Shiga toxin encoding bacteriophages and horizontal gene transfer in *Escherichia coli* O157

Eby Mazini Sim

A thesis submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy

July 2016

Certificate of Authorship/Originality

I certify that the work presented in this thesis has not previously been submitted for a degree, nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text.

I also certify that the written preparation of the thesis, and all experimental work associated with it, has been carried out solely by me, unless otherwise indicated.

Finally, I certify that all information sources and literature used are acknowledged in the text.

_____

Eby Mazini Sim

Acknowledgements

My journey as a PhD student, starting from The University of Queensland, Brisbane and the subsequent relocation to The University of Technology, Sydney, has been challenging, exciting and immensely rewarding. However, this thesis would not have been possible without the support and guidance from a number of amazing people along the way. First and foremost, I would like to acknowledge my principal supervisor Dr. Nico Petty. Nico not only supervised the work presented in this thesis and provided feedback on this thesis, she also guided and encouraged me through these years. Her passion for research is truly inspiring. I would also like to acknowledge my co-supervisor Dr. Kari Gobius. In spite of his responsibilities at the CSIRO, Kari always made time to mentor me. Both of my supervisors contributed enormously to my development as a scientist and I couldn't have asked for a better supervisory team for my PhD journey.

A huge thank you to all past and present members of the Petty lab and colleagues in the CSIRO, who have made this journey an amazing and productive one: Glen Mellor, Robert Barlow, Lesley Duffy, Kate McMillian, Salma Lahjar, Tuan Nyguen, Neralle Fegan, Scott Chandry, Leanne Haggerty, Alejandro Marquez, Lucy Martindale and Dylan Mansfield.

To my family, especially my Mum, thank you for your words of support over all these years. I am particularly grateful for your encouragement of my PhD work as well as my aspirations. To Melissa Hamwood, thank you for being there for me in both the good times and the hard times and for supporting me through this journey.

Lastly, I would like to acknowledge the financial assistance from the UQ International Scholarship awarded during my stay at UQ, the UTS International Research Scholarship awarded during my stay at UTS as well as funding from the CSIRO.

Content page

Certificate of Authorship/Originality ..... **i**

Acknowledgements ..... **ii**

Content page ..... **iii**

List of Figures ..... **vii**

List of Tables ..... **ix**

Abbreviations ..... **x**

Research outputs ..... **xii**

Publications xii
Conference proceedings xii
Book Chapter xii

Abstract ..... **xiii**

1. Introduction **1**

1.1. Bacteriophages 1
1.1.1 The structure of tailed phages 2
1.1.2. Phage genomes 4
1.1.3. Phage life cycle 5
1.1.3.1 Lytic life cycle 6
1.1.3.2. Lysogenic life cycle 7
1.1.3.3. Pseudolysogeny 9
1.1.3.4. Satellite phages- a subset of temperate phages 10
1.1.4 Coevolution of phages and bacteria 10
1.1.4.1. Inhibition of phage adsorption 10
1.1.4.2. Preventing DNA entry 11
1.1.4.3. CRISPR-Cas system 11
1.1.4.4. Restriction-modification systems 12
1.1.4.5. Abortive infection 12
1.1.5. Phages and horizontal gene transfer 13
1.1.5.1 Transduction 14
1.1.5.1.1. Generalised transduction 14
1.1.5.1.2. Specialised transduction 15
1.1.5.2. Lysogenic conversion 16
1.2. *Escherichia coli* 17
1.3. *Escherichia coli* O157 18
1.3.1. Transmission of STEC O157 19
1.3.2. Geographical variation in STEC O157 related disease 20
1.3.3. Diversity of STEC O157 21
1.3.4. Virulence factors of STEC O157 strains 23
1.3.4.1. The STEC O157 specific plasmid pO157 23
1.3.4.2 Locus of Enterocyte Effacement pathogenicity island 24
1.3.4.3. Shiga toxin 25
1.4. Shiga toxin-encoding phages 27
1.4.1. Stx phage genomes 27
1.4.2. Horizontal gene transfer and Stx phages 32

1.4.3. Induction of Stx phages 33
1.4.4. Persistence in the environment 33
1.4.5. Stx phage host range 33
1.4.6. The impact of Stx prophages on their bacterial host....36
1.4.7. Insertion sites of Stx phages 37
1.5. Australian STEC O157 and Stx phages 38
1.6. Objectives of this study....39

2. Materials and methods 40

2.1. Media, supplements and solutions 40
2.2. Bacterial strains, phages and culture conditions 41
2.3. Bacterial phenotype test....44
2.3.1. Differentiation of STEC O157 from other bacteria 44
2.3.2. Stx production....44
2.4. Phage methods 44
2.4.1. Isolation of Stx phages....44
2.4.2. Phage concentration and purification....45
2.4.2.1. Isopycnic ultracentrifugation 45
2.4.2.1. Ultrafiltration 45
2.4.3. Phage quantification....46
2.4.3.1. Double agar overlay plaque assay....46
2.4.3.1. Semi quantitative spot test 46
2.4.4 Test for lysogeny....46
2.4.4.1. Lysogenic conversion of bacterial strains on solid media 46
2.4.4.2. Lysogenic conversion of bacterial strains in liquid culture 47
2.4.5. Transduction assay....47
2.4.6. Phage electron microscopy 48
2.5. DNA work....48
2.5.1. Primers 48
2.5.2. PCR....51
2.5.2.1. Colony PCR 52
2.5.2.2. Quantitative PCR 52
2.5.2.4. Purification of PCR products....53
2.5.3. Gel electrophoresis....53
2.5.4. DNA extraction....53
2.5.5. DNA sequencing....54
2.6. Bioinformatics methods....54
2.6.1. Sequence quality control....54
2.6.2. Genome assembly 54
2.6.4. Genome annotation 55
2.6.5. Comparative genomic analyses....55

3. Characterisation of Stx1 phages from Australian *E. coli* O157 clinical isolates 56

3.1. Introduction....56
3.2. Results....56
3.2.1. Host genetic background affects Shiga toxin production 56
3.2.2. Stx1 phages belong to the *Podoviridae* family....58
3.2.3. Assembly of the Stx1 phage genomes 59

3.2.4. The genomes of phages AU5Stx1 and AU6Stx1 62
3.2.5. The genomes of AU5Stx1 and AU6Stx1 are highly similar to each other and to Stx2a phages 64
3.2.6. Integration of the Stx1 phages occurs at the 5' end of tRNA-*argW* 70
3.2.7. Host recognition by the Stx1 phages 72
3.2.8. Possible site specific insertion of the *stx* operon into Stx phages 79
3.3. Discussion 81
3.3.1. Both host and phage genetic background influences Stx1 production. 82
3.3.2. Structure of the Podovirus Stx phage 82
3.3.3. First mention of Stx phages integrating into the 5' end of a tRNA 84
3.3.4. The possibility for alternative tail fibres as an explanation of broad host range 85
3.3.5. Site-specific insertion of the *stx* operon 87
3.4. Conclusion 89

4. Characterisation of Stx prophages in non-clinical Australian isolates of *E. coli* O157 91

4.1. Introduction 91
4.2. Results 91
4.2.1. Variation in toxin and phage production in the three STEC O157 strains 91
4.2.2. Sequencing and assembly of the Stx prophages of EC14, EC571 and EC2441 93
4.2.3. Comparison of the Stx prophages of EC14, EC571 and EC2441 94
4.2.4. Comparison of the Stx1 prophages of EC14 and EC571 to known Stx phages 98
4.2.5. Comparison of the Stx2c prophages of EC14, EC571 and EC2441 to known Stx phage genomes 102
4.2.6. Translational frameshifting predicted within tail genes 105
4.2.7. The insertion site of the Stx2c prophages 109
4.2.8. DNA replication systems and predicted DNA packaging of the Stx2c prophages 111
4.2.9. The three Stx2c prophages carry a cargo gene which provides adaptation to magnesium starvation 112
4.2.10. The Stx2c prophages encode an anti-repression operon 112
4.3. Discussion 115
4.3.1. A dominant inducible Stx prophage within a bacterial strain harbouring two Stx prophages 115
4.3.2. Disruptions by IS elements contribute to variations in toxin production 117
4.3.3. DNA packaging and tail formation in the Stx2c prophages 118
4.4. Conclusion 118

5. Mobilisation of the Locus of Enterocyte Effacement by an Stx2a prophage 120

5.1. Introduction 120
5.2. Results 120
5.2.1. Viability of phages induced from EC1810 120
5.2.2. Morphology of the Stx2a prophage of EC1810 121

5.2.3. Host range of 1810Stx2a....123
5.2.4. Transduction of Δ*eae*::*cat* to recipient *E. coli* strains....124
5.2.5. Presence of 1810Stx2a is required for further dissemination of Δ*eae*::*cat* 127
5.2.6. Optimisation of transduction protocol to determine if transduction into non-STEC O157 strains could occur 129
5.2.7. Repeat of transduction using optimised transduction protocol....131
5.2.8. Genome assembly of STEC O157 strain FNS9, an attenuated derivative of EC1810 132
5.2.9. Genomic analysis of pp1810Stx2K 133
5.3. Discussion 138
5.3.1. A new DNA replication system for Stx phages 138
5.3.2. pp1810Stx2k carries genes that enhances its chances of infection....139
5.3.3. 1810Stx2a is likely to conform to the superinfection immunity paradigm 140
5.3.4. Evidence for transduction of the LEE....140
5.3.5. Mechanism of LEE mobilisation 141
5.4 Conclusions....144

6. Final discussion 145

6.1. Insights into the reduced virulence of Australian STEC O157....145
6.2. Broad host range of Stx phages leads to the potential of new Shiga toxin producing pathogens 148
6.3. Horizontal gene transfer: the possibility of more than just *stx* and phage encoded cargo genes 150
6.4. MITES as a possible cargo gene acquisition mechanism 152
6.5. Conclusions....153

References....155

Appendices....184

Appendix I: Functional annotation of AU5Stx1 184
Appendix II: Functional annotation of AU6Stx1....187
Appendix III: Functional annotation of pp14Stx1 189
Appendix IV: Functional annotation of pp571Stx1....191
Appendix V: Functional annotation of pp14Stx2c 193
Appendix VI: Functional annotation of pp571Stx2c 195
Appendix VII: Functional annotation of pp2441Stx2c....197
Appendix VIII: Functional annotation of pp1810Stx2k 199

List of Figures

Figure 1.1: Diversity of phages.....2
Figure 1.2: Phage life cycles.....5
Figure 1.3: Integration and excision of prophage.....8
Figure 1.4: Mechanisms of horizontal gene transfer in bacteria.....13
Figure 1.5: Overview of generalised transduction.....15
Figure 1.6: Overview of specialised transduction.....16
Figure 1.7: Annual incidence of STEC O157 related diseases.....20
Figure 1.8: Stepwise evolution model of STEC O157.....22
Figure 1.9: Graphical representation of the LEE pathogenicity island of STEC O157 strain Sakai.....25
Figure 1.10: Organisation of the functional modules in the genomes of phage lambda and Stx phages.....28

Figure 3.1: Comparison of Stx phage induction and Stx production in STEC O157 and K-12 lysogens.....57
Figure 3.2: Morphology of the two Stx1 phages.....59
Figure 3.3: Read coverage over assembly.....60
Figure 3.4: Graphical visualisation of the 454ContigGraph.txt output file generated by Newbler.....61
Figure 3.5: Stx1 phage genomes represented as linear genome in its prophage orientation.....63
Figure 3.6: Pairwise comparison between AU5Stx1 and AU6Stx1.....66
Figure 3.7: Comparison of Stx1 phages to published phages.....68
Figure 3.8: *att* sequence of the Stx1 phages AU5Stx1 and AU6Stx1.....72
Figure 3.9: Comparison of the tail tip protein of phage lambda, phage N15 and AU5Stx1.....73
Figure 3.10: Signatures for +1 translational frameshift in tail fibre formation in the majority of known Stx phages.....75
Figure 3.11: Subtypes of (A) cterm1, (B) cterm2 and (C) Pseudoknot sequences.....77
Figure 3.12: Possible site specific insertion of the *stx* gene in the Stx phages.....81
Figure 3.13: Proposed structure of the Podovirus Stx phage.....84

