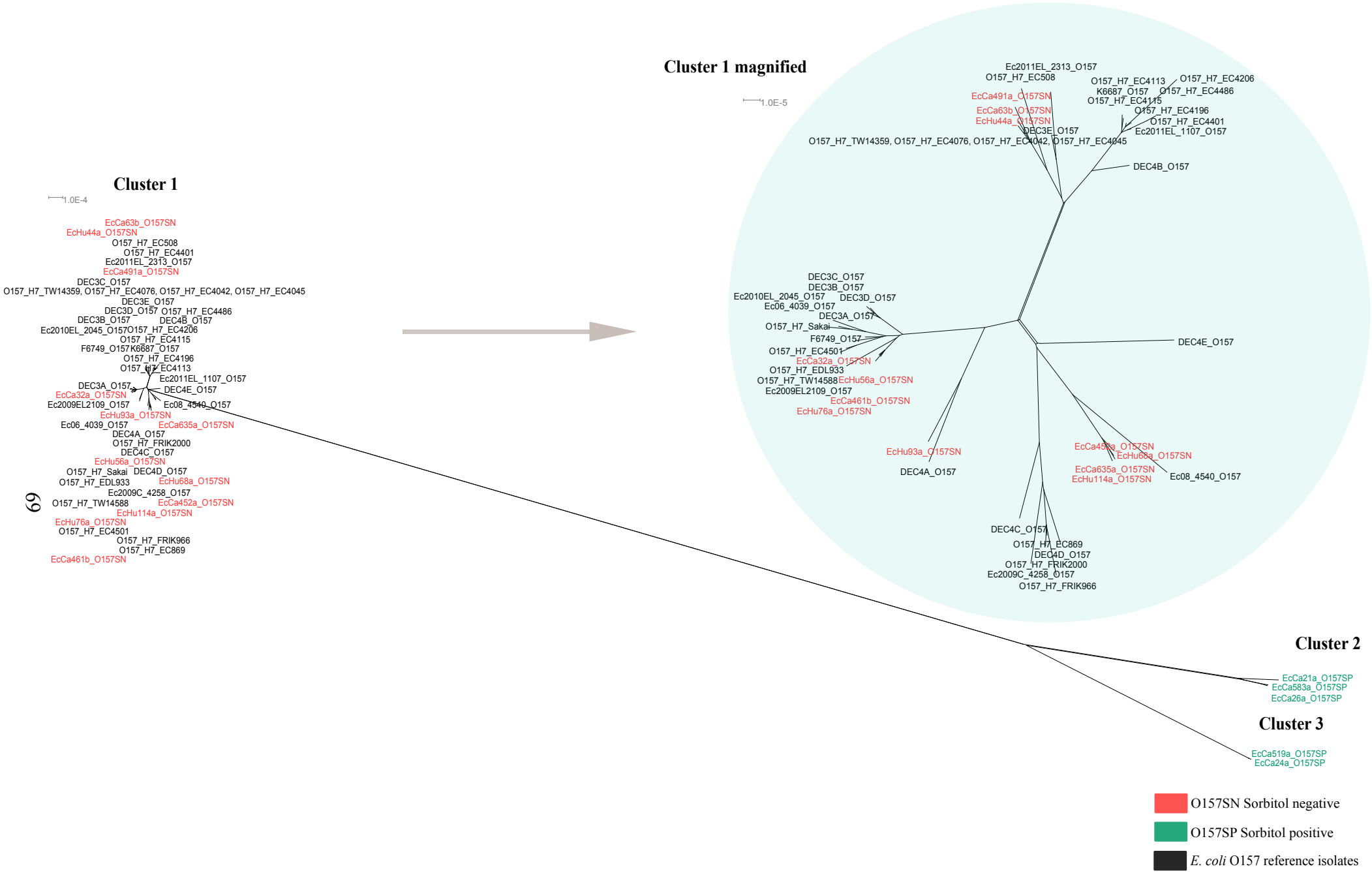

Figure 4.9. NeighborNet tree of *E. coli* O157 New Zealand isolates and *E. coli* O157 reference isolates. The tree was based on 1,513 genes all of the same length with a concatenated length of 475,774 amino acids.
SN: Sorbitol negative or non-fermenter, SP: Sorbitol positive or fermenter

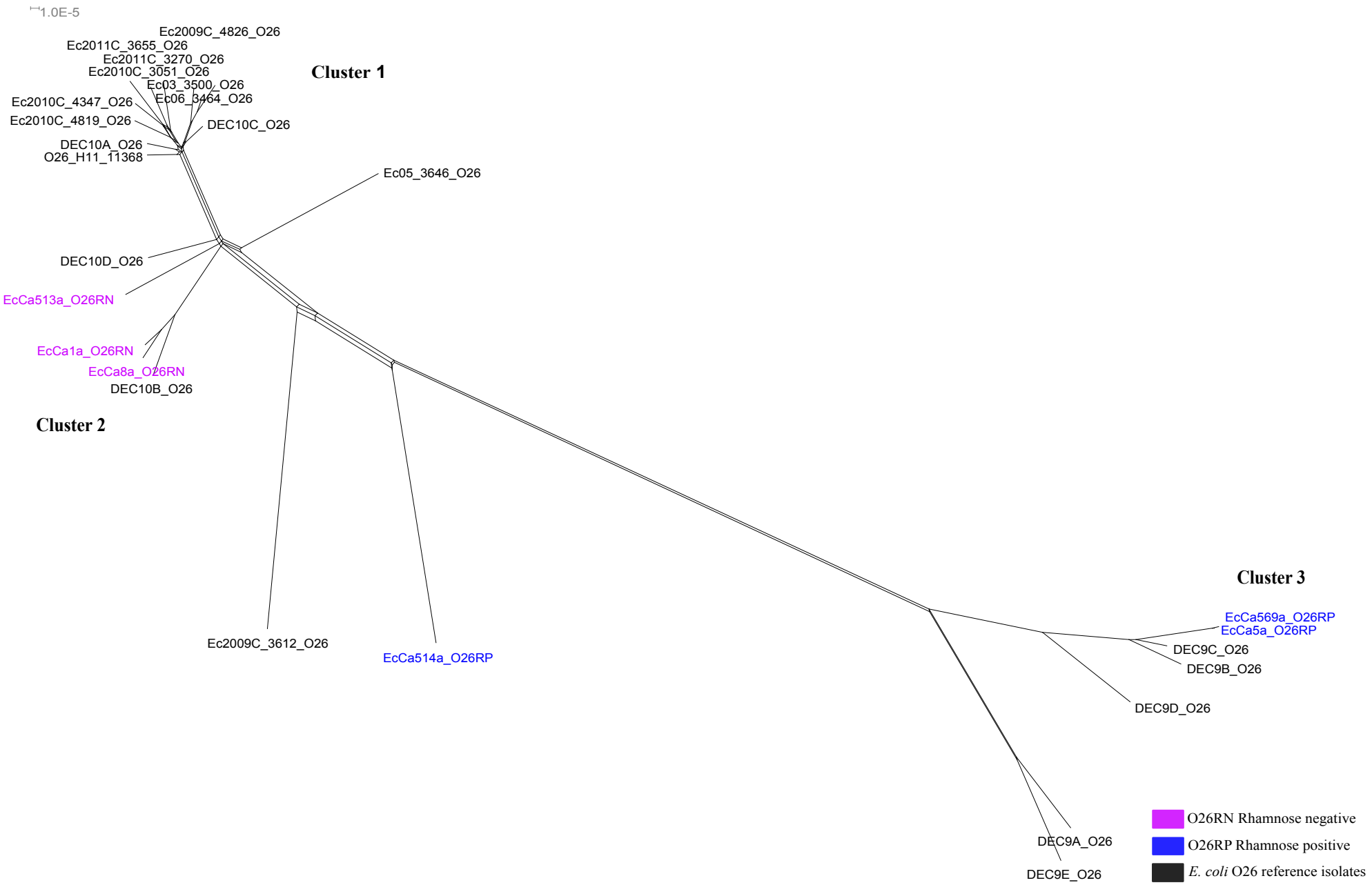

Figure 4.10. NeighborNet tree of *E. coli* O26 New Zealand isolates and *E. coli* O26 reference isolates. The tree was based on 1,012 genes all of the same length with a concatenated length of 364,626 amino acids.
RN: Rhamnose negative or non-fermenter, RP: Rhamnose positive or fermenter

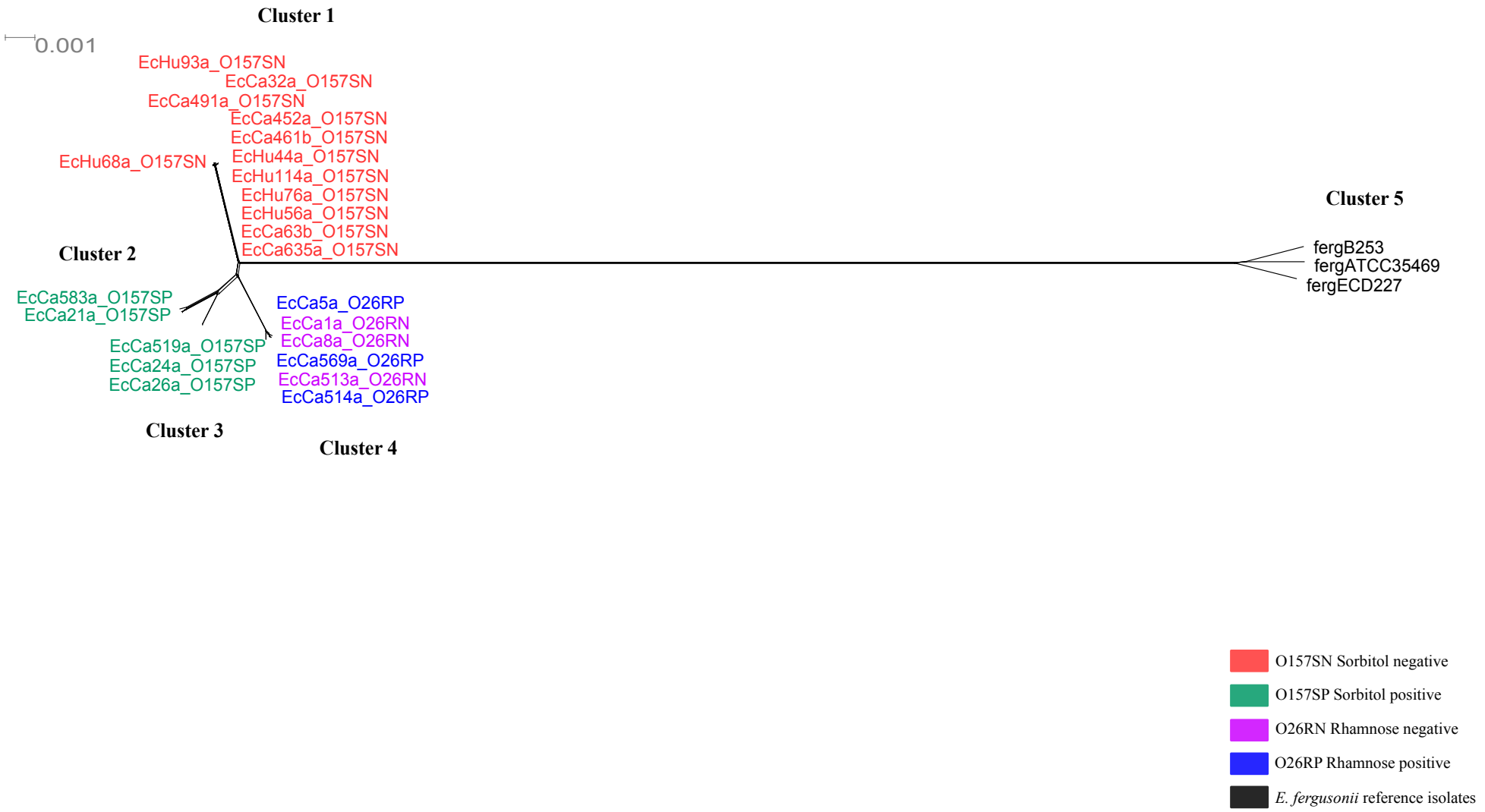

Figure 4.11. NeighborNet tree of *E. coli* O157 and O26 New Zealand isolates and *E. fergusonii* reference isolates. The tree was based on 1,860 genes all of the same length with a concatenated length of 583,909 amino acids.
SN: Sorbitol negative or non-fermenter, SP: Sorbitol positive or fermenter, RN: Rhamnose negative or non-fermenter, RP: Rhamnose positive or fermenter