Figure 4.1: Comparison of Stx production and selected gene copy numbers in three Australian STEC O157 strains.....92
Figure 4.2: Pairwise comparison of the Stx prophages.....95
Figure 4.3: Genome comparison of the Stx1 prophages to other Stx phages.....100
Figure 4.4: Genomic comparison of the Stx2c prophages to other Stx phages.....104
Figure 4.5: Comparison of the tail morphogenesis genes of pp571Stx2c, HK97 and lambda.....105
Figure 4.6: Multiple sequence alignment of cterm1 subtypes.....106
Figure 4.7: Sequences encoding tail fibres with alternative tips in Stx2c prophages.....107
Figure 4.8: The genetic region encoding Stx2c phage tail assembly.....108
Figure 4.9: Multiple amino acid sequence alignment of the integrases of the Stx2c prophages.....109
Figure 4.10: Schematic representation of the insertion site of Stx2c prophages.....110

Figure 4.11: Stx2c prophages are the only Stx phages which possess an anti-repression operon. 113
Figure 4.12: Comparison between the Stx1 prophages of EC3206 and EC14. 115

Figure 5.1: Investigation of 1810Stx2a stability over time, in various storage conditions. 121
Figure 5.2: Morphology of 1810Stx2a 122
Figure 5.3: Histogram showing the measurements of phage virions viewed under the TEM. 123
Figure 5.4: 1810Stx2a plaque formation on indicator strains. 124
Figure 5.5: Overview of transduction experiment. 126
Figure 5.6: PCR analysis of donor, recipients and transductants. 126
Figure 5.7: Detection of EC1810 cryptic prophage by PCR. 128
Figure 5.8: Schematic representation of transduction experiment to determine if 1810Stx2 is required for transduction. 129
Figure 5.9: Effect of diavlaent ion addition on rate of transduction. 130
Figure 5.10: Repeat of transduction experiments using optimised conditions. 132
Figure 5.11: Comparison of the chromosomes of STEC O157 strains FNS9 and Sakai. 133
Figure 5.12: Genome of pp1810Stx2K. 135
Figure 5.13: Protein folding prediction of the three unique genes carried by pp1810Stx2K. 138

List of Tables

Table 1.1: Completely sequenced Stx phage genomes and Stx prophages (in complete bacterial genomes) available in the public databases (GenBank/EMBL/DDBJ). 30
Table 1.2: Host range studies of Stx phages 34

Table 2.1: Media and solutions 40
Table 2.2: Bacterial strains used in this study 41
Table 2.3: Antibiotics 43
Table 2.4: Primers used in this study 49

Table 3.1: Sequencing and assembly statistics for phages AU5Stx1 and AU6Stx1. ..59
Table 3.2: Summary of the annotated phage genomes of AU5Stx1 and AU6Stx1. 62
Table 3.3: Subtypes of Stx phage tail fibre C terminal domains and pseudoknots. 77
Table 3.4: Variant C terminal domains and pseudoknots in the genomes of *Citrobacter rodentium* ICC168 and *Escherichia albertii* KF1 79

Table 4.1: Genome assembly statistics for the STEC O157 strains EC14, EC571 and EC2441 93
Table 4.2: Mean stx_1 gene copy number in inducing and non-inducing conditions. ...96
Table 4.3: *Enterobactericeae* genomes that harbour IS*Ec20*. 101
Table 4.4: Variant C terminal domains and pseudoknots encoded in the tail fibre genes of the five Stx prophages. 106

Table 5.1: Susceptibility of bacterial strains to 1810Stx2K 124
Table 5.2: Frequency of transduction (transductants per pfu) of Δ*eae*::*cat* from FNS8 to FNS9 with varied MOI and cell numbers. 131
Table 5.3: Genome assembly statistics of STEC O157 strain FNS9. 133

Abbreviations

Abbreviation	Meaning
att	Attachment site
bp	Base pair
BLAST	Basic Local Alignment Search Tool
CDS	Coding sequence
Cm	Chloramphenicol
Cfu	Colony forming unit
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DNA	Deoxyribonucleic acid
DNTP	Deoxynucleotide
ELISA	Enzyme-linked immunosorbent assay
G+C	Guanine+Cytosine
gDNA	Genomic DNA
HGT	Horizontal gene transfer
HUS	Haemolytic uremic syndrome
IR	Inverted repeat
IS	Insertion sequnce
Km	Kanamycin
kb	Kilo base
L	litre
LB	Luria broth
LEE	Locus of Enterocyte Effacement
M	Molar
mM	Millimolar
Mb	Mega base
ml	Millilitre
µl	Mircolitre
MITE	Miniature Inverted-repeat Transposition Element
MLEE	Multi-Locus Enzyme Electrophoresis
MLVA	Multiple- Locus Variable Number Tandem Repeat
MMC	Mitomycin C
MOI	Multiplicity of infection
nm	Nanometre
NCBI	National Centre for Biotechnology information
PCR	Polymerase chain reaction
Phage	Bacteriophage
pfu	Plaque forming unit
qPCR	Quantitative Polymerase chain reaction
RNA	Ribonucleic acid
R-M	Restriction modification
Rif	Rifampicin
Sm	Streptomycin
SMRT	Single Molecule Real Time
SNP	Single Nucleotide Polymorphism
Stx	Shiga toxin
STEC	Shiga toxigenic *E. coli*

Abbreviation	Meaning
T3SS	Type 3 Secretion System
TA	Toxin-antitoxin
TEM	Transmission electron microscopy
tRNA	Transfer ribonucleic acid
WT	Wildtype

Research outputs

Publications

Sim, E. M., P. S. Chandry, S. A. Beatson, K. S. Gobius & N. K. Petty (2016). Insights into the emergence of pathogens by genomic analysis of Shiga toxin 1 encoding bacteriophages originating from clinical Australian *Escherichia* coli O157 isolates (Manuscript in preparation).

Sim, E. M., K. S. Gobius & N. K. Petty (2016). Insights into the regulation of Australian Stx1 and Stx2c prophages in STEC O157 (Manuscript in preparation)

Conference proceedings

Sim, E.M., P. S. Chandry, S. A. Beatson, K. S. Gobius & N. K. Petty (January 2015) **POSTER.** Characterisation of Shiga Toxin 1 phages reveal insights for formation of new pathogens. *Bacteriophages 2015*, London, United Kingdom

Sim, E.M., P. S. Chandry, S. A. Beatson, N. K. Petty & K. S. Gobius (September 2013) **ORAL PRESENTATION.** Characterisation of novel Shiga toxin 1 encoding bacteriophages originating from Australian *E. coli* O157 isolates**.** *BacPath 12,* Brisbane, Australia.

Sim, E.M., N. K. Petty, T. J. Wells, M. Leitner, M. J. Sullivan, S. A. Beatson & K. S. Gobius (July 2012) **POSTER.** Phage mediated transfer of *Escherichia coli* O157 initimin gene (*eae*) encoded by the Locus of Enterocyte Effacement, *40th Annual Meeting of the Australian Society of Microbiology*, Brisbane, Australia.

Book Chapter

Turner, D., J. M. Sutton, D. M. Reynolds, **E. M. Sim** & N. K. Petty (2015) Visualisation of phage genomic data: comparative genomics and publication-quality diagrams. In *Bacteriophages: Methods & Protocols*. A. M. Kropinski & M. R. J. Clokie (eds). Humana Press (Submitted).

Abstract

Shiga toxigenic *Escherichia coli* (STEC), most notably serotype O157, is a food borne pathogen of global public health concern. The progression to a severe disease state following an STEC infection is associated with the production of Shiga toxin (Stx). The ability to produce Stx is conferred upon STEC strains by Stx–encoding bacteriophages (Stx phages), which infect and integrate into the host bacterial genome. These phages carry either stx_1 or stx_2 genes which encode two immunologically distinct toxins with similar biological functions. However, not all STEC O157 strains are equally pathogenic as Australian STEC O157 strains, are associated with lower incidence of clinical disease than strains from other countries. This led to the aim of this thesis, which was to investigate how Stx phages differentially influence the virulence of STEC O157 strains

The characterisation of two Stx1 phages, originating from clinical Australian STEC O157 strains, revealed that these phages are morphologically and genomically similar to Stx2a phages. It was also observed that the bacterial host genetic background could influence toxin production. Genomic analysis revealed that these phages can potentially induce a translational frameshift with two overlapping tail-coding sequences with different host recognition domains, which may account for the broad host range of Stx phages leading to the emergence of new Stx producing pathogens. In addition, *in-silico* analysis also revealed a possible mechanism on how the phage obtained the *stx* gene by means of a Miniature Inverted Transposable element.

Stx prophages from three non-clinical Australian STEC O157 isolates were also characterised in this thesis. Two STEC O157 strains harboured both Stx1 and Stx2c prophages and it was observed in these strains that one Stx prophage was induced at a higher titre over the other. Genomic analysis predicted for the first time that the Stx2c phages package their DNA via cohesive ends. Further interrogation of the genomes also showed that two translational frameshift events, in different genes, are required for tail assembly and extension of host range respectively. In addition, each of the Stx2c prophages studied carried an anti-repression operon that may counteract the repression of the Stx1 prophage in-*trans* which could be a possible explaination for the increase in the production of Stx. This could be the mechanism as to why STEC

strains that harbour a Stx2c prophage in conjunction with another Stx prophage, in particular the Stx2a prophage, appear to be more virulent than other combinations of Stx prophages.

This thesis also reports the first evidence for Stx phage-mediated horizontal transfer of the locus of enterocyte effacement (LEE) pathogenicity island, as well as the characterisation of this Stx prophage. The mechanism of LEE mobilisation was determined to be likely due to a combination of both generalised and specialised transduction, and the incorporation of the LEE in recipient strains is likely due to homologous recombination. In addition to the discovery that an Stx phage can mediate the mobilisation of the LEE, genomic characterisation of this Stx prophage also revealed that it has a number of phage encoded genes that are predicted to enhance its ability to infect other susceptible bacterial strains by evading bacterial toxin-anti toxin defences and infecting a broad host range.

Overall, the results presented in this thesis suggested that Stx phages contribute more to bacterial pathogenesis than just the conference of the *stx* genes and that there are other interactions between the host and the Stx phage, which have resulted in STEC O157 becoming a successful pathogen.

1. Introduction

The ability to produce Shiga toxin by *Escherichia coli* O157 is associated with the progression to a severe clinical manifestation following infection and this phenotype conferred to the bacterium by the acquisition of Shiga toxin-encoding bacteriophages. When compared to other countries, the incidence of disease caused by *Escherichia coli* O157 is low in Australia and it is likely that these bacteriophages are influencing the pathogenicity of the bacterium. This thesis will focus on Shiga-toxin encoding bacteriophages from Australian STEC O157 strains. Specifically these bacteriophages will be characterised and their genomes interrogated to explore how these phages, in addition to conferring a toxin producing phenotype, can influence the pathogenicity of their host bacterial strains.

1.1. Bacteriophages

Bacteriophages (phages), viruses that infect bacteria, were independently co-discovered by English bacteriologist Frederick Twort (Twort, 1915) and French-Canadian microbiologist Felix d'Herelle (D'herelle, 1917) in the 1910s. Phages are now known to be the most abundant biological entities on earth with an estimated population of 10^{31} phages (Bergh *et al.*, 1989). Not only is the estimated phage population large, it is also highly dynamic with an estimated 10^{23} infection events per second occurring globally (Suttle, 2007).

The international body known as the International Committee on Taxonomy of Viruses (ICTV; http://www.ictvonline.org) governs phage taxonomy and phages are classified based on nucleic acid content and morphology. More than 6300 phages have been examined under the electron microscope and a diversity of phage morphology have been described (tailed, polyhedral, filamentous or pleomorphic) (Ackermann & Prangishvili, 2012). Based on morphology and nucleic acid content (single stranded (ss) or doubled stranded (ds), RNA or DNA), all known phages are divided into 10 families (Figure 1.1). Amongst these, the most abundant are dsDNA tailed phages (belonging to the order *Caudovirales*) (Ackermann & Prangishvili, 2012) and this group of phages will be the topic for discussion throughout this thesis.

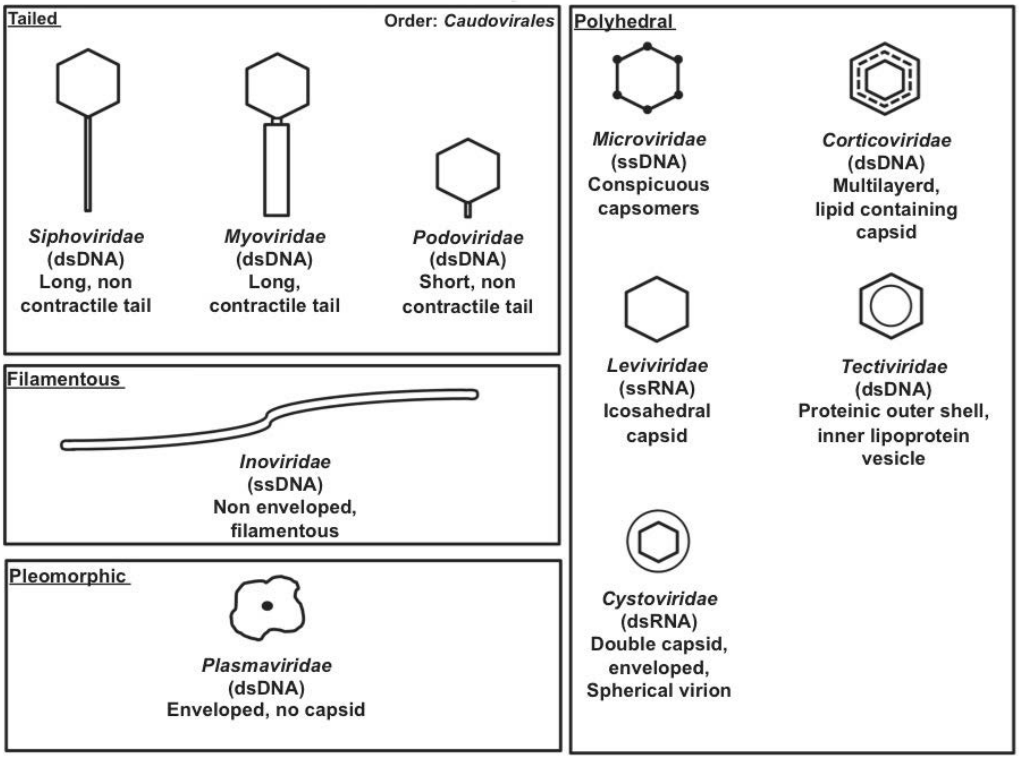


Figure 1.1: Diversity of phages. All known phage morphologies are shown with each a representative of the 10 phage families Figure adapted from Ackermann & Prangishvili (2012).

1.1.1 The structure of tailed phages

Members of the *Caudovirales* order possess a morphology comprising of a head (capsid) and a tail, a structure responsible for binding of the phage to a host bacterium and to deliver its DNA into the cell (Ackermann, 1998). The diameter of the capsid ranges from 30 nm to 160 nm which can be either icosahedral or an elongated icosahedron in shape (Ackermann & Prangishvili, 2012). Based on morphological difference in the phage tails, members of the *Caudovirales* order are subdivided into three different families: *Myoviridae* (long contractile tail), *Siphoviridae* (long non-contractile tail) and *Podoviridae* (short non-contractile tail) (Figure 1.1).

The tail of *Myoviridae* phages consist of a contractile sheath (Aksyuk *et al.*, 2009)surrounding a long, rigid tail tube, the length of which is controlled by a tape measure protein (Abuladze *et al.*, 1994). At the distal end of the tail lies a baseplate with emanating tail fibres (Leiman & Shneider, 2012). Once a myovirus attaches onto a bacterium, the baseplate undergoes a change in its conformation causing sheath contraction, bringing the phage closer to the bacteria surface and driving the tail tube through the bacterium (Leiman & Shneider, 2012). Unlike in *Myoviridae*, the tail tubes of

Siphoviridae may be flexible and are not surrounded by a sheath, but the length of the tail tube is also controlled by the tape measure protein (Katsura & Hendrix, 1984). In siphoviruses, the baseplate is usually referred to as the tail tip complex and this is where the tail fibres are attached (Davidson *et al.*, 2012). These fibres are named according to their position within the tail tip complex with fibres projected from the side of the tail tip complex named "side tail fibres" while the fibres attached to the centre of the tail tip complex named the "tail tip protein" which is responsible for penetrating the bacterial cell (Hendrix & Duda, 1992; Davidson *et al.*, 2012). In contrast to both *Myoviridae* and *Siphoviridae*, phages that belong to the *Podoviridae* family possess short tails. Podoviruses have been observed to harbour either tail spikes or tail fibres emanating from the short tail tube (Casjens & Molineux, 2012) although one podovirus has been document to harbour both tail spikes and tail fibres (Dai *et al.*, 2010). The tails of podoviruses are too short to penetrate the cell membrane but upon binding to the bacterial surface, podoviruses release a needle-like extension from their short tail in order to penetrate the host membrane (Casjens & Molineux, 2012).

The binding of tailed phages reversibly onto a primary bacterial surface receptor, mediating initial contact, is facilitated by the tail fibres (long thin trimeric intertwined β-strand tail appendages) or tail spikes (short, thick trimeric intertwined β-strand appendages) (Casjens & Molineux, 2012). These tail appendages recognise specific bacterial surface receptors with host recognition by the tail spikes/fibres dictated by the C terminal domain of these tail appendages (Fokine & Rossmann, 2014). Different phages have different tail fibres/spikes which recognise different surface molecules of Gram-positive and Gram-negative bacteria, including lipopolysaccharide (Lindberg *et al.*, 1970), outer membrane porins (Randall-Hazelbauer & Schwartz, 1973) teichoic acid (Xiang *et al.*, 2008), fimbriae (Roberts & Steitz, 1967), flagella (Lovett, 1972) and capsular polysaccharides (Taylor, 1965). Once initial contact is established, the tail tip protein will bind irreversibly to a secondary receptor, which will result in the phage injecting its DNA into the bacteria (Casjens & Molineux, 2012).

1.1.2. Phage genomes

The size of the encapsulated phage DNA is found in the range of 17 kb to 498 kb (Ackermann & Prangishvili, 2012). The genomes of dsDNA phages are highly organised with genes encoding proteins with related functions located close to each other, forming functional modules (e.g. DNA replication, bacterial lysis, head morphogenesis, tail morphogenesis) (Campbell, 1994; Hatfull, 2008). In addition to functional modules, there is also a general order in the genomic architecture of phage genome where the head morphogenesis modules are located upstream of the tail morphogenesis genes (Casjens, 2005), which suggested that this is an early feature in the evolution of tailed phages (Hatfull, 2008). The ends of the phage genome are generated by the cleavage mediated by the phage large terminases before packaging into phage heads, generating different genome termini with at least 6 types classified: (1) single stranded cohesive (*cos*) ends, (2) circular permutation of terminal redundancy (3) non-permuted exact short terminal redundancy, (4) non-permuted exact long redundancy, (5) host DNA fragments at phage termini and (6) covalent terminal proteins (Casjens & Gilcrease, 2008).

Another notable feature of phage genomes is their mosaic structures where each individual phage genome is considered as a combination of functional modules which can be exchanged among the population and no two phages, even among highly related phages, are identical (Hendrix *et al.*, 2003). Genomic mosaicism in phages can be observed at the nucleotide level by either heteroduplex DNA mapping under the electron microscope (Simon *et al.*, 1971; Highton *et al.*, 1990) or in-*silico* by comparative genomic analysis (Juhala *et al.*, 2000). At the resolution of nucleotide sequences, it was observed that clear homology breakpoints (regions of genomic differences flanked by regions of genomic similarity) are formed, corresponding to the boundaries of DNA segments with different evolutionary origins (Juhala *et al.*, 2000). Functional modular exchange was described in the pre-genomics era (Botstein, 1980) and it is now the most widely accepted hypothesis of phage genomic mosaicism with three possible recombination mechanisms for functional module exchange proposed. The first mechanism posits that short conserved DNA sequences at the homology breakpoint serves as a target for homologous recombination (Susskind & Botstein, 1978). The second mechanism suggests that the recombination events are not targeted and occur randomly by illegitimate recombination coupled with selection where non-functional phages after recombination are lost while

viable phages are selected (Hendrix *et al.*, 2000) while the third mechanism posits that homologous recombination is mediated by phage encoded recombinases (Martinsohn *et al.*, 2008). Genomic mosaicism in phages can also be observed at the predicted amino acid sequence level by amino-acid sequence comparison of phage encoded gene products (Hatfull *et al.*, 2006). This is an informative approach to view genomic mosaicism because many phages, inclusive of those that infect a common host, may not share nucleotide similarity and the protein sequence can reveal genes that share older ancestry allowing a researcher a possible glimpse of the evolutionary routes taken by a functional module (Hatfull, 2008).

1.1.3. Phage life cycle

Phages are obligate parasites and thus require a bacterial host for their propagation and can be classified as either virulent or temperate according to their life cycle (Figure 1.2).

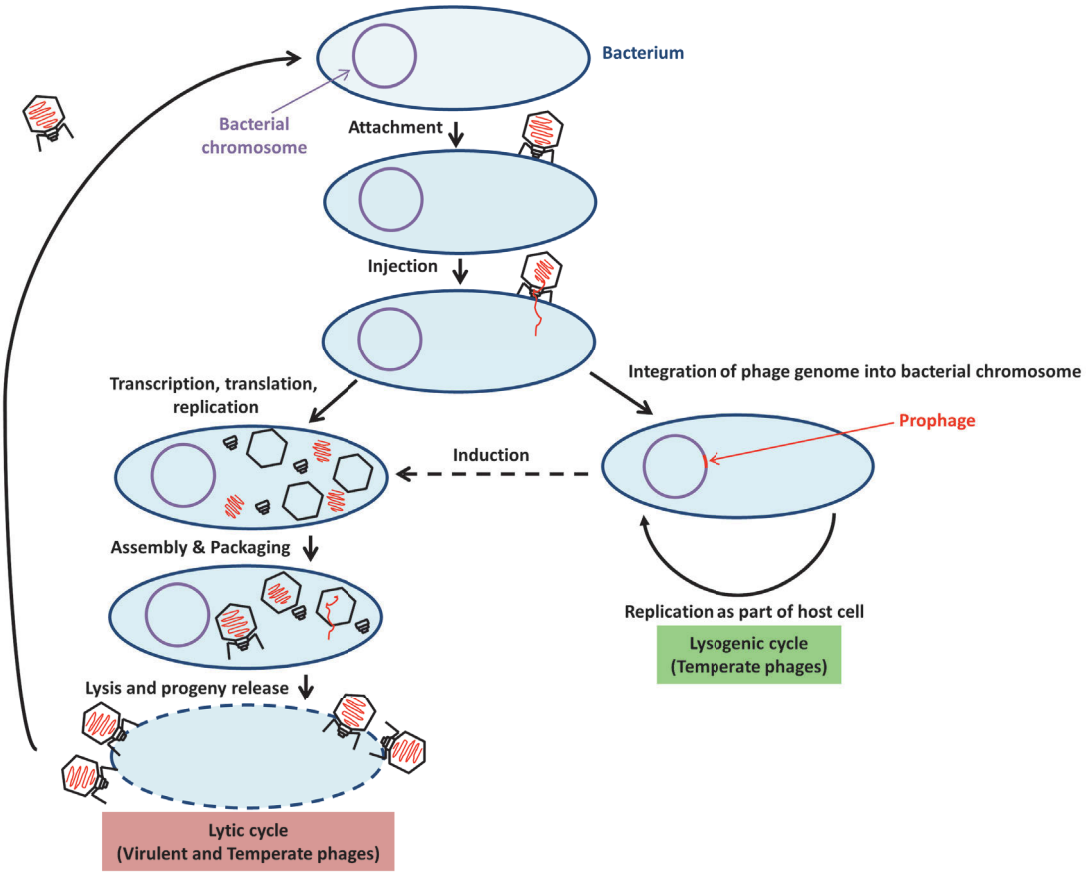


Figure 1.2: Phage life cycles. After infection of phage DNA, virulent phages enter the lytic cycle while temperate phages can either enter the lytic or lysogenic life cycle. In the lysogenic life cycle, the phage (now termed a prophage) replicates as part of its host cell.