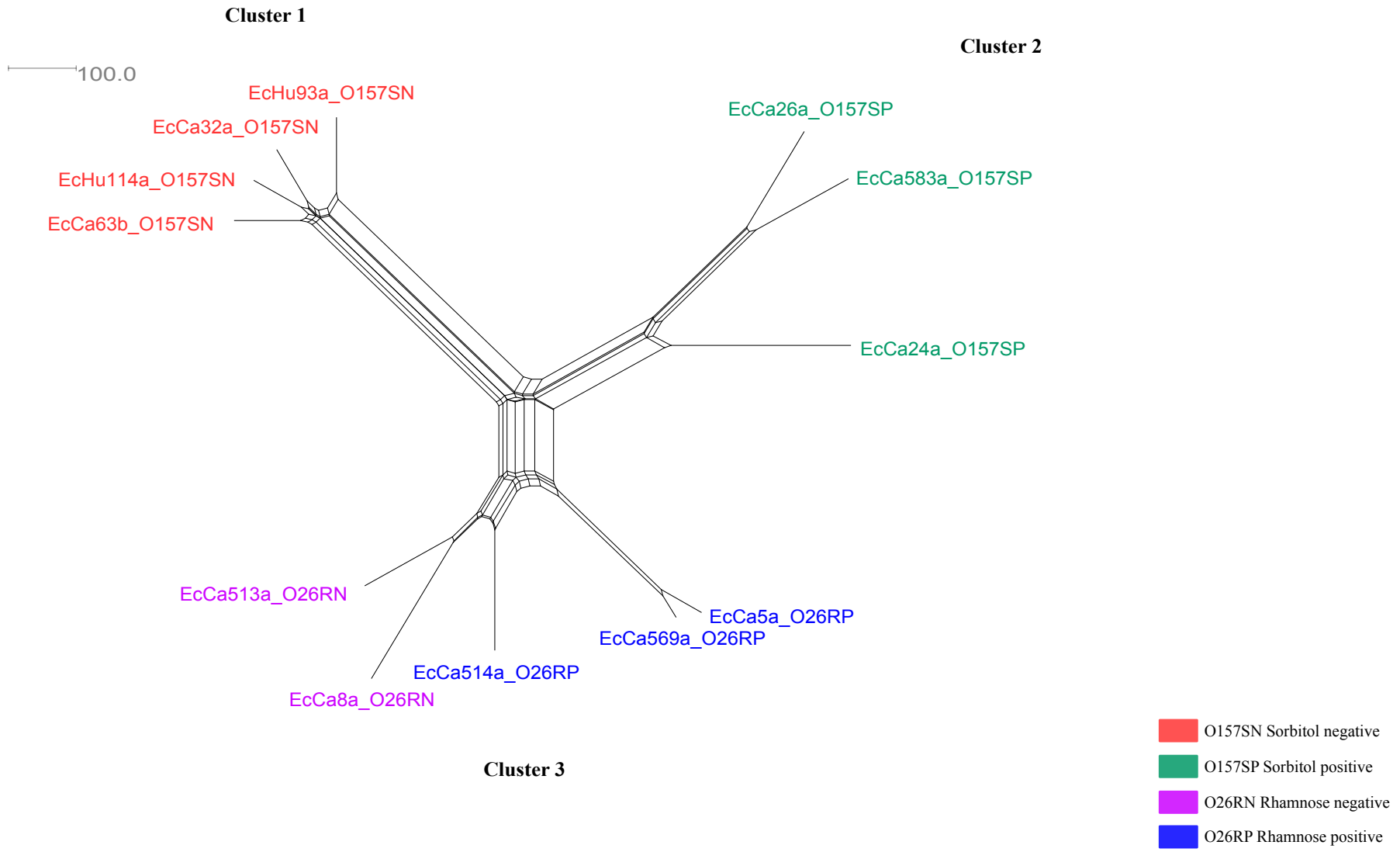

Figure 4.12. NeighborNet tree of *E. coli* O157 and O26 New Zealand isolates following COG analysis. The tree was based on 1,293 COGs that showed variable counts across the genomes under investigation.
SN: Sorbitol negative or non-fermenter, SP: Sorbitol positive or fermenter, RN: Rhamnose negative or non-fermenter, RP: Rhamnose positive or fermenter

Table 4.8. COG analysis of carbon metabolism in *E. coli* O157 isolates. Standard deviation was used to tabulate the number of differences in proteins/protein domains in sorbitol fermenting *E. coli* O157 isolates. A standard deviation of zero (green shaded area) showed either genes were absent or no differences were seen in the number of genes present for the proteins/domains between the isolates. The proteins/domains with standard deviation values greater than zero (purple shaded area) showed that either there were genes present or there were differences in the number of genes present related to the protein/domain between the isolates.

Protein/protein domain name	COG ID	COG category	Isolates			Standard deviation
			EcCa24a	EcCa26a	EcCa583a	
Flagellar motor protein MotB	COG1360	N	2	2	2	0
Type IV secretory pathway	COG3504	U	0	0	0	0
ATP-dependent Clp protease ATP-binding subunit ClpA	COG0542	O	2	4	3	1
Ribose/xylose/arabinose/galactoside ABC-type transport system	COG1172	G	9	8	8	0.577350269
ABC-type sugar transport system	COG1879	G	9	8	8	0.577350269
Phosphotransferase system	COG2893	G	4	3	3	0.577350269
Cob(I)alamin adenosyltransferase	COG2096	H	1	0	0	0.577350269
ABC-type transport system involved in multi-copper enzyme maturation	COG1277	O	0	1	0	0.577350269
Propanediol utilisation protein	COG4869	Q	1	0	0	0.577350269
Predicted ATP-dependent Lon-type protease	COG4930	O	0	1	0	0.577350269

COG: Clusters of Orthologous Groups of proteins

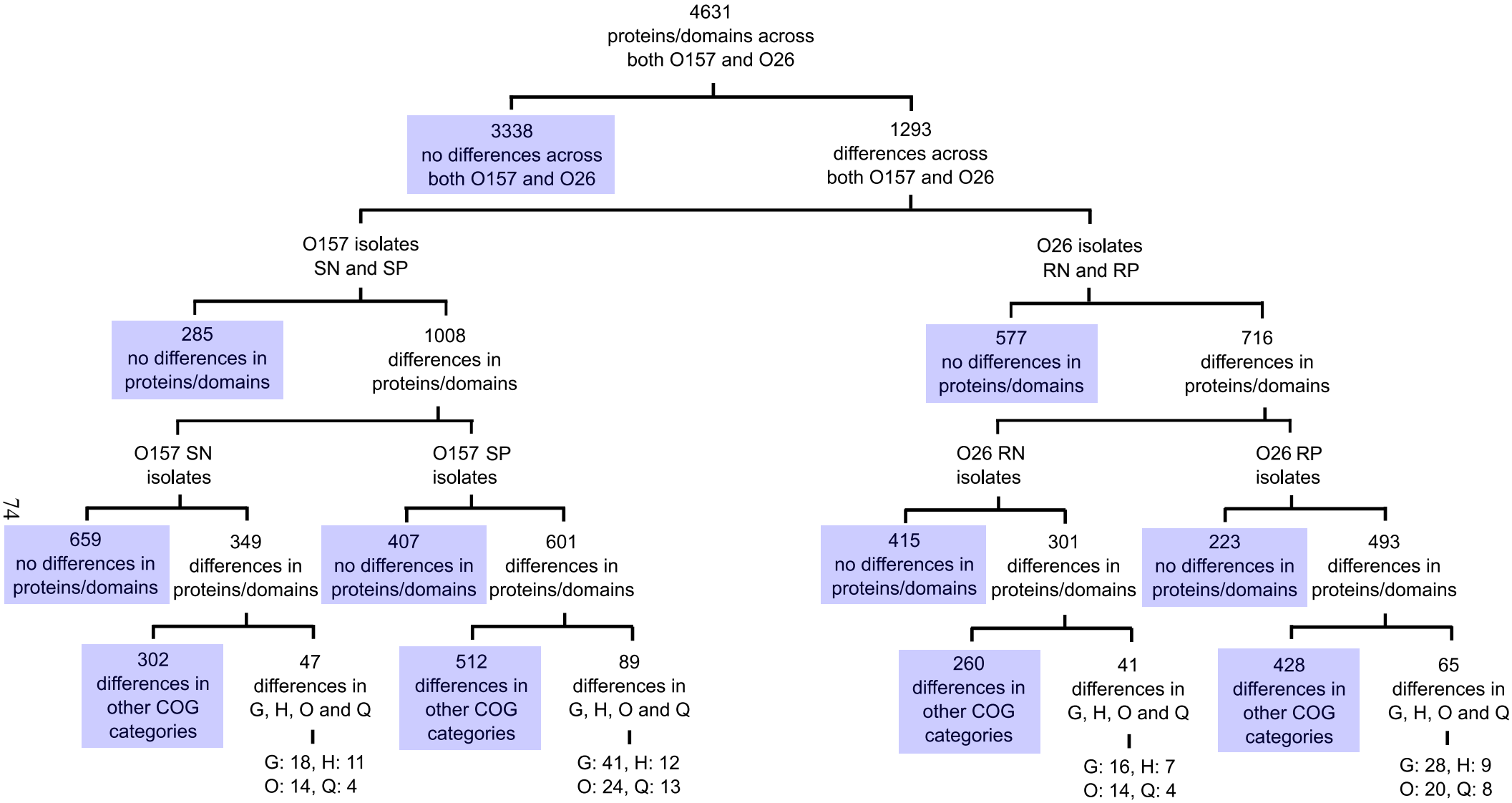

Figure 4.13. Number of differences in proteins/domains associated with carbon metabolism among *E. coli* O157 and O26 New Zealand isolates.The blue shaded areas indicate the endpoints at each stage of the flowchart. Clusters of Orthologous Groups of proteins (COGs) categories for carbon metabolism: G-Carbohydrate transport and metabolism, H-Coenzyme transport and metabolism, O-Posttranslational modification, protein turnover, chaperones, Q-Secondary metabolites biosynthesis, transport and catabolism

O157: *E. coli* O157 isolates, O26: *E. coli* O26 isolates, SN: Sorbitol negative or non-fermenter, SP: Sorbitol positive or fermenter, RN: Rhamnose negative or non-fermenter, RP: Rhamnose positive or fermenter

FIVE

Discussion

5.1 Introduction

The aims of the current study were to determine the phenotypic characteristics and the phylogeny of sorbitol fermenting or positive (SP) atypical *E. coli* O157 isolates and rhamnose fermenting or positive (RP) atypical *E. coli* O26 isolates. In order to achieve these objectives, the phenotypic behaviour and genetic profiles of sorbitol non-fermenting or negative (SN) *E. coli* O157 as well as rhamnose non-fermenting or negative (RN) *E. coli* O26 isolates needed to be determined. The differences and similarities observed in phenotypes and genotypes between both serogroups of *E. coli* isolates were determined by phenotypic microarray (PM) analysis, core genome comparisons and Clusters of Orthologous Groups of proteins (COGs) function analysis.

5.2 Shiga toxin-encoding bacteriophage insertion typing

Shiga toxin-encoding bacteriophage insertion typing has been widely used and studied in association with *E. coli* O157 strains and is useful in identifying virulent strains (Shringi *et al.*, 2012). The results of the current study showed that SBI typing was useful in demonstrating differences between and within *E. coli* O157 and O26 isolates. The findings showed strong diversity within *E. coli* O157 isolates. All the SN human and bovine *E. coli* O157 isolates possessed a Shiga toxin associated bacteriophage locus and at least one of the three *stx* genes. SBI typing revealed that the *stx2* gene was absent in half of the SN *E. coli* O157 isolates. However, the *stx2* gene was detected by preceding PCR suggesting that the *stx2* gene may have existed as a pseudogene or was a PCR generated artefact (Lerat & Ochman, 2005, Osborne *et al.*, 2005). Within the SN *E. coli* O157 isolates, the human isolates showed more diversity with six SBI genotypes compared to the bovine isolates with four SBI genotypes which was also reported in another study carried out on New Zealand *E. coli* O26 isolates (Irshad *et al.*, 2011).

The SP *E. coli* O157 isolates did not possess a SBI type. However, EcCa26a, a SP isolate possessed the *stx2* gene. Studies have shown that the presence of *stx* genes is associated with

STEC virulence. Eighty percent (12/15) of STEC isolates obtained from HUS and HC cases were found to possess *stx2* and 27% (4/15) of STEC isolates obtained from HUS cases possessed *stx1*, alone or in association with *stx2* (Pradel *et al.*, 2008). STEC isolates positive for the *stx2* gene are considered to be highly virulent due to their association with HUS and HC cases (Boerlin *et al.*, 1999, Friedrich *et al.*, 2002). The detection of the *stx2* gene in one SP and half of the SN *E. coli* O157 isolates suggests that they may be highly virulent and have the potential to cause disease.

Within the *E. coli* O26 bovine isolates, two RN and one RP isolates were shown to possess a Shiga toxin associated bacteriophage locus and only the RN isolates were positive for the *stx1* gene. The SBI typing revealed the presence of *stx1* gene in EcCa8a, a RN isolate. However, the *stx1* gene was not detected by preceding PCR suggesting that the *stx1* gene may have existed as a pseudogene in the genome or a part of the PCR primer sequence was missing (Lerat & Ochman, 2005, Dieffenbach *et al.*, 1993).