In addition, the prophage can be induced from the bacterial cell to enter the lytic cycle. Figure adapted from Campbell (2003a)

1.1.3.1 Lytic life cycle

The lytic lifecycle (Figure 1.2) starts with the attachment of a phage onto its specific bacterial cell surface receptor and the injection of phage DNA into the bacterial cell. The phage hijacks the bacterial host for its replication and transcription of phage genes and replication of the phage DNA then occurs and long concatamers of phage DNA are generated (Weigel & Seitz, 2006). Structural proteins are then synthesised and assembled into phage heads and tails.

During packaging of the phage DNA into the assembled phage heads, the concatemeric phage DNA is recognised by the small terminase subunit and subsequently cleaved by the large terminase subunit (Casjens & Gilcrease, 2008). Two mechanisms of packaging phage DNA into phage capsids are known to date: *cos* packaging and headful packaging. Both ends of a *cos* phage genome consist of complementary protruding single-strands (either 5' or 3') of identical length that are reported to be between 7 and 19 nucleotide in length for different phages (known as the *cos* site) (Hershey & Burgi, 1965; Padmanabhan *et al.*, 1974; Fitzmaurice *et al.*, 1984). These ends are generated when the large terminase makes a sequence specific cut in the first *cos* site which initialises DNA packaging and packaging is terminated at the other end where the large terminase specifically recognises the second *cos* site (Higgins *et al.*, 1988). This results in the phage heads containing 100% of the phage genome length (Feiss *et al.*, 1977). In headful packaging, DNA packaging is initiated when the terminase recognises a sequence specific site (known as a *pac* site) and packages DNA into the phage head to its full capacity (Casjens & Hayden, 1988). Once the phage head is filled with DNA, sequence independent cleavage occurs and packaging is terminated (Casjens & Hayden, 1988). This often leads to the packaging of 102% - 110% of the total genome size of the phage due to the lack of sequence specificity of its second cleavage and creates phage genomes with direct terminal repeats that vary from 2%- 10% of its genome length in different phages (Casjens & Gilcrease, 2008). This packaging of addition DNA results in the ends of each successive packaging event to be 'moved' along the genome sequence from the ends generated by the previous packaging event, leading to the phage genome to be terminally redundant (Tye *et al.*, 1974).

Once the DNA has been packaged, the neck protein substitutes the terminase proteins and this structure prevents the packaged DNA from leaking out from the phage head (Orlova *et al.*, 2003; Lhuillier *et al.*, 2009) and completing the head assembly. Subsequently, the phage tail proteins are attached to the assembled head (Aksyuk & Rossmann, 2011). Once assembly of the phage is completed, a phage encoded holin will create pores within the bacterial cytoplasmic membrane allowing the phage encoded endolysin to cause cell lysis (Young, 1992). This resulting lysis will release progeny phages to the environment that will be able to infect susceptible bacteria and the lytic cycle can start again in the new host.

1.1.3.2. Lysogenic life cycle

Temperate phages can undergo the same lytic life cycle but they can also undergo an alternative life cycle called the lysogenic life cycle (Figure 1.2) (Lwoff, 1953). The 'decision' to commit to either the lytic or lysogenic life cycle occurs after phage DNA injection is influenced by the environment and/or its host metabolic state (Ptashne, 2004). Some environmental factors influencing the decision to commit to the lysogenic life cycle include infecting carbon–starved bacterial cells (Kourilsky & Knapp, 1974) and in conditions whereby there are more infectious phages than susceptible bacteria (Kourilsky, 1973; Kourilsky & Knapp, 1974; Kobiler *et al.*, 2005)

In the lysogenic life cycle, the phage DNA can either integrate into the bacterial chromosome and replicate as part of the bacteria chromosome (Lwoff, 1953), or exist as a circular extra-chromosomal plasmid which self-replicates and is partitioned equally into both daughter cells during bacterial cell division (Ikeda & Tomizawa, 1968). Once integrated, the phage is termed a prophage and the bacterial host harbouring the prophage is known as a lysogen (Lwoff, 1953). Integration of the phage DNA into the host chromosome is mediated by the phage encoded integrase, which mediates site specific recombination between short, similar DNA sequences, the phage attachment site (*attP*; located in the phage genome) and the bacterial attachment site (*attB*; located in the bacteria genome) (Campbell, 1992; Groth & Calos, 2004). There exist two families of phage integrases, the tyrosine integrases utilise a catalytic tyrosine residue while serine integrases utilise a catalytic serine residue for DNA strand cleavage and binding the DNA to the integrase. (Groth & Calos, 2004). In addition to the catalytic residue, the

mechanism in which the different integrase families mediate site-specific recombination is different, with the tyrosine integrases forming a holiday junction to complete DNA exchange between *attP* and *attB* (Biswas *et al.*, 2005) while serine integrases make 2-bp staggered cuts in both *attP* and *attB* followed by ligation to complete DNA exchange (Smith *et al.*, 2004; Olorunniji *et al.*, 2012). Once integrated into the bacterial genome, the prophage is flanked by *attL* and *attR* sites which are hybrids containing half of *attB* and *attP* (Figure 1.3) (Groth & Calos, 2004).

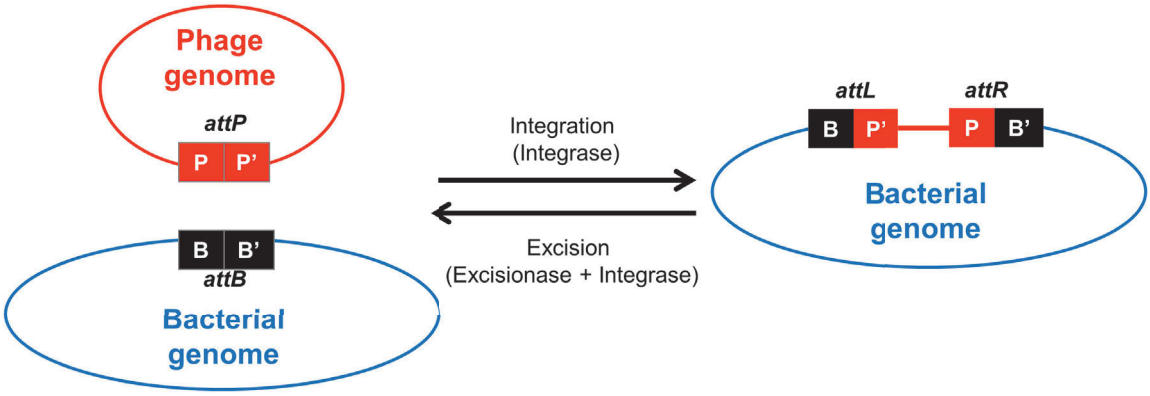


Figure 1.3: Integration and excision of prophage. (A) The red line represents the phage genome with the phage attachment site *attP* shown as a red box, while the blue line represents the bacterial genome with a black box representing the bacterial attachment site *attP*. The right side of the *att* is differentiated from the left side of *att* by the (') symbol. Integration of the phage genome will result in the formation of two hybrid *att* sites *attL* and *attR*. P and B represents phage and bacterial sequence respectively. Figure adapted from Groth & Calos (2004).

Once integrated into the bacterial genome, a prophage encoded repressor protein is expressed and inhibits transcription of most of the phage genes, including those required for the lytic cycle, and the prophage becomes quiescent with the repressor maintaining this lysogenic state (Echols & Green, 1971). In addition, the repressor also blocks expression of genes from incoming isogenic phages, granting the lysogen immunity to superinfection by phages carrying similar repressors (Dodd & Egan, 1996)

Although the lytic and lysogenic lifecycles are distinct, they are not mutually exclusive as the prophage can excise from the bacterial chromosome and switch to the lytic life cycle in a phenomenon known as prophage induction (Lwoff, 1953). Prophage induction generally requires the cleavage of the phage repressor by RecA, which occurs upon activation of the RecA-dependent bacterial SOS response following DNA damage from

agents like UV light and antibiotics (Lwoff, 1953). In addition, some phages also carry anti-repressor genes that encode proteins which recognises their cognate repressors or those of other prophages within the same bacterial host genome, which alleviates repression of the prophage, switching them into the lytic life cycle (Susskind & Botstein, 1975; Lemire *et al.*, 2011). All prophage excision events are catalysed by the integrase but a phage-encoded excisionase is also required to control the directionality of this excision reaction by facilitating the formation of a nucleoprotein complex through the alteration the DNA structure by bending, preventing reintegration of the excised prophage (Lewis & Hatfull, 2001; Sam *et al.*, 2004; Abbani *et al.*, 2007).

With the advent of genome sequencing, it has been observed that prophage sequences can constitute as much as 20% of the bacterial genome (Casjens, 2003). Some of these prophages are intact but some prophages, known as cryptic prophages, are defective due to inactivating mutations and may be irreversibly degraded and unable to switch to the lytic life cycle (Campbell, 1994; Casjens, 2003). However, some cryptic prophages have been observed to excise from the bacterial chromosome and be packaged into phage heads formed by other induced prophages within the bacterial genome (Asadulghani *et al.*, 2009; Wang *et al.*, 2010). Cryptic prophages, as well as intact prophages, have been shown to contribute to phage genomic mosciasm by recombination with other phage genomes, creating novel combinations of functional modules (De Paepe *et al.*, 2014). In addition, the harbouring of multiple related prophages in the bacterial chromosome is also believed to lead to large chromosomal rearrangements, deletions and inversions within the bacterial chromosome (Canchaya *et al.*, 2004).

1.1.3.3. Pseudolysogeny

Apart from the lytic and lysogenic life cycle, there is another phage life cycle called pseudolysogeny (Romig & Brodetsky, 1961). Pseudolysogeny is a state of "stalled development" where the injected phage DNA resides in a non-active state, neither eliciting a lytic response or integrating into the host genome in a stable fashion, until conditions such as increased availability of nutrients occur, triggering it to enter either into the lytic or lysogenic life cycle. It is likely that this carrier state could be a mechanism for survival of the phages in unfavourable conditions (Łoś & Węgrzyn, 2012).

1.1.3.4. Satellite phages- a subset of temperate phages

Subsets of temperate phages are the satellite phages, which are prophages that hijack another phage (helper phage) in order to produce functional virions. The prototypical satellite/helper phage system is the P4/P2 system (Christie & Calendar, 1990). The 11.6 kb satellite phage P4 can exist intracellularly as a prophage either integrated in the bacterial chromosome or as a multi copy plasmid (Calendar *et al.*, 1981; Deho *et al.*, 1984). While possessing the genes required for lysogeny , phage replication and prophage excision, P4 lacks the genes encoding phage structural proteins (Six, 1975). The lytic cycle of P4 requires the presence of its helper phage P2 and when both phages are present within the same bacterial cell and producing transcripts, reciprocal regulatory interactions from both P4 and P2 will lead to the mutual depression of both prophages as well as allowing P4 to manipulate P2 into producing phage virions (Christie & Calendar, 1990; Christie & Dokland, 2012). In addition, P4 is able to influence the capsid assembly process, forming smaller capsids commensurate with its smaller genome size (Dokland *et al.*, 1992). Lysis of the bacterial host will release two subpopulations of phages, at the same time, those with a smaller capsid containing P4 genome and those with a larger capsid containing P2 genome, both capable of infecting other susceptible bacterial cells (Christie & Calendar, 1990).