5.3 Phenotypic microarray technology and analysis

Sporadic cases of infection caused by *E. coli* through food (Espié *et al.*, 2006b, Espié *et al.*, 2006a, De Schrijver *et al.*, 2008), water (Swerdlow *et al.*, 1992), soil and sediment (Ogden *et al.*, 2001, Fremaux *et al.*, 2008) are a few examples that demonstrate the survival capabilities of *E. coli* to efficiently utilise diverse nutrients under challenging and dynamic conditions. Phenotypic characterisation of *E. coli* O157 and O26 New Zealand isolates based on carbon source utilisation revealed differences between and within the two serogroups. More carbon sources were utilised by SP *E. coli* O157 isolates compared to SN *E. coli* O157, RP and RN *E. coli* O26 isolates.

Phenotypic microarray analysis showed that the human isolates utilised over a third of the substrates while the animal isolates utilised about a quarter of the substrates. This suggests that human isolates have a greater metabolic repertoire than the animal isolates (Franz *et al.*, 2011). PM technology also showed that SP *E. coli* O157 isolates respired on D-serine while the SN *E. coli* O157 isolates did not respire on the same. The RN and RP *E. coli* O26 isolates respired on L-serine but did not respire on D-serine. Serine is abundantly present in the mucus of the bovine small intestine and its utilisation is important in the colonisation of the bovine intestine by *E. coli* (Montagne *et al.*, 2000, Bertin *et al.*, 2014). The SP *E. coli* O157, RN and RP *E. coli* O26

isolates were obtained from bovine samples. These isolates have the ability to survive in the bovine digestive tract by utilising serine. The SN *E. coli* O157 isolates were obtained from both bovine and human samples. PM technology indicated a correlation with sorbitol and D-serine metabolism in *E. coli* O157 isolates. However, no correlation was observed with rhamnose and D-serine or L-serine metabolism in *E. coli* O26 isolates.

The catabolism of D-serine by *E. coli* has been linked to virulence. *E. coli* catabolise D-serine to use as a nutrient and causes UTI (Alteri & Mobley, 2012, Roesch *et al.*, 2003). D-serine is synthesised by D-serine deaminase or dehydratase, *dsdA*, to pyruvate and further synthesis produces hydroxypyruvate which is used in the glycolytic pathway (McFall, 1964, Anfora & Welch, 2006, Schell, 2004). Genes encoding the D-serine tolerance locus, *dsdCXA*, implicates the catabolism of D-serine and triggers virulence (infection). The locus includes a transporter (*dsdX*) and a dehydratase (*dsdA*) which is regulated by a LysR-type transcriptional regulator (LTTR) (*dsdC*) (Nørregaard-Madsen *et al.*, 1995, Anfora & Welch, 2006, Connolly *et al.*, 2016). Also, a D-serine sensory locus *yhaOMKJ* includes an inner membrane transporter (*yhaO*) and a LTTR (*yhaJ*) which modulates the gene expression of virulence factors such as the pathogenicity island LEE (Connolly *et al.*, 2016, Connolly & Roe, 2016). However, the presence of D-serine can inhibit the expression of LEE and *yhaJ* and in turn affect the colonisation of *E. coli* in the intestinal tract of the host (Connolly *et al.*, 2015, Fitzhenry *et al.*, 2002, Connolly *et al.*, 2016). The SP *E. coli* O157 isolates utilised D-serine indicating the expression of *dsdCXA* and subsequent repression of the *yhaJ* gene. The SP *E. coli* O157 isolates may lack the ability to colonise but can survive in the intestinal tract of the host. Furthermore, the preceding PCR were negative for the *eae* gene (attaching and effacing lesions), which is a part of the LEE (Caprioli *et al.*, 2005), except for isolate EcCa26a which was found to be positive for the gene. However, the SN *E. coli* isolates did not respire on D-serine indicating the repression of *dsdCXA* and the expression of the *yhaJ* gene. The preceding PCR for the SN *E. coli* O157 isolates were positive for the *eae* gene. The SN *E. coli* O157 isolates may possess the ability to colonise the intestinal tract of the host and thus be more virulent than the SP *E. coli* O157 isolates.

Traditional testing methods, Cystine trypticase agar (CTA) and MacConkey medium were used to validate the PM findings. All the CTA medium with sorbitol tests produced similar findings as that of the PM technology. However, MacConkey agar with added D-serine failed to confirm the PM findings for all 10 isolates. On MacConkey with D-serine, the SN *E. coli* O157 isolates

produced a light growth of pink colonies whereas the SP *E. coli* O157 isolates produced a heavy growth of colourless colonies. The pink colonies indicated that the isolates fermented D-serine but the light growth indicated that D-serine may have also interfered with the growth of these isolates. D-serine inhibits the growth rate of bacteria but does not affect its viability (Davis & Maas, 1949, Maas & Davis, 1950). The metabolism of D-serine is driven by D-serine deaminase (*dsdA*). The addition of amino acids such as alanine or glycine to the growth medium or the expression of the functional *dsdA* could reverse the process of inhibition (Maas & Davis, 1950, Anfora & Welch, 2006). The expression of *dsdA* may have caused the SP *E. coli* O157 isolates to produce a heavy growth although D-serine was utilised and did not form acid.

Thirty percent of the isolates did not ferment the added glycolic acid, p-hydroxylphenylacetic acid (pH ADJ) or propionic acid added in the MacConkey agar which confirmed their respective findings from the PM technology. Isolates EcCa26a and EcCa583a produced heavy colourless growth on MacConkey agar with p-hydroxylphenylacetic acid (pH ADJ). The growth on the media was surrounded by black precipitate which made the media appear brownish black in colour. The accumulation of p-hydroxylphenylacetic acid in the media may have caused this change in colour (Carreira *et al.*, 2001, O'Connor *et al.*, 2001). No growth was observed for all the isolates on MacConkey agar with added p-hydroxyphenylacetic acid (not pH ADJ). With a pH of 4.1, the acidic medium may have been a harsh environment for the isolates and have inhibitory effects on bacterial growth (Eklund, 1989).

Future work could include pH and non-pH related tests to validate the findings of the PM analysis. Carbohydrate fermentation could be tested by pH related methods using Andrade peptone water (MacFaddin, 1985, HiMedia Laboratories, 2011, Sigma-Aldrich, no date). Non-pH dependent methods of chemotaxis assays such as the chemical-in-plug method and test-tube method can also be used to investigate carbohydrate fermentation (Tso & Adler, 1974).

5.4 Genome sequencing and analysis

Analysis of the genomes showed that there was diversity within the SN and SP *E. coli* O157 isolates whereas the RN and RP *E. coli* O26 isolates were found to be less diverse. The genome lengths of SN *E. coli* O157 isolates, RN and RP *E. coli* O26 isolates were larger than the SP *E. coli* O157 isolates. Studies have shown that the lengths of complete sequences of *E. coli* O26 serogroup are larger than those that belong to *E. coli* O157 serogroup (Hayashi *et al.*, 2001).

Prophages, insertion sequences and integrated elements contribute to these varying lengths and to the virulence repertoire of these isolates. The genomes have been shown to carry a core set of virulence factors such as pO157, haemolysin and LEE that are conserved across the various *E. coli* isolates (Perna *et al.*, 2001, Lim *et al.*, 2010b, Cooper *et al.*, 2014, Hayashi *et al.*, 2001).

The NeighborNet trees of the core genomes of the *E. coli* O157 isolates and *E. coli* O26 isolates showed conflicting signals between and within the two serogroups. The SN *E. coli* O157 isolates showed variations but were clustered with the *E. coli* O157 reference genomes while the SP *E. coli* O157 isolates were clustered further away. The RN and RP *E. coli* O26 isolates showed conflicting signals between isolates and when compared to the *E. coli* O26 reference genomes.

The phylogenetic analysis comparing the *E. coli* O157 and O26 isolates to *E. fergusonii* was performed to check that the fermenter isolates belonged to *E. coli* and not another species. The NeighborNet tree revealed that these *E. coli* O157 and O26 isolates were different to the *E. fergusonii* reference isolates. *E. fergusonii* was found to possess the O antigen of *E. coli* O157 suggesting that *E. fergusonii* acquired the O157 antigen cluster from *E. coli* O157 through horizontal gene transfer (HGT) (Fegan *et al.*, 2006).

The NeighborNet tree showed the *E. coli* O157 New Zealand isolates, except the SP *E. coli* O157 isolates, clustered near known and established *E. coli* O157 isolates derived from *E. coli* outbreaks that occurred in various parts of the world. Recombination or mutation events, like non-synonymous mutations within the same gene, HGT of DNA and mobility elements, like the pathogenicity island and plasmids, between bacteria giving rise to diversions from the evolutionary lineages and hence the conflicting signals within the *E. coli* O157 New Zealand and reference isolates (Manning *et al.*, 2008, Barrick *et al.*, 2009, Bentley & Parkhill, 2015, Wick *et al.*, 2005, Laing *et al.*, 2012). The concordant signals following the divergence suggest that the isolates and the acquired phage, gene or mobile elements co-evolved steadily (Laing *et al.*, 2012).

In the current study, the SP *E. coli* O157 had shorter genome lengths compared to the RP *E. coli* O26 isolates. Despite having larger genome lengths, the RP *E. coli* O26 isolates utilised fewer carbon substrates than the SP *E. coli* O157 isolates. Based on the carbon utilisation profile, the SP *E. coli* O157 isolates were shown to be metabolically more diverse than the

remaining isolates tested in this research. This phenotypic characteristic and the genes that encode the carbon metabolism might serve the SP *E. coli* O157 isolates as traits for fitness and enable their survival. Durso *et al*. (2004) reported that the *E. coli* O157:H7 human and cattle strains oxidised fewer carbon substrates compared to the commensal *E. coli* strains. Although the *E. coli* O157 strains had larger genomes in comparison to the commensal strains, it did not enhance the metabolic flexibility of the strains (Durso *et al.*, 2004).

Bacteria adapt their metabolism to available nutrients and environmental conditions to assist their survival and replication in any given hosts. Specific metabolic pathways enable the transportation and use of nutrients (Eisenreich *et al.*, 2010). Proteins and protein domains among other transcriptional regulators, encoded by genes, may directly or indirectly assist in metabolic pathways. The phosphotransferase system (PTS) along with cyclic AMP (cAMP) biosynthetic enzyme (adenylate cyclase), regulatory cAMP receptor protein (CRP, which includes chaperone proteins), glucose-specific IIA protein (IIA^{glc}), high energy glycolytic metabolite phosphoenolpyruvate (PEP) and other sugar-specific enzyme II complexes catalyses and controls the transport of carbohydrates and sugar derivatives in *E. coli* (Deutscher, 2008, Saier Jr *et al*., 1995, Saier, 1998, Brückner & Titgemeyer, 2002, Gosset *et al*., 2004). The COG categories G, H, O and Q were selected based on these protein or protein domains and regulators which are related to carbon metabolism.

For the phenotype-genotype analysis, the Clusters of Orthologous Groups of proteins database was used to get an insight into the functions, the number of genes and protein domains associated with carbohydrate metabolism in the *E. coli* O157 and O26 isolates. The NeighborNet distance matrix showed that there were differences in COG functions between the *E. coli* O157 and O26 isolates. Conflicting signals were observed between the *E. coli* O157 isolates. The findings showed 42 more differences in the presence or number of genes related to the protein or protein domains associated with carbon metabolism in the SP *E. coli* O157 isolates compared to those seen in the SN *E. coli* O157 isolates. The differences observed in the RN and RP *E. coli* O26 isolates were fewer compared to those seen in the *E. coli* O157 isolates. The differences in number of genes related to the protein or protein domains may be caused due to loss or gain of genes encoding these protein or protein domains through events of homologous recombination or mutations. The evolutionary process of gene duplication caused the duplication of only one out of 451 enzymes in *E. coli* since its divergence from *Salmonella* ~100 million years ago (Pál *et al.*, 2005). Phylogenetic analysis of proteobacteria

species including *E. coli* K12 estimated that HGT of 15-32 genes occurred during the divergence of *E. coli* from Salmonella over the same period of time. Seven percent of these genes belonged to the metabolic network of *E. coli* (Pál *et al.*, 2005). The large gene number changes seen in the current research suggests that multiple events of HGT and duplication of genes may have occurred during the divergence of SP *E. coli* O157, RN and RP *E. coli* O26 strains from the SN *E. coli* O157 strain over a period of time. The transfer of genes through evolution or condition-specific demands may contribute to the fitness or adaptation of bacteria in new environments. However, not all function is preserved during protein evolution. Therefore, it would be erroneous to associate gene function solely on protein similarity (Konopka & Crabbe, 2004).