1.1.4 Coevolution of phages and bacteria

Phages are constantly in an evolutionary arms race with their bacterial hosts. Bacteria evolve mechanisms to defend against phage infection by blocking various steps of its life cycle while phages evolve mechanisms to counter the defences put forth by their bacterial host to ensure their survival. These mechanisms are listed below:

1.1.4.1. Inhibition of phage adsorption

Binding to the surface receptor of its host bacterium is the first step of phage infection. However, bacteria can prevent the phage from accessing its receptor by utilising phase variation to alter their cell surface or producing extracellular polymers to form a physical barrier between the phage and its receptor (Hammad, 1999; Holst Sorensen *et al.*, 2012; Kim & Ryu, 2012). Phages can circumvent this by either evolving to recognise new surface receptors (Meyer *et al.*, 2012), changing the specificity of their host recognition

proteins through tail fibre modification (Stockdale *et al.*, 2013) or by producing enzymes to degrade extracellular polymers masking the surface receptor (Cornelissen *et al.*, 2012).

1.1.4.2. Preventing DNA entry

Proteins that are either anchored on or associated with the cell membrane can block the entry of phage DNA into bacterial cells after the phage has bound to its receptor and injected its DNA. This manner of blocking is known as superinfection exclusion and the genes encoding these proteins are harboured by prophages within the bacterial cell. This form of protection works by either blocking the transfer of phage DNA into the bacterial cytoplasm by changing the conformation of the injection site (Lu *et al.*, 1993) or by preventing the degradation of the peptidoglycan by the phage, trapping the phage DNA between the peptidoglycan layer and the outer membrane (Lu & Henning, 1994)

1.1.4.3. CRISPR-Cas system

Once injected into the bacterial cell, the phage DNA can be subjected to destruction by a CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated) system, a form of bacterial adaptive immunity that functions in three general steps (Barrangou *et al.*, 2007). The first step is called adaptation, where short stretches of non-repetitive DNA (spacers) acquired from foreign DNA (protospacers; 26-72 bp) are incorporated into the CRISPR loci between the CRISPR repeat sequences (21-48 bp direct repeats) (Sorek *et al.*, 2008). The second step is biogenesis where the CRISPR locus is transcribed to produce short crRNA which contains segments of a CRISPR repeat and a protospacer and subsequently complexes with the Cas proteins to form the CRISPR-Cas complex (Sorek *et al.*, 2008). The third general step is interference where the CRISPR-Cas complex seeks out foreign DNA with nucleotide sequence identical to the protospacer and cleaves the foreign DNA (Sorek *et al.*, 2008).

Phages can evade the CRISPR-Cas system by either having mutations in their protospacers to prevent detection by the CRISPR-Cas complex (Deveau *et al.*, 2008), or encoding an anti-CRISPR protein that interfere with the CRISPR-Cas systems (Bondy-Denomy *et al.*, 2013) or by encoding a phage borne CRISPR-Cas system to inactivate the host antiviral defence (Seed *et al.*, 2013).

1.1.4.4. Restriction-modification systems

Restriction-modification (R-M) systems are another bacterial defence where restriction endonucleases degrade foreign DNA, upon entry into the cell, at specific restriction sites (Pingoud *et al.*, 2005). Modification of the restriction sites in the host DNA by its partner methyltransferase protects the host DNA from cleavage by the restriction endonuclease (Tock & Dryden, 2005). Phage mechanisms to evade R-M systems include possessing few restriction sites (Kruger *et al.*, 1988), modification of restriction sites by the incorporation of modified bases (Warren, 1980), phage-encoded methyltransfereases to protect the phage genome against corresponding R-M systems (Drozdz *et al.*, 2012), masking its restriction site (Iida *et al.*, 1987), enhancing host methyltransferase to methylate phage DNA (Loenen & Murray, 1986) or producing an enzyme to degrade a co-factor required by the R-M system (Studier & Movva, 1976).

1.1.4.5. Abortive infection

Abortive infection (Abi) systems are anti-phage defences, which induce bacterial cell death; an altruistic mechanism that allows non-infected bacteria to escape infection. The genes encoding the single protein or protein complexes of the Abi systems, are usually found in prophages and plasmids (Samson *et al.*, 2013). Abi systems documented to stop phage propagation include changing of the physiology of the bacterium to disrupt phage DNA replication (Schmitt *et al.*, 1991; Parma *et al.*, 1992), disruption of the bacterial translational machinery preventing the phage from hijacking it (Snyder, 1995) and cleaving of tRNAs in their anticodon loops which halts translation (Snyder, 1995). The toxin-antitoxin (TA) system is a subgroup of Abi system, which acts after the injection of the phage DNA into the bacterial cell, leading to both the death of the bacterial cell and the abortion of phage infection (Pecota & Wood, 1996; Yamaguchi *et al.*, 2011). The TA system consists of a toxin and its neutralising antitoxin (Yamaguchi *et al.*, 2011). The antitoxin is more labile than the toxin and is rapidly degraded when stress (e.g. phage infection) occurs, allowing the toxin to induce cell death and aborting the phage infection (Fineran *et al.*, 2009a; Yamaguchi *et al.*, 2011). To escape abortive infection by the TA system some phages either encode their own antitoxin or a pseudo antitoxin mimic to prevent the killing effects of the toxin (Fineran *et al.*, 2009a; Otsuka & Yonesaki, 2012).

1.1.5. Phages and horizontal gene transfer

Horizontal gene transfer (HGT) is the phenomenon of physical intercellular transfer of DNA from one bacterium to another, independent of cell division, and its subsequent stable incorporation into the recipient bacterial genome either by recombination, site specific integration or as extrachromosomal DNA which will be subsequently be passed on to each daughter cell vertically via cellular division (Ochman *et al.*, 2000). HGT occurs via four main mechanisms: conjugation, transformation, transduction and lysogenic conversion (Ochman *et al.*, 2000; Brussow *et al.*, 2004) (Figure 1.4). Conjugation is a contact dependent mechanism of DNA transfer mediated by a specialised apparatus known as the conjugative pilus through which plasmids and integrative conjugative elements are transferred to other bacterial strains (Cabezon *et al.*, 2015). Transformation is the uptake of extracellular DNA from the environment that occurs when bacteria enter a physiological state of competence (Johnston *et al.*, 2014)The last two mechanisms of HGT, transduction and lysogenic conversion, are mediated by phages and these processes are linked to the phage life cycle, as described below.

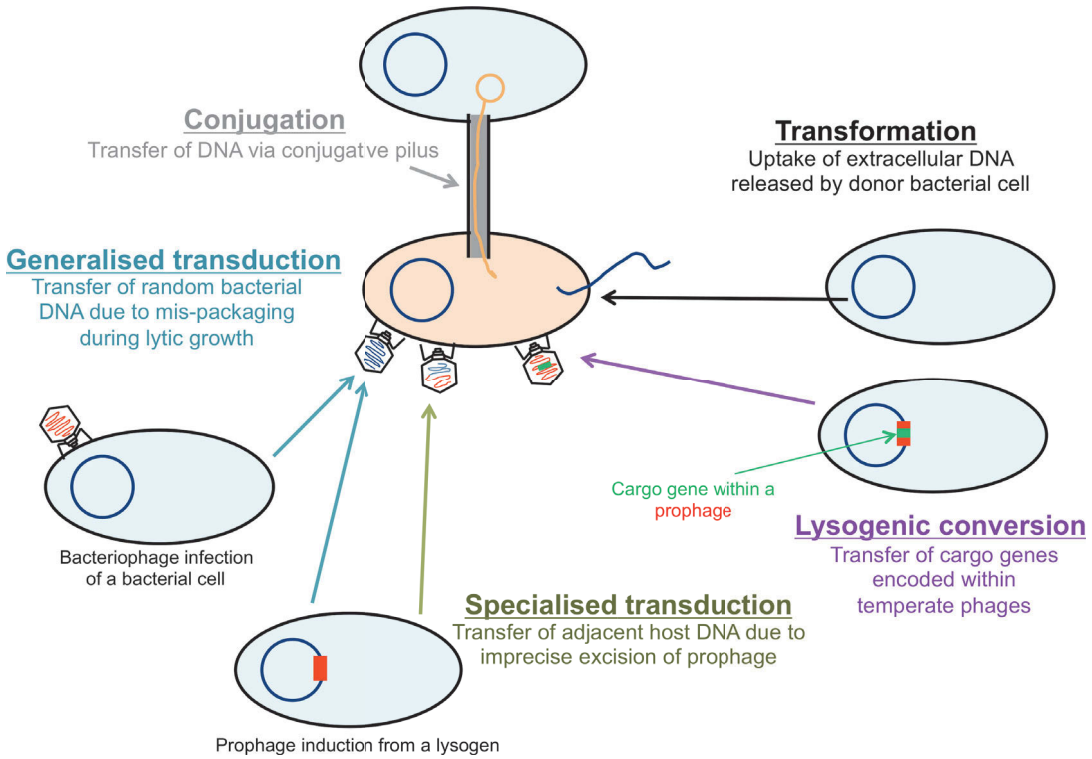


Figure 1.4: Mechanisms of horizontal gene transfer in bacteria. The recipient bacterial cell is coloured in orange while the donor bacterial cells are coloured in blue. Brief descriptions of the different mechanisms of HGT are colour coded according to the arrows in the figure. Figure adapted from Zaneveld *et al*., (2008), Fineran *et al*., (2009b) and Brussow *et al*., (2004).

1.1.5.1 Transduction

Transduction, which is the mispackaing of bacterial DNA into phage heads, can be differentiated into generalised transduction and specialised transduction, based on the DNA content that is packaged within the phage heads.

1.1.5.1.1. Generalised transduction

Generalised transduction is a HGT event whereby any section of the bacterial DNA (chromosome or plasmid) is transferred from one bacterium to another via a phage virion (Figure 1.5) (Fineran *et al.*, 2009b). Generalised transducing phages can be either virulent or temperate phages that package DNA using the headful packaging mechanism. During the phage lytic cycle, random bacterial DNA is mispackaged at low frequency into phage heads, instead of phage DNA forming transducing particles (Fineran *et al.*, 2009b). Once filled with bacterial DNA and virion assembly is completed, these transducing particles are able to infect susceptible bacteria (Figure 1.5). Once DNA is injected into a susceptible bacterial cell, three possible fates await. The first is abortive transduction where the transduced DNA is not integrated into the bacterial chromosome and unable to replicate among the bacterial progeny and subsequently diluted out during cell division (Stocker, 1956). The second fate is the recycling of nucleotides where the transduced DNA is degraded by the host into component nucleotides that can be incorporated during DNA repair (Ebel-Tsipis *et al.*, 1972). The third possible fate is recombination into the bacterial DNA by homologous recombination, a rare event which results in the successful transfer of foreign DNA which can be stably maintained in the bacterial chromosome (Fineran *et al.*, 2009b).

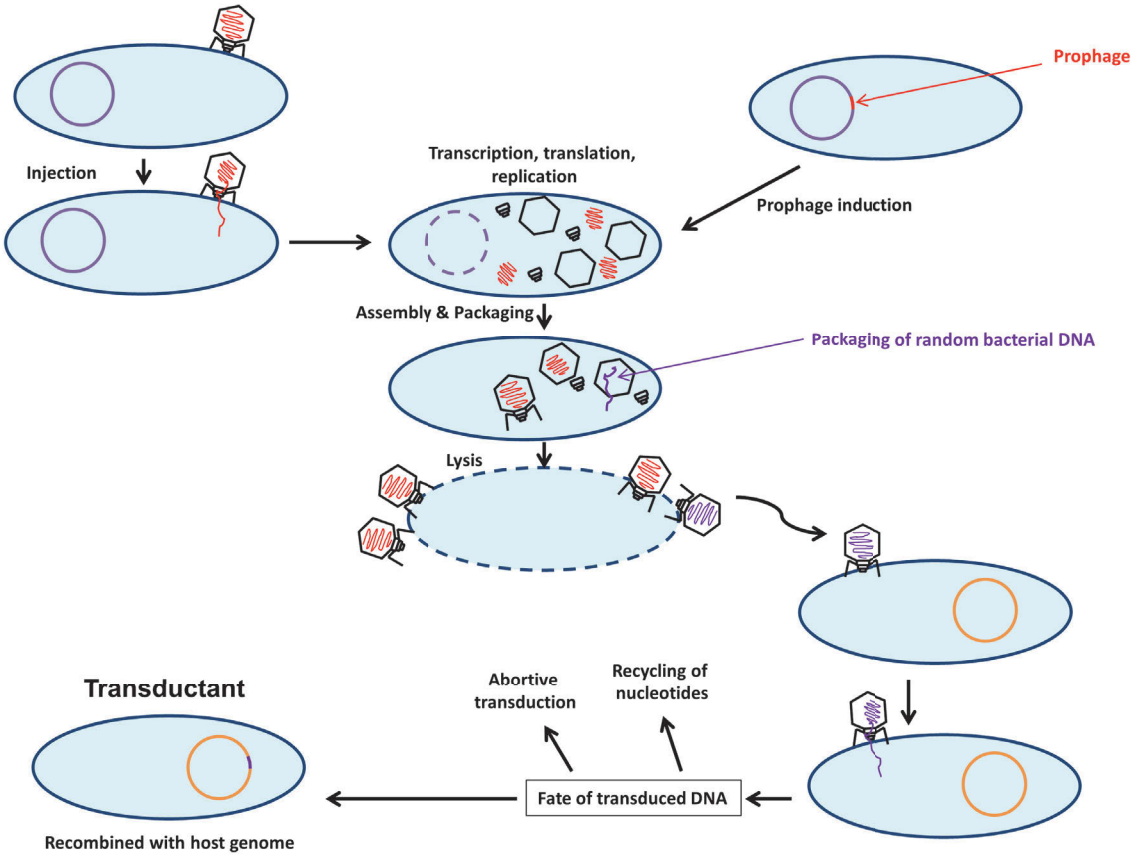


Figure 1.5: Overview of generalised transduction. This occurs during the lytic life cycle where mispackaging of random bacterial DNA at low frequency into phage heads occurs. Transducing particles can adsorb to susceptible bacterial cells and inject the packaged bacterial DNA. There are three possible outcomes after infection of DNA with only recombination leading to the formation of stable transductants. Phage DNA is shown in red, donor DNA in purple and recipient DNA in orange. Figure adapted from Fineran *et al.*, (2009b).

1.1.5.1.2. Specialised transduction

Specialised transduction, is a phenomenon that occurs during the aberrant excision of a prophage from the bacterial chromosome (Figure 1.6). The difference between specialised transduction and generalised transduction is the type of DNA that is mispackaged into phage heads; mispackaged bacterial DNA in generalised transduction is random but mispackaged bacterial DNA during specialised transduction is localised to adjacent bacterial DNA flanking the integrated prophage (Morse *et al.*, 1956). The excision of prophages is usually a precise event but occasionally, prophages may excise imprecisely leading to the excision of adjacent bacterial along with phage DNA which is then packaged into phage heads (Morse *et al.*, 1956). On injection of this DNA into a new susceptible bacterial cell, the adjacent bacterial DNA from the previous bacterial host will

be integrated into the new bacterial chromosome along with the prophage, during the lysogenic cycle (Figure 1.6) (Fineran *et al.*, 2009b). Due to imperfect excision, some phage gene(s) might be left behind and depending on the type of gene(s) left behind; the phage could be defective (Fineran *et al.*, 2009b).

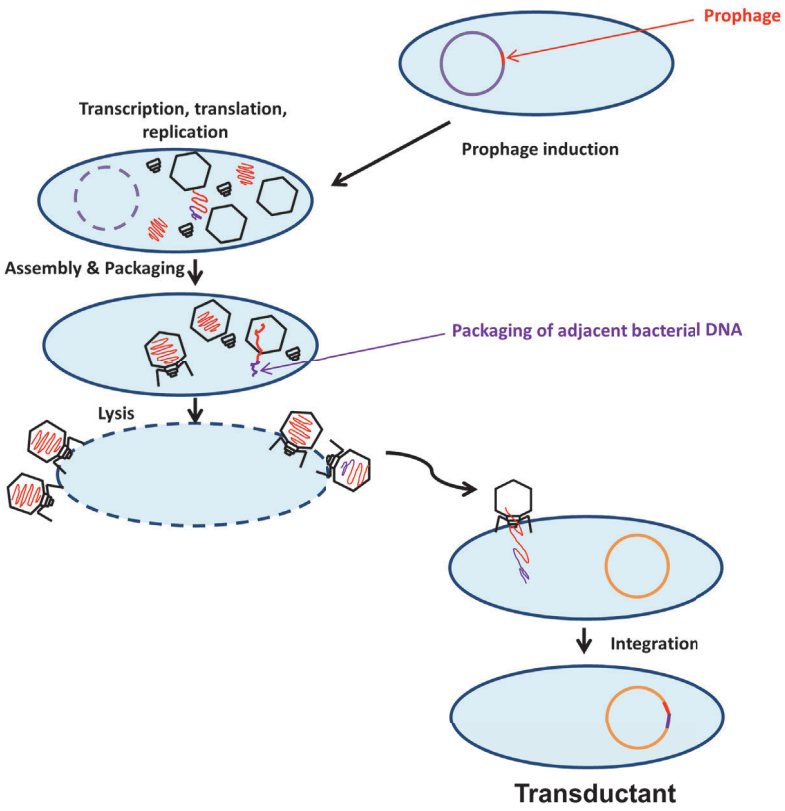


Figure 1.6: Overview of specialised transduction. During prophage induction, a mistake in excision can lead to flanking bacterial DNA being excised and packaged together with the prophage into phage heads. A new recipient strain can acquire this foreign bacterial DNA via lysogeny. Phage DNA is shown in red, donor bacterial DNA in purple and recipient bacterial DNA in orange. Figure adapted from Fineran *et al.*, (2009b).

1.1.5.2. Lysogenic conversion

Some temperate phages carry extra genes, which are not required for the biological activity of the phage but which during the prophage state confer upon their bacterial host a phenotypic change, a process called lysogenic conversion (Brussow *et al.*, 2004). These extra genes, known as cargo genes, are additions to a phage genome with a G+C content that differing from the surrounding phage genes (indicative of HGT) and flanked by an upstream promoter and a downstream terminator which allows for the expression of the cargo genes without interfering adjacent prophage genes (Hendrix *et al.*, 2000). In

addition, these cargo genes are usually found in a 'cargo hold', a location that causes minimal/no disruptions to a phage gene and its typically within a functional module. This 'cargo hold' is conserved amongst related phages but the cargo gene it carries may vary between phages (Nilsson *et al.*, 2004; Petty *et al.*, 2011). The localisation of the cargo genes, within functional modules suggested that these cargo genes are exchangeable with other phages, together with the functional module, based on the modular theory of phage evolution (Botstein, 1980; Brussow *et al.*, 2004). However, the origins of how these cargo genes initially entered the phage genome is still currently unknown but a mechanism of site-specific insertion of cargo genes into phage genomes has been previously postulated (Nilsson *et al.*, 2004).

Some outcomes of lysogenic conversion include resistance to antibiotics (Billard-Pomares *et al.*, 2014), serum resistance (Barondess & Beckwith, 1995), enhanced adhesion to eukaryotic cells (Vica Pacheco *et al.*, 1997), defence against phagocyotosis (Meltz Steinberg & Levin, 2007) and the transfer of virulence genes which contribute to bacterial pathogenesis (Brussow *et al.*, 2004). A notable example of lysogenic conversion contributing to bacterial pathogenesis is the conversion of bacteria into toxin producing pathogens. For example, the Shiga toxin genes of Shiga toxigenic *Escherichia coli* are encoded on a prophage (Kaper *et al.*, 2004).

1.2. *Escherichia coli*

A member of the *Enterobactericeae* family, *Escherichia coli* (*E. coli*) was first discovered in the human colon in 1885 by Theodor Escherich (Escherich, 1885). It is a Gram negative, rod shaped, facultative anaerobe, usually found as an intestinal commensal of warm blooded animals and humans (Kaper *et al.*, 2004). In the pre-molecular era, serotyping studies revealed that the somatic (O) antigen, flagella antigen (H) and the capsular antigen (K; to a lesser extend) are useful in distinguishing between different strains of *E. coli* and has since been adopted as a way of classifying *E. coli* strains (Ørskov *et al.*, 1977). Years later, the development of a technique known as multi-locus enzyme electrophoresis (MLEE), which looked at the electrophoretic mobility of enzymes in *E. coli* strains, revealed more diversity within the *E. coli* population and in the years to follow, a hypothesis was formed where *E. coli* populations evolve in a clonal manner with recombination playing a minor role and that the O, H and K serotypes do not correlate

with the population diversity of *E. coli* (Ochman & Selander, 1984; Caugant *et al.*, 1985). The adoption of MLEE also led to the classification of *E. coli* into five major phylogenetic groups (A, B1, B2, D and E) (Herzer *et al.*, 1990).

These usually benign commensals can evolve into pathogenic strains by acquiring virulence factors encoded within mobile genetic elements such as plasmids, genomic islands and bacteriophages (Kaper *et al.*, 2004). Pathogenic *E. coli* are commonly grouped into pathotypes according to their clinical manifestation (Kaper *et al.*, 2004). Pathotypes of pathogenic *E. coli* include: Meningitis/sepsis-associated *E. coli* (MNEC), Uropathogenic *E. coli* (UPEC), Enteroinvasive *E. coli* (EIAC), Enteroaggrevative *E. coli* (EAggEC), Diffusely adherent *E. coli* (DIAC), Enterotoxigenic *E. coli* (ETEC), Enteropathogenic *E. coli* (EPEC) and lastly, Shiga toxigenic *E. coli* (STEC). The STEC pathotype includes a subgroup known as Enterohaemorrhagic *E. coli* (EHEC) which consist of STEC strains that causes haemorrhagic colitis or haemolytic uremic syndrome in addition to producing Stx (Kaper *et al.*, 2004).

1.3. *Escherichia coli* O157

A notorious member of the EHEC/STEC pathogroup is *E. coli* O157. For consistency, *E. coli* O157 will be referred to as STEC O157:H7 for the rest of this thesis. STEC O157:H7 first gained notoriety in 1983 when it was associated with an outbreak of haemorrhagic colitis from the consumption of undercooked beef in Oregon, Michigan (Riley *et al.*, 1983). Since then, the status of STEC O157 has transited from an emerging infectious disease into a global public health concern (Snedeker *et al.*, 2009). Only a small infectious dose of 10 -100 colony forming units was predicted to establish an infection (Tuttle *et al.*, 1999) and clinical outcomes following infection ranges from asymptomatic patients to death (Cowden *et al.*, 2001; Mannix *et al.*, 2005). Associated symptomatic clinical manifestation include neurological complications (Hamano *et al.*, 1993), diarrhoea, haemorrhagic colitis and a progression to a severe disease state known as haemolytic uremic syndrome (HUS), characterised by haemolytic anaemia, thrombocytopenia and acute kidney damage, which in some cases can lead to kidney failure or death in some individuals (Peacock *et al.*, 2001). The progression to HUS occurs in 5 to 7% of infected patients (disproportionally affecting children and the elderly) and usually within 5 to 13 days post diarrhoea onset (Tarr *et al.*, 2005). HUS is recognised as a common cause for

childhood acute renal failure with a significant 5% mortality rate and a 25% risk for subsequent development of chronic renal and extra-renal complications (Scheiring *et al.*, 2008). In addition, elderly patients have the highest rate of death associated with STEC O157 infection (Gould *et al.*, 2009). There is currently no treatment for STEC infections and management of infected patients replies on supportive therapy and hydration (Tarr *et al.*, 2005).