The genes associated with the carbon metabolic pathway can be investigated through for example, the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (Kanehisa, 1995) for future work. The sequences for the genes encoding regulatory enzymes and subsequent protein domains can be obtained from KEGG. PhenoLink is another web-tool which can be used in any future research to determine any correlations between the phenotype and the genotype for a given isolate (Bayjanov *et al.*, 2012).

In summary, the *E. coli* O157 isolates demonstrated that differences in their carbon source utilisation may be correlated to the differences observed in their genomic characteristics. Furthermore, the SP *E. coli* O157 isolates demonstrated a greater metabolic repertoire and possessed shorter genomes compared to the SN *E. coli* O157, RN and RP *E. coli* O26 isolates. Although, the *E. coli* O26 isolates showed similar phenotypic profiles, differences were observed in their genotypic profiles.

5.5 Limitations of the research

The limitations of the current study are that the sample set for *E. coli* O26 was much smaller compared to that of *E. coli* O157. For the PM tests, different batches of plates, inoculating fluids and dye A were used between assays for both PM1 and PM2A. The number of replicate assays was not set a priori for both serogroups in PM technology. The substrates were added to the wells of the PM plates at the time of manufacture which made it difficult to determine the plate contents. The concentrations of the substrates may have been low or in a manner

unavailable to the *E. coli* isolates. Environmental conditions such as temperature may have affected the temperature in the OmniLog® incubator during an overnight incubation.

5.6 Conclusion

The findings of the current research have helped to advance our understanding of the behaviour and pathogenicity of *E. coli* O157 and O26 isolates and are important from both a public health and economic perspective. Consumption of contaminated food and water by pathogenic *E. coli* O157 and non-O157 serogroups like *E. coli* O26 can cause life threatening illnesses. The findings of this research highlight the need for rapid, sensitive and specific detection methods to detect atypical *E. coli* O157, typical and atypical *E. coli* O26 strains in the meat industry to identify contaminated products that can then be withdrawn from processing and to minimise the risk of food related *E. coli* outbreaks. More specific detection methods are also required to aid in identification, diagnosis and treatment of atypical *E. coli* O157, typical and atypical *E. coli* O26 related infections in humans.

References

Aidelberg, G., Towbin, B.D., Rothschild, D., Dekel, E., Bren, A. & Alon, U., (2014) Hierarchy of non-glucose sugars in *Escherichia coli*. *BMC systems biology* **8**: 133.

Aleksić, S., Karch, H. & Bockemühl, J., (1992) A biotyping scheme for Shiga-like (vero) toxin-producing *Escherichia coli* O157 and a list of serological cross-reactions between O157 and other gram-negative bacteria's. *Zentralblatt für bakteriologie* **276**: 221-230.

Alfredo, C., Alberto, E.T., Gianfranco, R. & Helge, K., (1997) Non-O157 Shiga toxin-producing *Escherichia coli* infections in Europe. *Emerging infectious disease* **3**: 578.

Alteri, C.J. & Mobley, H.L.T., (2012) *Escherichia coli* physiology and metabolism dictates adaptation to diverse host microenvironments. *Current opinion in microbiology* **15**: 3-9.

Anderson, R.L. & Wood, W.A., (1969) Carbohydrate metabolism in microorganisms. *Annual review of microbiology* **23**: 539-578.

Anfora, A.T. & Welch, R.A., (2006) DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073. *Journal of bacteriology* **188**: 6622-6628.

Anonymous, (no date-a) *Escherichia coli* O157 latex test. In.: Thermo Fisher Scientifc, https://www.thermofisher.com/order/catalog/product/DR0620M, accessed 2017.

Anonymous, (no date-b) Monera (bacteria). In.: Leaving certificate biology, http://leavingbio.net/, accessed 2015.

Armstrong, G.L., Hollingsworth, J. & Morris Jr, J.G., (1996) Emerging foodborne pathogens: *Escherichia coli* O157: H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiologic reviews* **18**: 29.

Baker, M., Eyles, R., Bennett, J., Nicol, C., Wong, W. & Garrett, N., (1999) Emergence of verotoxigenic *Escherichia coli* (VTEC) in New Zealand. *New Zealand Public Health Report* **6**: 9-12.

Baljer, G., (1997) EHEC animal reservoirs. *Fortschritte der medizin* **115**: 84.

Barlow, R.S. & Mellor, G.E., (2010) Prevalence of enterohemorrhagic *Escherichia coli* serotypes in Australian beef cattle. *Foodborne pathogens and disease* **7**: 1239-1245.

Barrett, T.J., Kaper, J.B., Jerse, A.E. & Wachsmuth, I.K., (1992) Virulence factors in Shiga-like toxin-producing *Escherichia coli* isolated from humans and cattle. *Journal of infectious diseases* **165**: 979-980.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E. & Kim, J.F., (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**: 1243-1247.

Bayjanov, J.R., Molenaar, D., Tzeneva, V., Siezen, R.J. & van Hijum, S.A., (2012) PhenoLink-a web-tool for linking phenotype to ~omics data for bacteria: application to gene-trait matching for *Lactobacillus plantarum* strains. *BMC genomics* **13**: 170.

Becton Dickinson and Company, (2012) CTA medium™ with sorbitol (QC/PI Manual). Maryland, USA, pp. 3.

Becton Dickinson and Company, (no date) Difco™ & BBL™ manual (2nd edition): MacConkey agar. Maryland, USA, pp. 5.

Bell, B.P., Goldoft, M., Griffin, P.M., Davis, M.A., Gordon, D.C., Tarr, P.I., Bartleson, C.A., Barrett, T.J., Wells, J.G., Baron, R. & Kobayashi, J., (1994) A multistate outbreak of *Escherichia coli* O157:H7—associated bloody diarrhea and hemolytic uremic syndrome from hamburgers: the Washington experience. *Journal of the American medical association* **272**: 1349-1353.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L. & Bignell, H.R., (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.

Bentley, S.D. & Parkhill, J., (2004) Comparative genomic structure of prokaryotes. *Annual review of genetics* **38**: 771-791.

Bentley, S.D. & Parkhill, J., (2015) Genomic perspectives on the evolution and spread of bacterial pathogens. *Proceedings of the Royal Society B: Biological Sciences* **282**.

Bertin, Y., Deval, C., de la Foye, A., Masson, L., Gannon, V., Harel, J., Martin, C., Desvaux, M. & Forano, E., (2014) The gluconeogenesis pathway is involved in maintenance of enterohaemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content. *PLOS ONE* **9**: e98367.

Bertrand, R. & Roig, B., (2007) Evaluation of enrichment-free PCR-based detection on the *rfbE* gene of *Escherichia coli* O157-application to municipal wastewater. *Water research* **41**: 1280-1286.

Besser, R.E., Griffin, P.M. & Slutsker, L., (1999) *Escherichia coli* O157: H7 gastroenteritis and the hemolytic uremic syndrome: an emerging infectious disease *Annual review of medicine* **50**: 355-367.

Besser, T.E., Shaikh, N., Holt, N.J., Tarr, P.I., Konkel, M.E., Malik-Kale, P., Walsh, C.W., Whittam, T.S. & Bono, J.L., (2007) Greater diversity of Shiga toxin-encoding bacteriophage insertion sites among *Escherichia coli* O157: H7 isolates from cattle than in those from humans. *Applied and environmental microbiology* **73**: 671-679.

Bettelheim, K.A., (1998) A spectrum of verocytotoxigenic *Escherichia coli*. *Journal of medical microbiology* **47**: 1037-1038.

Bettelheim, K.A., (2007) The non-O157 Shiga-toxigenic (verocytotoxigenic) *Escherichia coli*; under-rated pathogens. *Critical reviews in microbiology* **33**: 67-87.

Beutin, L., (2006) Emerging enterohaemorrhagic *Escherichia coli*, causes and effects of the rise of a human pathogen. *Journal of veterinary medicine, series B* **53**: 299-305.

Beutin, L., Geier, D., Steinrück, H., Zimmermann, S. & Scheutz, F., (1993) Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. *Journal of clinical microbiology* **31**: 2483-2488.

Bhunia, A.K., (2008) *Escherichia coli* In: Foodborne microbial pathogens: mechanisms and pathogenesis. Springer New York, pp. 183-200.

Bielaszewska, M., Zhang, W., Tarr, P.I., Sonntag, A.-K. & Karch, H., (2005) Molecular profiling and phenotype analysis of *Escherichia coli* O26:H11 and O26:NM: secular and geographic consistency of enterohemorrhagic and enteropathogenic isolates. *Journal of clinical microbiology* **43**: 4225-4228.

Bilge, S.S., Vary, J.C., Dowell, S.F. & Tarr, P.I., (1996) Role of the *Escherichia coli* O157:H7 O side chain in adherence and analysis of an *rfb* locus. *Infection and immunity* **64**: 4795-4801.

Bochner, B.R., (2009) Global phenotypic characterization of bacteria. *FEMS microbiology reviews* **33**: 191-205.

Bochner, B.R., Gadzinski, P. & Panomitros, E., (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome research* **11**: 1246-1255.

Bochner, B.R. & Savageau, M.A., (1977) Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms. *Applied and environmental microbiology* **33**: 434-444.

Boerlin, P., McEwen, S.A., Boerlin-Petzold, F., Wilson, J.B., Johnson, R.P. & Gyles, C.L., (1999) Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *Journal of clinical microbiology* **37**: 497-503.

Bower, C.K., McGuire, J. & Daeschel, M.A., (1996) The adhesion and detachment of bacteria and spores on food-contact surfaces. *Trends in food science & technology* **7**: 152-157.

Brooks, H., Bettelheim, K., Todd, B. & Holdaway, M., (1997) Non-O157 vero cytotoxin producing *Escherichia coli*: aetiological agents of diarrhoea in children in Dunedin, New Zealand. *Comparative immunology, microbiology and infectious diseases* **20**: 163-170.

Brooks, J.T., Sowers, E.G., Wells, J.G., Greene, K.D., Griffin, P.M., Hoekstra, R.M. & Strockbine, N.A., (2005) Non-O157 Shiga toxin–producing *Escherichia coli* infections in the United States, 1983–2002. *Journal of infectious diseases* **192**: 1422-1429.

Brown, S.A., Palmer, K.L. & Whiteley, M., (2008) Revisiting the host as a growth medium. *Nature reviews microbiology* **6**: 657-666.

Brückner, R. & Titgemeyer, F., (2002) Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS microbiology letters* **209**: 141-148.

Brunder, W., Schmidt, H. & Karch, H., (1997) EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Molecular microbiology* **24**: 767-778.

Bryant, D. & Moulton, V., (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution* **21**: 255-265.

Brzuszkiewicz, E., Thürmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F.-D., Boelter, J., Petersen, H., Gottschalk, G. & Daniel, R., (2011) Genome sequence analyses of two isolates from the recent *Escherichia coli* outbreak in Germany reveal the emergence of a new pathotype: entero-aggregative-haemorrhagic *Escherichia coli* (EAHEC). *Archives of microbiology* **193**: 883-891.

Buermans, H.P.J. & den Dunnen, J.T., (2014) Next generation sequencing technology: advances and applications. *Biochimica et biophysica acta (BBA) - molecular basis of disease* **1842**: 1932-1941.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G. & Knight, R., (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal* **6**: 1621-1624.

Caprioli, A., Morabito, S., Brugère, H. & Oswald, E., (2005) Enterohaemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Veterinary research* **36**: 289-311.

Carreira, A., Ferreira, L.s.M. & Loureiro, V.l., (2001) Brown pigments produced by *Yarrowia lipolytica* result from extracellular accumulation of homogentisic acid. *Applied and environmental microbiology* **67**: 3463-3468.

Cavanagh, P.D., (2003) Exports. In: New Zealand official yearbook. P.D. Cavanagh (ed). Auckland, NZ: Statistics New Zealand: David Bateman Ltd. In, pp. 533-537.

Chapman, P.A., Siddons, C.A., Wright, D.J., Norman, P., Fox, J. & Crick, E., (1993) Cattle as a possible source of verocytotoxin-producing *Escherichia coli* O157 infections in man. *Epidemiology & infection* **111**: 439-448.

Clark, D.P. & Cronan, J.E., (2005) Two-carbon compounds and fatty acids as carbon sources. *EcoSal plus* **1**.

Cloud, J.L., Harmsen, D., Iwen, P.C., Dunn, J.J., Hall, G., LaSala, P.R., Hoggan, K., Wilson, D., Woods, G.L. & Mellmann, A., (2010) Comparison of traditional phenotypic identification methods with partial 5′ 16S rRNA gene sequencing for species-level identification of nonfermenting gram-negative bacilli. *Journal of clinical microbiology* **48**: 1442-1444.

Collins, F. & Galas, D., (1993) A new five-year plan for the U.S. human genome project. *Science* **262**: 43-46.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. & Walters, L., (1998) New goals for the U.S. human genome project: 1998-2003. *Science* **282**: 682-689.

Connolly, J.P., Goldstone, R.J., Burgess, K., Cogdell, R.J., Beatson, S.A., Vollmer, W., Smith, D.G. & Roe, A.J., (2015) The host metabolite D-serine contributes to bacterial niche specificity through gene selection. *The ISME journal* **9**: 1039-1051.

Connolly, J.P.R., Gabrielsen, M., Goldstone, R.J., Grinter, R., Wang, D., Cogdell, R.J., Walker, D., Smith, D.G.E. & Roe, A.J., (2016) A highly conserved bacterial D-serine uptake system links host metabolism and virulence. *PLOS pathogens* **12**: e1005359.

Connolly, J.P.R. & Roe, A.J., (2016) When and where? Pathogenic *Escherichia coli* differentially sense host D-serine using a universal transporter system to monitor their environment. *Microbial cell* **3**: 181-184.

Cookson, A.L., Taylor, S.C.S. & Attwood, G.T., (2006a) The prevalence of Shiga toxin-producing *Escherichia coli* in cattle and sheep in the lower North Island, New Zealand. *New Zealand veterinary journal* **54**: 28-33.

Cookson, A.L., Taylor, S.C.S., Bennett, J., Thomson-Carter, F. & Attwood, G.T., (2006b) Serotypes and analysis of distribution of Shiga toxin-producing *Escherichia coli* from cattle and sheep in the lower North Island, New Zealand. *New Zealand veterinary journal* **54**: 78-84.

Coombes, B.K., Gilmour, M.W. & Goodman, C.D., (2011) The evolution of virulence in non-O157 Shiga toxin-producing *Escherichia coli*. *Front Microbiol* **2**.

Coombes, B.K., Wickham, M.E., Mascarenhas, M., Gruenheid, S., Finlay, B.B. & Karmali, M.A., (2008) Molecular analysis as an aid to assess the public health risk of non-O157 Shiga toxin-producing *Escherichia coli* strains. *Applied and environmental microbiology* **74**: 2153-2160.

Cooper, K.K., Mandrell, R.E., Louie, J.W., Korlach, J., Clark, T.A., Parker, C.T., Huynh, S., Chain, P.S., Ahmed, S. & Carter, M.Q., (2014) Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7. *BMC genomics* **15**: 17.

Cox, M.P., Peterson, D.A. & Biggs, P.J., (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics* **11**: 485.

Crump, J.A., Murdoch, D.R. & Baker, M.G., (2001) Emerging infectious diseases in an island ecosystem: the New Zealand perspective. *Emerging infectious diseases* **7**: 767-772.

Davis, B.D. & Maas, W.K., (1949) Inhibition of *E. coli* by D-serine and the production of serine-resistant mutants. *Journal of the American chemical society* **71**: 1865-1865.

De Schrijver, K., Buvens, G., Possé, B., Van den Branden, D., Oosterlynck, O., De Zutter, L., Eilers, K., Piérard, D., Dierick, K. & Van Damme-Lombaerts, R., (2008) Outbreak of verocytotoxin-producing *E. coli* O145 and O26 infections associated with the consumption of ice cream produced at a farm, Belgium, 2007. *Eurosurveillance* **13**: 61-64.

Deutscher, J., (2008) The mechanisms of carbon catabolite repression in bacteria. *Current opinion in microbiology* **11**: 87-93.

Dieffenbach, C., Lowe, T. & Dveksler, G., (1993) General concepts for PCR primer design. *PCR methods and applications* **3**: S30-S37.

Doyle, M.P. & Schoeni, J.L., (1984) Survival and growth characteristics of *Escherichia coli* associated with hemorrhagic colitis. *Applied and environmental microbiology* **48**: 855-856.

Durso, L.M., Bono, J.L. & Keen, J.E., (2005) Molecular serotyping of *Escherichia coli* O26:H11. *Applied and environmental microbiology* **71**: 4941-4944.

Durso, L.M., Smith, D. & Hutkins, R.W., (2004) Measurements of fitness and competition in commensal *Escherichia coli* and *E. coli* O157:H7 strains. *Applied and environmental microbiology* **70**: 6466-6472.

Eisenreich, W., Dandekar, T., Heesemann, J. & Goebel, W., (2010) Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nature reviews microbiology* **8**: 401-412.

Eklund, T., (1989) Organic acids and esters. In: Mechanisms of action of food preservation procedures. G.W. Gould (ed). pp. 161-200.

Elliott, E.J., Robins-Browne, R.M., O'Loughlin, E.V., Bennett-Wood, V., Bourke, J., Henning, P., Hogg, G.G., Knight, J., Powell, H., Redmond, D. & Contributors to the Australian paediatric surveillance unit, (2001) Nationwide study of haemolytic uraemic syndrome: clinical, microbiological, and epidemiological features. *Archives of disease in childhood* **85**: 125-131.

Entner, N. & Doudoroff, M., (1952) Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. *Journal of biological chemistry* **196**: 853-862.

Erdem, A.L., Avelino, F., Xicohtencatl-Cortes, J. & Girón, J.A., (2007) Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. *Journal of bacteriology* **189**: 7426-7435.

Espié, E., Grimont, F., Mariani-Kurkdjian, P., Bouvet, P., Haeghebaert, S., Filliol, I., Loirat, C., Decludt, B., Minh, N.N.T., Vaillant, V. & de Valk, H., (2008) Surveillance of hemolytic uremic syndrome in children less than 15 years of age, a system to monitor O157 and non-O157 Shiga toxin-producing *Escherichia coli* infections in France, 1996–2006. *Pediatric infectious disease journal* **27**: 595-601.

Espié, E., Mariani-Kurkdjian, P., Grimont, F., Pihier, N., Vaillant, V., Francart, S., De Valk, H. & Vernozy-Rozand, C., (2006a) Shigatoxin producing *Escherichia coli* O26 infection and unpasteurised cows cheese, France 2005. In: 6th International symposium on STEC, Melbourne.

Espié, E., Vaillant, V., Mariani-Kurkdjian, P., Grimont, F., Martin-Schaller, R., De Valk, H. & Vernozy-Rozand, C., (2006b) *Escherichia coli* O157 outbreak associated with fresh unpasteurized goats' cheese. *Epidemiology and infection* **134**: 143-146.

ESR, (2016) Notifiable diseases in New Zealand: annual report 2015. Porirua, New Zealand: The Institute of Environmental Science and Research Ltd, pp. 52-54.

Farmer, J.J. & Davis, B.R., (1985) H7 antiserum-sorbitol fermentation medium: a single tube screening medium for detecting *Escherichia coli* O157:H7 associated with hemorrhagic colitis. *Journal of clinical microbiology* **22**: 620-625.

Farmer, J.J., Fanning, G.R., Davis, B.R., O'hara, C.M., Riddle, C., Hickman-Brenner, F.W., Asbury, M.A., Lowery, V.A. & Brenner, D.J., (1985) *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of Enterobacteriaceae isolated from clinical specimens. *Journal of clinical microbiology* **21**: 77-81.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G., (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic acids research* **34**: e22.

Fegan, N., Barlow, R.S. & Gobius, K.S., (2006) *Escherichia coli* O157 somatic antigen is present in an isolate of *E. fergusonii*. *Current microbiology* **52**: 482-486.

Fenchel, T. & Finlay, B.J., (2006) *The diversity of microbes: resurgence of the phenotype*, Philosophical transactions of the Royal Society B-Biological Sciences **361**: 1965-1973.

Feng, P., Lampel, K.A., Karch, H. & Whittam, T.S., (1998) Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *Journal of infectious diseases* **177**: 1750-1753.

Fields, P.I., Blom, K., Hughes, H.J., Helsel, L.O., Feng, P. & Swaminathan, B., (1997) Molecular characterization of the gene encoding H antigen in *Escherichia coli* and development of a PCR-restriction fragment length polymorphism test for identification of *E. coli* O157: H7 and O157: NM. *Journal of clinical microbiology* **35**: 1066-1070.

Fitzhenry, R.J., Pickard, D.J., Hartland, E.L., Reece, S., Dougan, G., Phillips, A.D. & Frankel, G., (2002) Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic *Escherichia coli* O157:H7. *Gut* **50**: 180.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A. & Merrick, J.M., (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.

França, L.T.C., Carrilho, E. & Kist, T.B.L., (2002) A review of DNA sequencing techniques. *Quarterly reviews of biophysics* **35**: 169-200.

Franz, E., van Hoek, A.H.A.M., Bouw, E. & Aarts, H.J.M., (2011) Variability of *Escherichia coli* O157 strain survival in manure-amended soil in relation to strain origin, virulence profile, and carbon nutrition profile. *Applied and environmental microbiology* **77**: 8088-8096.

Fremaux, B., Prigent-Combaret, C., Delignette-Muller, M., Mallen, B., Dothal, M., Gleizal, A. & Vernozy-Rozand, C., (2008) Persistence of shiga toxin producing *Escherichia coli* O26 in various manure amended soil types. *Journal of applied microbiology* **104**: 296-304.

Friedrich, A.W., Bielaszewska, M., Zhang, W.-L., Pulz, M., Kuczius, T., Ammon, A. & Karch, H., (2002) *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *Journal of infectious diseases* **185**: 74-84.

Gaastra, W., Kusters, J.G., van Duijkeren, E. & Lipman, L.J.A., (2014) *Escherichia fergusonii*. *Veterinary microbiology* **172**: 7-12.

Gadiel, D. & Abelson, P., (2010) The economic cost of foodborne disease in New Zealand. In: New Zealand Food Safety Authority. Sydney, NSW, pp. 21-23.

Galiero, G., Conedera, G., Alfano, D. & Capriolo, A., (2005) Isolation of verocytotoxin-producing *Escherichia coli* O157 from water buffaloes (*Bubalus bubalis*) in southern Italy. *Veterinary record* **156**: 382-383.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I. & Koonin, E.V., (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic acids research* **43**: D261-D269.

Gill, C.O., (1998) Microbiological contamination of meat during slaughter and butchering of cattle, sheep and pigs. In: The microbiology of meat and poultry. A. Davies & R. Board (eds). London, UK: Blackie academic & professional, pp. 118-157.

Gogarten, J.P. & Townsend, J.P., (2005) Horizontal gene transfer, genome innovation and evolution. *Nature reviews microbiology* **3**: 679-687.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A. & Saier, M.H., (2004) Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *Journal of bacteriology* **186**: 3516-3524.

Gould, L.H., Bopp, C., Strockbine, N., Atkinson, R., Baselski, V., Body, B., Carey, R., Crandall, C., Hurd, S., Kaplan, R., Neill, M., Shea, S., Somsel, P., Tobin-D'Angelo, M., Griffin, P.M. & Gerner-Smidt, P., (2009) Recommendations for diagnosis of Shiga toxin-producing *Escherichia coli* infections by clinical laboratories. *Morbidity and mortality weekly report: recommendations and reports* **58**: 1-14.

Gould, L.H., Mody, R.K., Ong, K.L., Clogher, P., Cronquist, A.B., Garman, K.N., Lathrop, S., Medus, C., Spina, N.L., Webb, T.H., White, P.L., Wymore, K., Gierke, R.E., Mahon, B.E. & Griffin, P.M., (2013) Increased recognition of non-O157 Shiga toxin–producing *Escherichia coli* infections in the United States during 2000–2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne pathogens and disease* **10**: 453-460.

Griffin, P.M., Ostroff, S.M., Tauxe, R.V., Greene, K.D., Wells, J.G., Lewis, J.H. & Blake, P.A., (1988) Illnesses associated with *Escherichia coli* O157:H7 infections: a broad clinical spectrum. *Annals of internal medicine* **109**: 705-712.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C. & Gelbart, W.M., (2000) Genes, the environment, and the organism. In: An introduction to genetic analysis. 7th edition. W.H. Freeman (ed). New York.

Gunzer, F., Böhm, H., Rüssmann, H., Bitzan, M., Aleksic, S. & Karch, H., (1992) Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. *Journal of clinical microbiology* **30**: 1807-1810.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. & Shinagawa, H., (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA research* **8**: 11-22.

Hedican, E.B., Medus, C., Besser, J.M., Juni, B.A., Koziol, B., Taylor, C. & Smith, K.E., (2009) Characteristics of O157 versus non-O157 Shiga toxin-producing *Escherichia coli* infections in Minnesota, 2000–2006. *Clinical infectious diseases* **49**: 358-364.

Heuvelink, A.E., van den Biggelaar, F.L.A.M., Zwartkruis-Nahuis, J.T.M., Herbes, R.G., Huyben, R., Nagelkerke, N., Melchers, W.J.G., Monnens, L.A.H. & de Boer, E., (1998) Occurrence of verocytotoxin-producing *Escherichia coli* O157 on Dutch dairy farms. *Journal of clinical microbiology* **36**: 3480-3487.

HiMedia Laboratories, (2011) Andrade peptone water (M885). Mumbai, India: HiMedia Laboratories Pvt. Ltd., pp. 2.

Hiramatsu, R., Matsumoto, M., Miwa, Y., Suzuki, Y., Saito, M. & Miyazaki, Y., (2002) Characterization of Shiga toxin-producing *Escherichia coli* O26 strains and establishment of selective isolation media for these strains. *Journal of clinical microbiology* **40**: 922-925.