1.3.1. Transmission of STEC O157

The primary reservoir of STEC O157 has been determined to be the gastrointestinal tract of ruminant animals, particularly cattle, where it is harboured in the lower intestine (Low *et al.*, 2005). Even though ruminants carry the bacterium, they have been shown to be asymptomatic and are thus vectors of direct or indirect transmission of STEC O157 to humans (Caprioli *et al.*, 2005). The initial observation of a vehicle for STEC O157 transmission was associated with undercooked ground beef (Riley *et al.*, 1983). Ground beef is the most common transmission vehicle due to the ease of cross contamination in food production (Ferens & Hovde, 2011). Food borne transmission of STEC O157 are not only limited to contaminated ground beef products as a myriad of other food products have been previously implicated to be a vehicle of transmission of STEC O157 which includes, but not limited to, pork (Trotz-Williams *et al.*, 2012), deer jerky (Keene *et al.*, 1997), salami (Macdonald *et al.*, 2004), lettuce (Hilborn *et al.*, 1999), watercress (Launders *et al.*, 2013), raw milk (Guh *et al.*, 2010), unpasteurised cheese (Honish *et al.*, 2005), unpreserved apple cider (Besser *et al.*, 1993), cookie dough (Neil *et al.*, 2012), radish sprouts (Michino *et al.*, 1999) and spinach (Wendel *et al.*, 2009). Waterborne outbreaks of STEC O157 have been associated with both drinking (Swerdlow *et al.*, 1992) and recreational (Verma *et al.*, 2007) water. Contact with ruminant animals has also been associated with STEC O157 outbreaks (Crump *et al.*, 2002; Crump *et al.*, 2003). STEC O157:H7 can also spread from humans to humans via a secondary infection due to close contact and poor hygiene practices (Reida *et al.*, 1994; Werber *et al.*, 2008). Infections in laboratory workers have also been documented, highlighting the need for proper biosafety containment and handling of this pathogen (Spina *et al.*, 2005).

Apart from the first documented outbreak of STEC O157 in the USA, in 1982, there were also two other notable outbreaks. The first was an outbreak in Sakai, Japan in the year

1996 associated with radish sprouts, which led to the largest outbreak recorded with 7966 cases of STEC O157 infection with 106 of these patients progressing to HUS with 3 deaths (Michino *et al.*, 1999). The second was a multistate outbreak in the USA in the year 2006, associated with spinach, where 51% of infected patients were hospitalised and 15% of infected patients progressed to HUS and kidney failure (Kulasekara *et al.*, 2009; Wendel *et al.*, 2009).

1.3.2. Geographical variation in STEC O157 related disease

While STEC O157 is a pathogen of global public health concern, the incidence of STEC O157-associated infection varies geographically with the highest incidence coming from Argentina while the country with the lowest incidence of STEC O157 related disease is Australia (Figure 1.7).

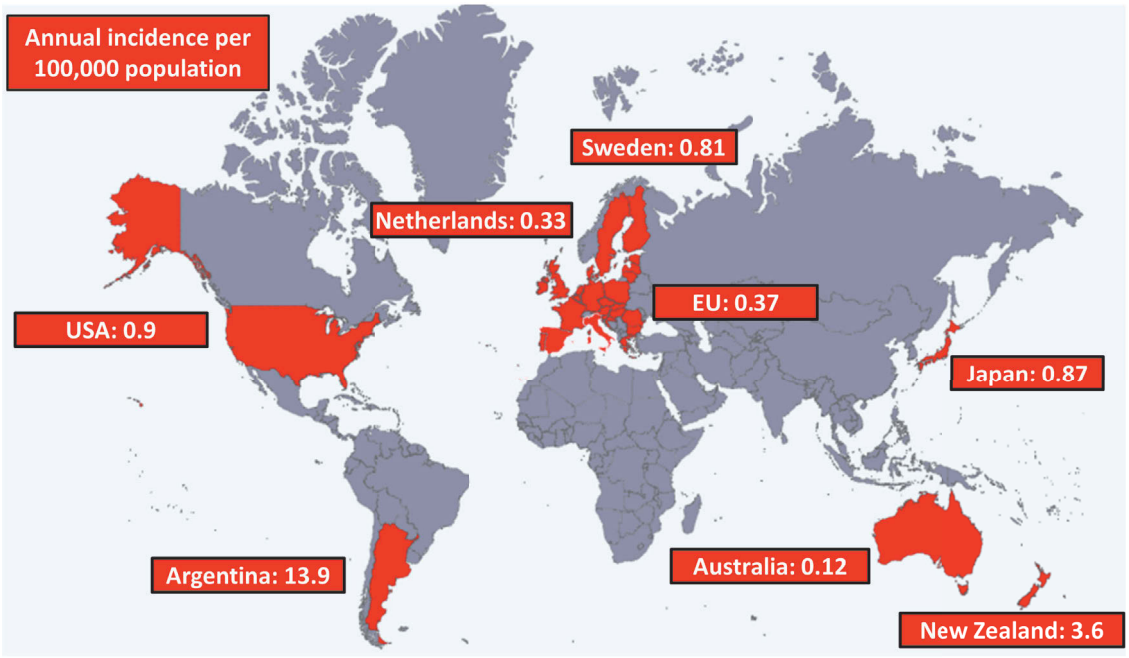


Figure 1.7: Annual incidence of STEC O157 related diseases. Figure adapted from Franz *et al*.,(2014).

These differences in the incidence of STEC O157 associated disease can be influenced by a number of factors. One of the most important factors is surveillance, different countries have different surveillance practices and this is imperative to keep that in mind when comparing the incidence rate (Rivas *et al.*, 2014). Social demographics is also a contributing factor where higher incidence of disease is observed in rural areas as opposed

to urban areas (Byrne *et al.*, 2015) as well as in places with a lower socio-economical demographic (Bentancor *et al.*, 2012) due to more frequent contact with cattle, hindered access to healthcare, and poorer awareness of proper food handling practices (Simonsen *et al.*, 2008; Whitney *et al.*, 2015). Apart from socio-economic factors, cultural factors can also further influence incidence rates with the case in point being Argentina, the country with the highest beef consumption per capita in the world (Rivas *et al.*, 2014). This geographical variation may also be influenced by genetic diversity in STEC O157 strains.

1.3.3. Diversity of STEC O157

One key biochemical phenotype of STEC O157 strains is its general inability to ferment sorbitol and lack of β-glucuronidase activity, unlike wild type *E. coli* (Feng, 1995), although some variants of non-motile STEC O157 have been shown ferment sorbitol and express β-glucuronidase (Karch *et al.*, 1993). Using MLEE on a panel of diarrhoeagenic *E. coli* strains inclusive of EHEC, EPEC and ETEC strains, it was documented that STEC O157 strains fall into phylogenetic group E, are clonal in nature and that STEC O157 was genetically closely related to the EPEC O55:H7 clone, which suggested that STEC O157 was derived from O55:H7 (Whittam *et al.*, 1993). This subsequently led to the development of a step-wise evolution model of STEC O157, which is now the currently accepted model of STEC O157 evolution (Figure 1.8) (Feng *et al.*, 1998). Briefly this model postulates that the most recent common ancestor of STEC O157 was an O55:H7 strain (A1 in Figure 1.8) that acquired a Stx prophage resulting in a Stx-producing O55:H7 strain (A2 in Figure 1.8). The next step was a somatic antigen conversion to O157 (A3 in Figure 1.8) where from this stage, two separate evolution events happened. The first branch resulted in the loss of motility (A4 in Figure 1.8) while the second branch led to the gain of another Stx prophage and the loss of both its ability to ferment sorbitol (A5 in Figure 1.8) and express β-glucuronidase (A6 in Figure 1.8) and that the clonal descendants of A6 account for the global spread of STEC O157. However, recently, the utilisation of Stx prophages as a marker in the evolution of STEC O157 has been argued against due to their mobile nature, and even with the exclusion of Stx prophages as a marker, there was no change to the step-wise evolution model (Kyle *et al.*, 2012).

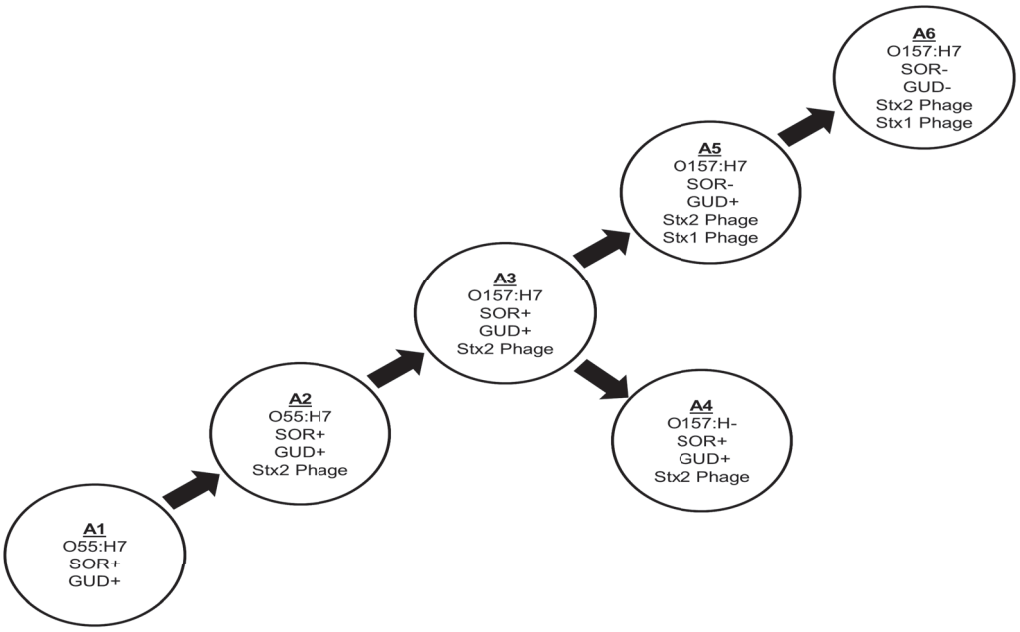


Figure 1.8: Stepwise evolution model of STEC O157. SOR and GUD is the abbreviation of Sorbitol and β-glucuronidase respectively. Figure adapted from Feng *et al.,*(1998).

While genetically highly similar, STEC O157 strains have been shown to have genetic heterogeneity, due to polymorphisms in conserved *E. coli* genes, as well as each isolate possessing different mobile genetic elements. Due to this heterogeneity, multiple typing schemes have been developed over the years to determine evolutionary relationships and population diversity of STEC O157 strains. These subtyping schemes range from traditional subtyping methods like pulse field gel electrophoresis (PFGE) (Bohm & Karch, 1992) , multilocus sequence typing (Noller *et al.*, 2003), multiple- locus variable number tandem repeat (MLVA) (Hyytia-Trees *et al.*, 2010) to STEC O157 specific subtyping schemes such as octamer based genome scanning assay (Kim *et al.*, 1999), lineage specific polymorphism assay (LSPA-6) (Yang *et al*., 2004) and single nucleotide polymorphism (SNP) assays (Manning *et al.*, 2008; Bono *et al.*, 2012; Jung *et al.*, 2013).

Studies have shown that genetic heterogeneity of STEC O157 is associated with host distribution, geographical distribution and virulence. Typing by LSPA-6, which groups STEC O157 into three different lineages- lineage I (LI), lineage II (LII) and lineage I/II (LI/II) showed different geographical distribution of STEC O157 strains (Kim *et al.*, 2001; Lee *et al.*, 2011; Franz *et al.*, 2012; Mellor *et al.*, 2012). LII isolates are dominant amongst cattle while LI and LI/II are dominant in human population in the United States,

Netherlands, Japan and Canada while LI/II strains are predominant from strains isolated from both cattle and humans in Australia and Argentina (Kim *et al.*, 2001; Lee *et al.*, 2011; Franz *et al.*, 2012; Mellor *et al.*, 2012). One SNP assay, which categorised clinical strains into nine distinct clades, showed that hypervirulent strains were localised within a single clade and the authors suggested that these strains have acquired genetic factors that made them more virulent (Manning *et al.*, 2008). In addition, SNP assays have also been used to infer evolutionary relationships from supposedly unrelated STEC O157 strains of cattle or human origin due the non-random distribution of STEC O157 lineages, where either cattle or clinical isolates dominate (Bono *et al.*, 2012; Jung *et al.*, 2013).