Hughes, J.M., Wilson, M.E., Johnson, K.E., Thorpe, C.M. & Sears, C.L., (2006) The emerging clinical importance of non-O157 Shiga toxin-producing *Escherichia coli*. *Clinical infectious diseases* **43**: 1587-1595.

Huson, D.H., (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**: 68-73.

Huson, D.H. & Bryant, D., (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**: 254-267.

Ihaka, R. & Gentleman, R., (1996) R: a language for data analysis and graphics. *Journal of computational and graphical statistics* **5**: 299-314.

Illumina Inc., (2014) Using a PhiX control for HiSeq® sequencing runs. In: Technical note: sequencing. San Diego, California, U.S.A.

Irshad, H., (2013) Molecular epidemiology of Shiga toxin-producing *Escherichia coli* (STEC) O157 and non-O157 STEC in calves in the North Island of New Zealand: a thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Massey University, Palmerston North, New Zealand.

Irshad, H., Cookson, A.L., Hotter, G., Besser, T.E., On, S.L.W. & French, N.P., (2011) Epidemiology of Shiga toxin-producing *Escherichia coli* O157 in very young calves in the North Island of New Zealand. *New Zealand veterinary journal* **60**: 21-26.

Jackson, M.P., Newland, J.W., Holmes, R.K. & O'Brien, A.D., (1987) Nucleotide sequence analysis of the structural genes for Shiga-like toxin I encoded by bacteriophage 933J from *Escherichia coli*. *Microbial pathogenesis* **2**: 147-153.

Jaros, P., (2014) Epidemiological investigations of Shiga toxin-producing *Escherichia coli* (STEC) O157 and STEC O26 in New Zealand slaughter cattle, and the source attribution of human illness: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Veterinary Science at Massey University, Palmerston North, New Zealand.

Jaros, P., Cookson, A.L., Campbell, D.M., Besser, T.E., Shringi, S., Mackereth, G.F., Lim, E., Lopez, L., Dufour, M., Marshall, J.C., Baker, M.G., Hathaway, S., Prattley, D.J. & French, N.P., (2013) A prospective case–control and molecular epidemiological study of human cases of Shiga toxin-producing *Escherichia coli* in New Zealand. *BMC infectious diseases* **13**: 450.

Jaros, P., Cookson, A.L., Reynolds, A., Prattley, D.J., Campbell, D.M., Hathaway, S. & French, N.P., (2016) Nationwide prevalence and risk factors for faecal carriage of *Escherichia coli* O157 and O26 in very young calves and adult cattle at slaughter in New Zealand. *Epidemiology and infection* **144**: 1736-1747.

Jelacic, J.K., Damrow, T., Chen, G.S., Jelacic, S., Bielaszewska, M., Ciol, M., Carvalho, H.M., Melton-Celsa, A.R., O'Brien, A.D. & Tarr, P.I., (2003) Shiga toxin–producing *Escherichia coli* in Montana: bacterial genotypes and clinical profiles. *Journal of infectious diseases* **188**: 719-729.

Jett, J.H., Keller, R.A., Martin, J.C., Marrone, B.L., Moyzis, R.K., Ratliff, R.L., Seitzinger, N.K., Shera, E.B. & Stewart, C.C., (1989) High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. *Journal of biomolecular structure and dynamics* **7**: 301-309.

Johnson, R.P., Clarke, R.C., Wilson, J.B., Read, S.C., Rahn, K., Renwick, S.A., Sandhu, K.A., Alves, D., Karmali, M.A., Lior, H., McEwen, S.A., Spika, J.S. & Gyles, C.L., (1996) Growing concerns and recent outbreaks involving non-O157:H7 serotypes of verotoxigenic *Escherichia coli*. *Journal of food protection* **59**: 1112-1122.

Jung, W.K., Bono, J.L., Clawson, M.L., Leopold, S.R., Shringi, S. & Besser, T.E., (2013) Lineage and genogroup-defining single nucleotide polymorphisms of *Escherichia coli* O157:H7. *Applied and environmental microbiology* **79**: 7036-7041.

Kalai Chelvam, K., Yap, K.P., Chai, L.C. & Thong, K.L., (2015) Variable responses to carbon utilization between planktonic and biofilm cells of a human carrier strain of *Salmonella enterica* serovar Typhi. *PLOS ONE* **10**: e0126207.

Kanehisa, M., (1995) KEGG: kyoto encyclopedia of genes and genomes.

Kaper, J.B., Nataro, J.P. & Mobley, H.L.T., (2004) Pathogenic *Escherichia coli*. *Nature reviews microbiology* **2**: 123-140.

Kaper, J.B. & O'Brien, A.D., (1998) *Escherichia coli O157: H7 and other Shiga toxin-producing E. coli strains*. American Society for Microbiology, Washington DC, USA.

Karch, H., Schubert, S., Zhang, D., Zhang, W., Schmidt, H., Ölschläger, T. & Hacker, J., (1999) A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Escherichia coli* clonal lineages. *Infection and immunity* **67**: 5994-6001.

Karmali, M., Petric, M., Steele, B. & Lim, C., (1983) Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools *Lancet* **321**: 619-620.

Keeling, P.J. & Palmer, J.D., (2008) Horizontal gene transfer in eukaryotic evolution. *Nature reviews genetics* **9**: 605-618.

Kilian, M. & Bülo, P., (1976) Rapid diagnosis of enterobacteriaceae. *Acta pathologica microbiologica scandinavica section B microbiology* **84**: 245-251.

Konopka, A.K. & Crabbe, M.J.C., (2004) *Compact handbook of computational biology*. CRC Press.

Kotewicz, M.L., Mammel, M.K., LeClerc, J.E. & Cebula, T.A., (2008) Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* **154**: 3518-3528.

Kotrba, P., Inui, M. & Yukawa, H., (2001) Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *Journal of bioscience and bioengineering* **92**: 502-517.

Kuhnert, P., Boerlin, P. & Frey, J., (2000) Target genes for virulence assessment of *Escherichia coli* isolates from water, food and the environment. *FEMS microbiology reviews* **24**: 107-117.

Kulkarni, S. & Pfeifer, J., (2014) Clinical genomics: a guide to clinical next generation sequencing. In. S. Kulkarni & J. Pfeifer (eds). Academic Press, pp. 11,299,300.

Kunau, W.H., Dommes, V. & Schulz, H., (1995) Beta-oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress. *Progress in lipid research* **34**: 267-342.

Laing, C.R., Zhang, Y., Gilmour, M.W., Allen, V., Johnson, R., Thomas, J.E. & Gannon, V.P.J., (2012) A comparison of Shiga-toxin 2 bacteriophage from classical enterohemorrhagic *Escherichia coli* serotypes and the German *E. coli* O104:H4 outbreak strain. *PLOS ONE* **7**: e37362.

Lapierre, P. & Gogarten, J.P., (2009) Estimating the size of the bacterial pan-genome. *Trends in genetics* **25**: 107-110.

Lawrence, J.G., (2005) Common themes in the genome strategies of pathogens. *Current opinion in genetics & development* **15**: 584-588.

Lenski, R.E., Rose, M.R., Simpson, S.C. & Tadler, S.C., (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American naturalist* **138**: 1315-1341.

Lerat, E. & Ochman, H., (2005) Recognizing the pseudogenes in bacterial genomes. *Nucleic acids research* **33**: 3125-3132.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S., (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li, L., Stoeckert, C.J. & Roos, D.S., (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* **13**: 2178-2189.

Lim, J., Hong, J., Sheng, H., Shringi, S., Kaul, R., Besser, T. & Hovde, C., (2010a) Phenotypic diversity of *Escherichia coli* O157:H7 strains associated with the plasmid O157. *Journal of microbiology* **48**: 347-357.

Lim, J.Y., Yoon, J. & Hovde, C.J., (2010b) A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *Journal of microbiology and biotechnology* **20**.

Louise, C.B. & Obrig, T.G., (1995) Specific interaction of *Escherichia coli* O157:H7-derived Shiga-like toxin II with human renal endothelial cells. *Journal of infectious diseases* **172**: 1397-1401.

Lukjancenko, O., Wassenaar, T.M. & Ussery, D.W., (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microbial ecology* **60**: 708-720.

Maas, W.K. & Davis, B.D., (1950) Pantothenate studies I: Interference by D-serine and L-aspartic acid with pantothenate synthesis in *Escherichia coli*. *Journal of bacteriology* **60**: 733.

MacFaddin, J.F., (1985) Media for isolation-cultivation identification-maintenance of medical bacteria. Baltimore, MD: Williams & Wilkins.

MacPherson, L., (2014) Overseas merchandise trade: August 2014. Statistics New Zealand.

Madic, J., Vingadassalon, N., de Garam, C.P., Marault, M., Scheutz, F., Brugère, H., Jamet, E. & Auvray, F., (2011) Detection of Shiga toxin-producing *Escherichia coli* serotypes O26:H11, O103:H2, O111:H8, O145:H28, and O157:H7 in raw-milk cheeses by using multiplex real-time PCR. *Applied and environmental microbiology* **77**: 2035-2041.

Manning, S.D., Motiwala, A.S., Springman, A.C., Qi, W., Lacher, D.W., Ouellette, L.M., Mladonicky, J.M., Somsel, P., Rudrik, J.T., Dietrich, S.E., Zhang, W., Swaminathan, B., Alland, D. & Whittam, T.S., (2008) Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proceedings of the National Academy of Sciences* **105**: 4868-4873.

March, S.B. & Ratnam, S., (1986) Sorbitol-MacConkey medium for detection of *Escherichia coli* O157: H7 associated with hemorrhagic colitis. *Journal of clinical microbiology* **23**: 869-872.

Mardis, E.R., (2008a) The impact of next-generation sequencing technology on genetics. *Trends in genetics* **24**: 133-141.

Mardis, E.R., (2008b) Next-generation DNA sequencing methods. *Annual review of genomics and human genetics* **9**: 387-402.

Martin-Gallardo, A., McCombie, W., Gocayne, J., FitzGerald, M., Wallace, S., Lee, B., Lamerdin, J., Trapp, S., Kelley, J. & Liu, L., (1992) Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13. 3. *Nature genetics* **1**: 34-39.

Maxam, A.M. & Gilbert, W., (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**: 560-564.

McClelland, M., Florea, L., Sanderson, K., Clifton, S.W., Parkhill, J., Churcher, C., Dougan, G., Wilson, R.K. & Miller, W., (2000) Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typhimurium, Typhi and Paratyphi. *Nucleic acids research* **28**: 4974-4986.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S. & Kaper, J.B., (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proceedings of the National Academy of Sciences* **92**: 1664-1668.

McFall, E., (1964) Genetic structure of the D-serine deaminase system of *Escherichia coli*. *Journal of molecular biology* **9**: 746-753.

McMaster, C., Roch, E., Willshaw, G., Doherty, A., Kinnear, W. & Cheasty, T., (2001) Verocytotoxin-producing *Escherichia coli* serotype O26: H11 outbreak in an Irish crèche. *European journal of clinical microbiology & infectious diseases* **20**: 430-432.

Mead, P.S. & Griffin, P.M., (1998) *Escherichia coli* O157:H7. *The Lancet* **352**: 1207-1212.

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M. & Tauxe, R.V., (1999) Food-related illness and death in the United States. *Emerging infectious diseases* **5**: 607-625.

Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R., (2005) The microbial pan-genome. *Current opinion in genetics & development* **15**: 589-594.

Mellor, G.E., Besser, T.E., Davis, M.A., Beavis, B., Jung, W., Smith, H.V., Jennison, A.V., Doyle, C.J., Chandry, P.S., Gobius, K.S. & Fegan, N., (2013) Multilocus genotype analysis of *Escherichia coli* O157 isolates from Australia and the United States provides evidence of geographic divergence. *Applied and environmental microbiology* **79**: 5050-5058.

Meng, J. & Schroeder, C.M., (2007) *Escherichia coli*. In: Foodborne diseases. S. Simjee (ed). Humana Press, Totowa, New Jersey, pp. 1-25.

MPI, (2015) Situation and outlook for primary industries 2015. Wellington, New Zealand: Ministry for Primary Industries, pp. 27-34.

MPI, (no date) National Microbiological Database (NMD) programme. Wellington, New Zealand: Ministry for Primary Industries, http://www.foodsafety.govt.nz/industry/general/nmd/, accessed October 2018.

Minoche, A.E., Dohm, J.C. & Himmelbauer, H., (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome biology* **12**: R112.

Montagne, L., Toullec, R. & Lalles, J., (2000) Calf intestinal mucin: isolation, partial characterization, and measurement in ileal digesta with an enzyme-linked immunosorbent assay. *Journal of dairy science* **83**: 507-517.

Mukherjee, A., Jackson, S.A., LeClerc, J.E. & Cebula, T.A., (2006) Exploring genotypic and phenotypic diversity of microbes using microarray approaches. *Toxicology mechanisms and methods* **16**: 121-128.

Muñoz-Elías, E.J. & McKinney, J.D., (2006) Carbon metabolism of intracellular bacteria. *Cellular microbiology* **8**: 10-22.

Natale, D.A., Shankavaram, U.T., Galperin, M.Y., Wolf, Y.I., Aravind, L. & Koonin, E.V., (2000) Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). *Genome biology* **1**: research0009.0001.

Nataro, J.P. & Kaper, J.B., (1998) Diarrheagenic *Escherichia coli*. *Clinical microbiology reviews* **11**: 142-201.

Neuwirth, E., (2011) RColorBrewer: ColorBrewer palettes. *R package version*: 1.0-5.

Nielsen, E.M., Skov, M.N., Madsen, J.J., Lodal, J., Jespersen, J.B. & Baggesen, D.L., (2004) Verocytotoxin-producing *Escherichia coli* in wild birds and rodents in close proximity to farms. *Applied and environmental microbiology* **70**: 6944-6947.

Nørregaard-Madsen, M., McFall, E. & Valentin-Hansen, P., (1995) Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus. *Journal of bacteriology* **177**: 6456-6461.

Nyrén, P. & Lundin, A., (1985) Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical biochemistry* **151**: 504-509.

O'Connor, K.E., Witholt, B. & Duetz, W., (2001) p-hydroxyphenylacetic acid metabolism in *Pseudomonas putida* F6. *Journal of bacteriology* **183**: 928-933.

Ochman, H., Lawrence, J.G. & Groisman, E.A., (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Oeggerli, M., (2009) *Escherichia coli* In: Visuals unlimited, Inc. 3018348.JPG (ed).

Ogden, I.D., Fenlon, D.R., Vinten, A.J.A. & Lewis, D., (2001) The fate of *Escherichia coli* O157 in soil and its potential to contaminate drinking water. *International journal of food microbiology* **66**: 111-117.

Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., Kodama, T., Abe, H., Nakayama, K., Kurokawa, K., Tobe, T., Hattori, M. & Hayashi, T., (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proceedings of the National Academy of Sciences* **106**: 17939-17944.

Osborne, C.A., Galic, M., Sangwan, P. & Janssen, P.H., (2005) PCR-generated artefact from 16S rRNA gene-specific primers. *FEMS microbiology letters* **248**: 183-187.

Pál, C., Papp, B. & Lercher, M.J., (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature genetics* **37**: 1372-1375.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E. & Blot, M., (1999) Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences* **96**: 3807-3812.

Parry, S.M. & Salmon, R.L., (1998) Sporadic STEC O157 infection: secondary household transmission in Wales. *Emerging infectious diseases* **4**: 657.

Perna, N.T., Plunkett, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J. & Kirkpatrick, H.A., (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157: H7. *Nature* **409**: 529-533.

Piggott, N.E. & Marsh, T.L., (2004) Does food safety information impact U.S. meat demand? *American journal of agricultural economics* **86**: 154-174.

Postma, P. & Lengeler, J., (1985) Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria. *Microbiological reviews* **49**: 232.

Postma, P.W., Lengeler, J.W. & Jacobson, G.R., (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiological reviews* **57**: 543-594.

Pradel, N., Bertin, Y., Martin, C. & Livrelli, V., (2008) Molecular analysis of Shiga toxin-producing *Escherichia coli* strains isolated from hemolytic-uremic syndrome patients and dairy samples in France. *Applied and environmental microbiology* **74**: 2118-2128.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. & Gu, Y., (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* **13**: 341.

Rahal, E.A., Kazzi, N., Nassar, F.J. & Matar, G.M., (2012) *Escherichia coli* O157:H7-clinical aspects and novel treatment approaches. *Frontiers in cellular and infection microbiology* **2**: 138.

Rangel, J.M., Sparling, P.H., Crowe, C., Griffin, P.M. & Swerdlow, D.L., (2005) Epidemiology of *Escherichia coli* O157: H7 outbreaks, United States, 1982–2002.

Rasko , D.A., Webster , D.R., Sahl , J.W., Bashir , A., Boisen , N., Scheutz , F., Paxinos , E.E., Sebra , R., Chin , C.-S., Iliopoulos , D., Klammer , A., Peluso , P., Lee , L., Kislyuk , A.O., Bullard , J., Kasarskis , A., Wang , S., Eid , J., Rank , D., Redman , J.C., Steyert , S.R., Frimodt-Møller , J., Struve , C., Petersen , A.M., Krogfelt , K.A., Nataro , J.P., Schadt , E.E. & Waldor , M.K., (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic–uremic syndrome in Germany. *New England journal of medicine* **365**: 709-717.

Reference laboratory for *Escherichia coli* (EcL), (2004) Pathogenic *E. coli*. Faculty of Veterinary Medicine, Université de Montréal, Canada.

Rice, E.W., Allen, M.J., Brenner, D.J. & Edberg, S.C., (1991) Assay for beta-glucuronidase in species of the genus *Escherichia* and its applications for drinking-water analysis. *Applied and environmental microbiology* **57**: 592-593.

Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., Blake, P.A. & Cohen, M.L., (1983) Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *New England journal of medicine* **308**: 681-685.

Rochan, M., (2013) India backs private investments in abattoirs. SupportBiz bureau, pp. 1.

Roesch, P.L., Redford, P., Batchelet, S., Moritz, R.L., Pellett, S., Haugen, B.J., Blattner, F.R. & Welch, R.A., (2003) Uropathogenic *Escherichia coli* use d-serine deaminase to modulate infection of the murine urinary tract. *Molecular microbiology* **49**: 55-67.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. & Nyrén, P., (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry* **242**: 84-89.

Rowen, L. & Koop, B.F., (1994) Zen and the art of large-scale genomic sequencing. In: Automatic DNA sequencing and analysis. M.D. Adams, C. Fields & J.C. Venter (eds). San Diego: Academic Press, pp. 167-174.

Russo, T.A. & Johnson, J.R., (2000) Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *Journal of infectious diseases* **181**: 1753-1754.

Ryan, C.A., Tauxe, R.V., Hosek, G.W., Wells, J.G., Stoesz, P.A., McFadden, H.W., Smith, P.W., Wright, G.F. & Blake, P.A., (1986) *Escherichia coli* O157: H7 diarrhea in a nursing home: clinical, epidemiological, and pathological findings. *Journal of infectious diseases* **154**: 631-638.

Sabri, S., Nielsen, L.K. & Vickers, C.E., (2013) Molecular control of sucrose utilization in *Escherichia coli* W, an efficient sucrose-utilizing strain. *Applied and environmental microbiology* **79**: 478-487.

Saier Jr, M.H., Chauvaux, S., Deutscher, J., Reizer, J. & Ye, J.-J., (1995) Protein phosphorylation and regulation of carbon metabolism in Gram-negative versus Gram-positive bacteria. *Trends in biochemical sciences* **20**: 267-271.

Saier, M.H., (1998) Multiple mechanisms controlling carbon metabolism in bacteria. *Biotechnology and bioengineering* **58**: 170-174.

Sanger, F. & Coulson, A.R., (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94**: 441-448.

Sanger, F., Nicklen, S. & Coulson, A.R., (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**: 5463-5467.

Sarkar, D., (2008) Lattice: multivariate data visualization with R. Springer, New York.

Schell, M.J., (2004) The *N*–methyl D–aspartate receptor glycine site and D–serine metabolism: an evolutionary perspective. *Philosophical transactions of the Royal Society of London. Series B: biological sciences* **359**: 943-964.

Schmidt, H., Beutin, L. & Karch, H., (1995) Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. *Infection and immunity* **63**: 1055-1061.

Schubert, S., Rakin, A., Karch, H., Carniel, E. & Heesemann, J., (1998) Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infection and immunity* **66**: 480-485.

Scotland, S.M., Willshaw, G.A., Smith, H.R. & Rowe, B., (1990) Properties of strains of *Escherichia coli* O26:H11 in relation to their enteropathogenic or enterohemorrhagic classification. *Journal of infectious diseases* **162**: 1069-1074.

Seemann, T., (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*: btu153.

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N. & Whittam, T.S., (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and environmental microbiology* **51**: 873-884.

Shaikh, N. & Tarr, P.I., (2003) *Escherichia coli* O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications. *Journal of bacteriology* **185**: 3596-3605.

Shringi, S., Schmidt, C., Katherine, K., Brayton, K.A., Hancock, D.D. & Besser, T.E., (2012) Carriage of *stx2a* differentiates clinical and bovine-biased strains of *Escherichia coli* O157. *PLOS ONE* **7**.

Shringi, S., Schmidt, C., Katherine, K., Brayton, K.A., Hancock, D.D. & Besser, T.E., (2012, pp. 4) Carriage of *stx2a* differentiates clinical and bovine-biased strains of *Escherichia coli* O157. *PLOS ONE* **7**.

Sigma-Aldrich, (no date) Andrade peptone water (A0715). St. Louis, MO, USA: Sigma-Aldrich Co. LLC., pp. 1.

Smith, J.L., (1998) Foodborne illness in the elderly. *Journal of food protection* **61**: 1229-1239.

Spickler, A.R., Roth, J.A., Galyon, J. & Lofstedt, J., (2010) *Emerging and exotic diseases of animals*. Center for Food Security and Public Health (CFSPH) Iowa State University.

Stager, C.E. & Davis, J.R., (1992) Automated systems for identification of microorganisms. *Clinical microbiology reviews* **5**: 302-327.

Statistics New Zealand, (2013) Global New Zealand – International trade, investment, and travel profile: Year ended December 2013. Wellington: Ministry of Foreign Affairs and Trade and Statistics New Zealand.

Stevens, M.P., van Diemen, P.M., Dziva, F., Jones, P.W. & Wallis, T.S., (2002) Options for the control of enterohaemorrhagic *Escherichia coli* in ruminants. *Microbiology* **148**: 3767-3778.

Strachan, N.J.C., Fenlon, D.R. & Ogden, I.D., (2001) Modelling the vector pathway and infection of humans in an environmental outbreak of *Escherichia coli* O157. *FEMS microbiology letters* **203**: 69-73.

Studier, F.W., (1989) A strategy for high-volume sequencing of cosmid DNAs: random and directed priming with a library of oligonucleotides. *Proceedings of the National Academy of Sciences* **86**: 6917-6921.

Stülke, J. & Hillen, W., (1999) Carbon catabolite repression in bacteria. *Current opinion in microbiology* **2**: 195-201.

Sussman, M., (1985) Theodor Escherich (1857-1911): a biographical note. In: The virulence of *Escherichia coli*: reviews and methods. M. Sussman (ed). London: Academic Press.

Swerdlow, D.L., Woodruff, B.A., Brady, R.C., Griffin, P.M., Tippen, S., Donnell, J.H.D., Geldreich, E., Payne, B.J., Meyer, J.A., Wells, J.G., Greene, K.D., Bright, M., Bean, N.H. & Blake, P.A., (1992) A waterborne outbreak in Missouri of *Escherichia coli* O157:H7 associated with bloody diarrhea and death. *Annals of Internal Medicine* **117**: 812-819.

Tarr, P.I. & Neill, M.A., (1996) Perspective: the problem of non-O157:H7 Shiga toxin (verocytotoxin)-producing *Escherichia coli*. *The Journal of infectious diseases* **174**: 1136-1139.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. & Koonin, E.V., (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic acids research* **29**: 22-28.

Tauxe, R.V., (2002) Emerging foodborne pathogens. *International Journal of Food Microbiology* **78**: 31-41.

Textor, S., Wendisch, V.F., De Graaf, A.A., Muller, U., Linder, M.I., Linder, D. & Buckel, W., (1997) Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. *Archives of microbiology* **168**: 428-436.

Tilden Jr, J., Young, W., McNamara, A.-M., Custer, C., Boesel, B., Lambert-Fair, M.A., Majkowski, J., Vugia, D., Werner, S. & Hollingsworth, J., (1996) A new route of transmission for *Escherichia coli*: infection from dry fermented salami. *American journal of public health* **86**: 1142-1145.

Trevisani, M., Mancusi, R., Riu, R. & Serraino, A., (2013) Quantitative detection of Shiga toxin-producing and enteropathogenic *Escherichia coli* serotypes O157 and O26 in bulk raw milk. *Food analytical methods* **6**: 1750-1758.

Tso, W.-W. & Adler, J., (1974) Negative chemotaxis in *Escherichia coli*. *Journal of Bacteriology* **118**: 560-576.

Turcatti, G., Romieu, A., Fedurco, M. & Tairi, A.-P., (2008) A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic acids research* **36**: e25.

Vaas, L.A.I., Sikorski, J., Michael, V., Göker, M. & Klenk, H.P., (2012) Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLOS ONE* **7**: e34846.

van Diemen, P.M., Dziva, F., Stevens, M.P. & Wallis, T.S., (2005) Identification of enterohemorrhagic *Escherichia coli* O26: H− genes required for intestinal colonization in calves. *Infection and immunity* **73**: 1735-1743.

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. & Swings, J., (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological reviews* **60**: 407-438.

Venter, J.C., Smith, H.O. & Hood, L., (1996) A new strategy for genome sequencing. *Nature* **381**: 364-366.

Virdi, J. & Sachdeva, P., (2005) Genetic diversity of pathogenic microorganisms: basic insights, public health implications and the Indian initiatives. *Current science* **89**.

Warnes, G.R., Bolker, B., Bonebakker, L., Huber, W., Liaw, A., Lumley, T., Magnusson, A., Moeller, S., Schwartz, M. & Venables, B., (2013) gplots: various R programming tools for plotting data. 2013. *R package version* **2**.

Werber, D., Beutin, L., Pichner, R., Stark, K. & Fruth, A., (2008) Shiga toxin–producing *Escherichia coli* serogroups in food and patients, Germany. *Emerging infectious diseases* **14**: 1803-1806.

Werber, D., Fruth, A., Heissenhuber, A., Wildner, M., Prager, R., Tschäpe, H. & Ammon, A., (2004) Shiga toxin-producing *Escherichia coli* O157 more frequently cause bloody diarrhea than do non-O157 strains. *Journal of infectious diseases* **189**: 1335-1336.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.-z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A. & Rothberg, J.M., (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**: 872-876.

Whittam, T.S., Wachsmuth, I.K. & Wilson, R.A., (1988) Genetic evidence of clonal descent of *Escherichia coli* O157:H7 associated with hemorrhagic colitis and hemolytic uremic syndrome. *Journal of infectious diseases* **157**: 1124-1133.

Whittam, T.S., Wolfe, M.L., Wachsmuth, I.K., Orskov, F., Orskov, I. & Wilson, R.A., (1993) Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infection and immunity* **61**: 1619-1629.

Wick, L.M., Qi, W., Lacher, D.W. & Whittam, T.S., (2005) Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *Journal of bacteriology* **187**: 1783-1791.

Wickham, M.E., Brown, N.F., Boyle, E.C., Coombes, B.K. & Finlay, B.B., (2007) Virulence is positively selected by transmission success between mammalian hosts. *Current biology* **17**: 783-788.

Wieler, L.H., McDaniel, T.K., Whittam, T.S. & Kaper, J.B., (1997) Insertion site of the locus of enterocyte effacement in enteropathogenic and enterohemorrhagic *Escherichia coli* differs in relation to the clonal phylogeny of the strains. *FEMS microbiology letters* **156**: 49-53.

Wright, J., Fraser, D. & Baker, M., (1993) *Escherichia coli* O157: H7 infection: first New Zealand case report. *Communicable diseases New Zealand* **93**: 113-116.

Wright, S., (1982) Character change, speciation, and the higher taxa. *Evolution*: 427-443.

Zaw, M.T., Yamasaki, E., Yamamoto, S., Nair, G.B., Kawamoto, K. & Kurazono, H., (2013) Uropathogenic specific protein gene, highly distributed in extraintestinal uropathogenic *Escherichia coli*, encodes a new member of HNH nuclease superfamily. *Gut pathogens* **5**: 13.

Zerbino, D.R. & Birney, E., (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome research* **18**: 821-829.

Zhang, Y., Laing, C., Steele, M., Ziebell, K., Johnson, R., Benson, A., Taboada, E. & Gannon, V., (2007) Genome evolution in major *Escherichia coli* O157: H7 lineages. *BMC genomics* **8**: 121.

Zimbro, M.J., Power, D.A., Miller, S.M., Wilson, G.E. & Johnson, J.A., (2009) Difco™ & BBL™ manual: manual of microbiological culture media. In: Second Edition. BD Diagnostics (ed). Sparks, Maryland: Becton, Dickinson and Company, pp. 328-331.

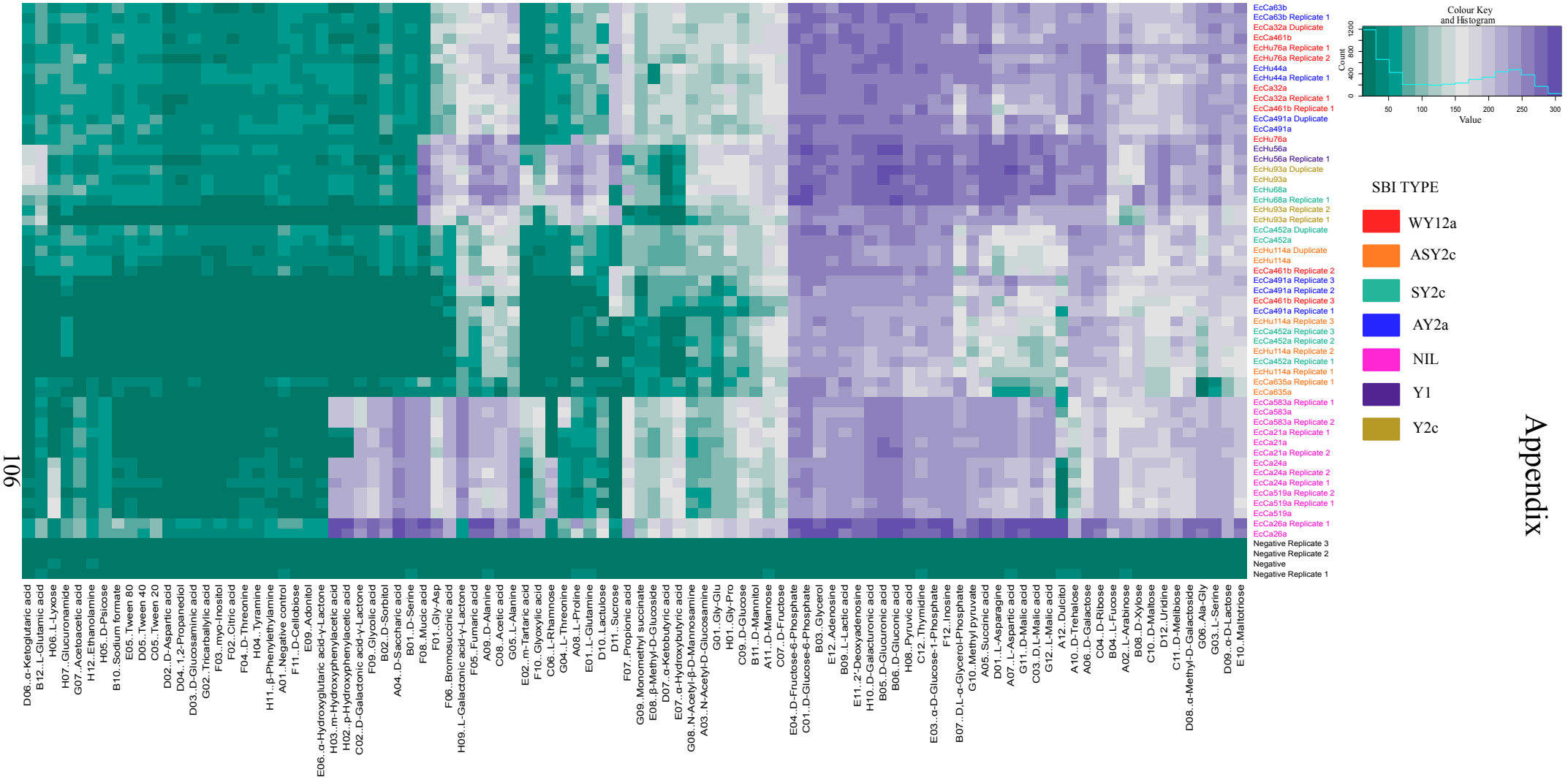

Figure A1. Heatmap of *E. coli* O157 respiration changes on PM1 including duplicate and replicate assays. The heatmap illustrates the cell respiration of 96 phenotypes by 17 *E. coli* O157 isolates. The purple shaded areas indicate better respiration; the grey areas indicate minimal respiration and the green areas indicate no respiration of the isolates. The column on the right represents the isolates tested for respiration and the row at the bottom represents the 96 wells of PM1 (A01 through H12). The graph on the right includes the colour key and histogram of the heatmap and the key on the right indicates the SBI types of the various *E. coli* O157 isolates.

Figure A2. Heatmap of *E. coli* O157 respiration changes on PM2A including duplicate and replicate assays. The heatmap illustrates the cell respiration of 96 phenotypes by 17 *E. coli* O157 isolates. The purple shaded areas indicate better respiration; the grey areas indicate minimal respiration and the green areas indicate no respiration of the isolates. The column on the right represents the isolates tested for respiration and the row at the bottom represents the 96 wells of PM2A (A01 through H12). The graph on the right includes the colour key and histogram of the heatmap and the key on the right indicates the SBI types of the various *E. coli* O157 isolates.

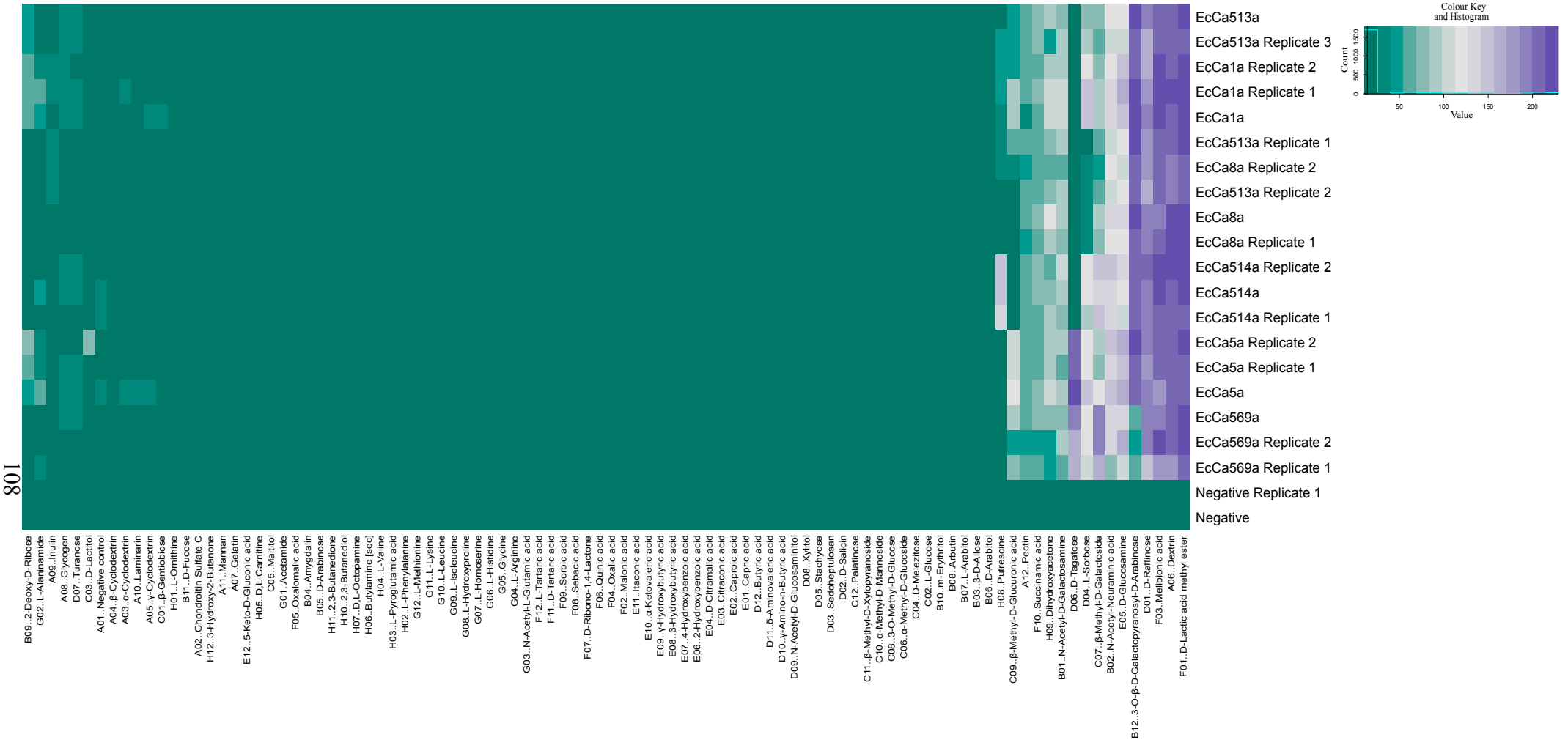

Figure A3. Heatmap of *E. coli* O26 respiration changes on PM2A. The heatmap illustrates the cell respiration of 96 phenotypes by six *E. coli* O26 isolates. The purple areas indicate better respiration, the grey areas indicate minimal respiration and the green areas indicate no respiration of the isolates. The column on the right represents the isolates tested for respiration and the row at the bottom represents the 96 wells of PM2A (A01 to H12). The graph on the right includes the colour key and histogram of the heatmap.