1.3.4. Virulence factors of STEC O157 strains

The genomes of the STEC O157:H7 strain associated with the 1982 outbreak in Oregon, Michigan, USA and the STEC O157 strain associated with the 1996 outbreak in Sakai, Japan were published in the early 2000s (although the genome of EDL933 was only completed in 2014) (Hayashi *et al.*, 2001; Perna *et al.*, 2001; Latif *et al.*, 2014). The chromosome of STEC O157 is approximately 5.5 Mb, larger than that of the laboratory strains of *E. coli* and this is due to the carriage of stretches of unique DNA sequences (genomic islands) that are not present in laboratory strains (Hayashi *et al.*, 2001; Perna *et al.*, 2001). In addition to STEC O157 strains EDL933 and Sakai, three other outbreak STEC O157 strains TW14359 (2006 outbreak in USA), EC4115 (outlier of the 2006 USA outbreak) and Xuzhou21 (1999 outbreak in China) were completely sequenced (Kulasekara *et al.*, 2009; Eppinger *et al.*, 2011; Xiong *et al.*, 2012). The genomes of all five STEC O157 strains are highly similar to each other, with regions of differences localised to mobile genetic elements and gene polymorphisms which can be up to 1.8% of the total genome size (Hayashi *et al.*, 2001; Kulasekara *et al.*, 2009; Xiong *et al.*, 2012). In addition, all five STEC O157 genomes harboured three prominent virulence factors that are described below:

1.3.4.1. The STEC O157 specific plasmid pO157

The plasmid pO157 is a plasmid that is found in almost all (> 99%) of clinical STEC O157 isolates (Levine *et al.*, 1987; Ostroff *et al.*, 1989; Kaper *et al.*, 2004). The complete sequence of pO157 from both EDL933 and Sakai are both 92 kb in length (99% nucleotide similarity) and encodes several virulence factors (Burland *et al.*, 1998; Makino

et al., 1998). This plasmid carries genes that encode three enzymes including a serine protease, a catalase peroxidase and a metalloprotease which are implicated in mucosal haemorrhage, enhancing adhesion to host intestines and causing proinflammatory response respectively (Brunder *et al.*, 1996; Brunder *et al.*, 1997; Lathem *et al.*, 2002) This plasmid also encodes an enterohaemolysin which stimulates the host inflammatory response (Zhang *et al.*, 2012), a putative adhesin (Tatsuno *et al.*, 2001) as well as a type II secretion system (Schmidt *et al.*, 1997). While the genes encoded within pO157 have been studied to a certain extend, there are still currently no published studies correlating carriage of pO157 to severity of clinical disease.

1.3.4.2 Locus of Enterocyte Effacement pathogenicity island

Pathogens that can form attaching and effacing lesions (A/E lesions) in the intestinal tract are known as A/E pathogens (Moon *et al.*, 1983). STEC O157, along with other A/E pathogens, can form A/E lesions by binding intimately to the intestinal epithelial cells mediated by the type III secretion system (T3SS). Once the bacterium comes into contact with the intestinal epithelial cell, the translocator proteins form a translocation pore and effector proteins are injected into the cell subverting the cytoskeletal and signalling machinery of the cell (Schmidt, 2010). One of the secreted effector proteins is the translocated intimin receptor (Tir), which following injection, is translocated onto the surface of the host cell and acts as a receptor for intimin which is expressed on the surface of the bacterium (Jerse *et al.*, 1990; Freeman *et al.*, 2000). This binding also triggers a downstream cascade leading to the effacement of the microvilli and the formation of an actin rich pedestal directly underneath the bacterium (Frankel & Phillips, 2008).

The ability to form A/E lesions is conferred onto all A/E pathogens by the Locus of Enterocyte Effacement (LEE). First described in EPEC (Mcdaniel *et al.*, 1995), the LEE is a 35 kb pathogenicity island with the majority of the genes localised in five operons (LEE1 to LEE5) encoding the machinery for the T3SS and its regulators, chaperones, intimin, translocator proteins and effector proteins (Figure 1.9) (Wong *et al.*, 2011). In addition to the effector genes encoded on the LEE, genes encoding other T3SS effector proteins are found outside the LEE and are called non-LEE effectors (Deng *et al.*, 2004), the majority of which are encoded within prophages (Tobe *et al.*, 2006).

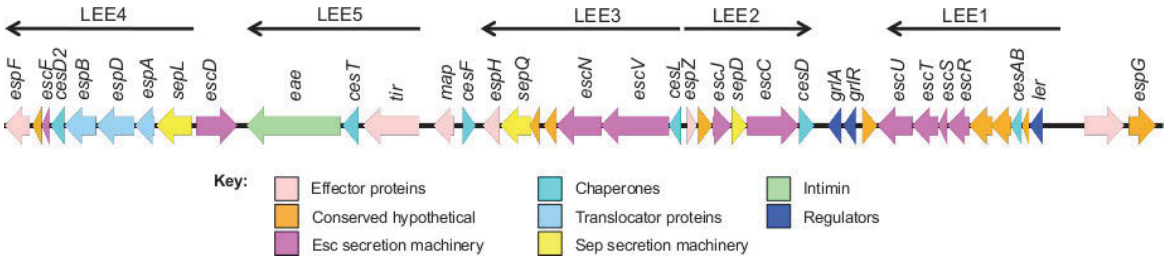


Figure 1.9: Graphical representation of the LEE pathogenicity island of STEC O157 strain Sakai. Genes are colour coded according to their function as shown in the key. The LEE1-LEE5 operons are also represented in the figure as black arrows. Figure was adapted from Wong *et al.*,(2011) and generated using Easyfig (Sullivan *et al.*, 2011)

The LEE is found in all A/E pathogens, not limited to EHEC and EPEC strains but also in rabbit diarrheagenic *E. coli* (Zhu *et al.*, 2001) the emerging pathogen *Escherichia albertii* (Ooka *et al.*, 2012), the mouse pathogen *Citrobacter rodentium* (Deng *et al.*, 2001) and have also recently been found in two poultry isolates of *Salmonella enterica* subspecies *salamae* serovar Sofia (Chandry *et al.*, 2012). The G+C content of the LEE is ~35% which is lower than the G+C content of all the bacterial genomes (~50%) in which it has been found. Previous investigations of the genomic regions flanking the LEE have reported that the LEE is flanked by either prophages (Perna *et al.*, 1998), chromosomal genes (Tauschek *et al.*, 2002) or IS elements (Perna *et al.*, 1998; Deng *et al.*, 2001). In addition, the LEE is found in different genomic locations in different bacterial strains (Rumer *et al.*, 2003). These observations have led to previous speculations that the LEE has been acquired by HGT, but the mechanism of transfer has yet to be elucidated (Deng *et al.*, 2001; Castillo *et al.*, 2005; Petty *et al.*, 2010).

1.3.4.3. Shiga toxin

The progression to HUS is associated with the level of production of Shiga toxin (Stx) by STEC O157 and other STEC strains (Karmali *et al.*, 1983; Neupane *et al.*, 2011). The cytotoxin Stx, first identified in *Shigella dysenteriae* serotype (Conradi, 1903; Neisser & Shiga, 1903), is a 6 subunit, A_1B_5 configured protein consisting of an enzymatically active A subunit, anchored via its C-terminus, onto the pentameric receptor binding B subunits (O'brien & Holmes, 1987) which binds to globotriaosylceramide (Gb3) (Lingwood *et al.*, 1987). In humans, expression of Gb3 is localised to the endothelial cells in the intestine, the endothelial and epithelial layer of the kidney, immune cells and the endothelial cells in the central nervous system (Lingwood, 1994; Van Setten *et al.*, 1996;

Schuller *et al.*, 2007; Obata *et al.*, 2008; Khan *et al.*, 2009). Stx can delivered to epithelial cells by different endocytic mechanisms, either clathrin-dependent, clathrin-independent, or caveolae-derived tabulation after binding to Gb3 (Bergan *et al.*, 2012).

Once internalised, Stx is transported to the Golgi apparatus and to the Endoplasmic reticulum via retrograde transport (Sandvig *et al.*, 1992), through which the A subunit is cleaved into two fragments, A1 (catalytically active) and A2 fragments (Garred *et al.*, 1995). From the endoplasmic reticulum, the A1 fragment is translocated into the cytosol where it inhibits protein synthesis and induces apoptosis, which has been previously shown to be a causal mechanism for renal injury (Lieberthal & Levine, 1996; Savill *et al.*, 1996; Tesh, 2010). In addition to cytotoxicity, Stx can also increase the expression of nucleolin and other surface receptors on the intestinal epithelial cell, which can bind to intimin and aid in mediating bacterial adhesion to the intestinal cell (Robinson *et al.*, 2006; Liu *et al.*, 2010).

There are two structurally similar but immunologically distinct Stx produced by STEC strains: Stx1 and Stx2. Based on nucleotide sequence differences, Stx1 has been further subtyped into Stx1a (henceforth referred to as Stx1), Stx1c and Stx1d while Stx2 can be further subtyped into Stx2a, Stx2b Stx2c, Stx2d, Stx2e, Stx2f and Stx2g (Scheutz *et al.*, 2012). The carriage of different Stx subtypes has been previously suggested to influence the pathogenicity of STEC O157 with subtypes Stx2a, Stx2c and Stx2d being more likely to cause the progression to HUS than the other subtypes. (Friedrich *et al.*, 2002; Orth *et al.*, 2007). In terms of toxin potency, Stx2a was found to be a more potent toxin than Stxl in animal models (Tesh *et al.*, 1993; Siegler *et al.*, 2003; Stearns-Kurosawa *et al.*, 2010), while although associated with the development of HUS, Stx2c has been shown to have lower potency when compared to Stx2a (Fuller *et al.*, 2011). The production of Stx amongst different STEC strains, carrying the same type and number of *stx* genes, also appears to vary which suggested that the amount of Stx produced is strain specific with STEC strains isolated from patients suffering severe human disease producing more toxin (Neupane *et al.*, 2011; Mellor *et al.*, 2015).

1.4. Shiga toxin-encoding phages

The first evidence of the conversion of laboratory non-STEC strains into Stx producers following co-incubation with STEC O26:H11 strain H19 via an undetermined mechanism, was documented in a study by Smith & Linggood (1971). Subsequently, studies by two independent groups published in 1983 showed that the genes encoding Stx in STEC O26:H11 strain H19 were carried by phages (henceforth referred to as Stx phages) (Scotland *et al.*, 1983; Smith *et al.*, 1983). The Stx phage isolated from the study by Smith *et al*., (1983), H-19B, was determined to be a siphovirus that carried the genes *stxA1* and *stxB1*, which forms an operon (*stx* operon) encoding the A subunit and the B subunit of Stx respectively (O'brien *et al.*, 1984; Willshaw *et al.*, 1987). To date, all Stx phages observed under the microscope were tailed phages with all three *Caudovirales* morphology types observed- podovirus (Plunkett *et al.*, 1999; Allison *et al.*, 2003; Asadulghani *et al.*, 2009; Beutin *et al.*, 2012; Sváb *et al.*, 2015), siphovirus (O'brien *et al.*, 1984; Strauch *et al.*, 2001) and myovirus (Muniesa *et al.*, 2000).

1.4.1. Stx phage genomes

Genome sequencing of the Stx1 phage of STEC O157 strain Sakai (VT1-Sakai) (Yokoyama *et al.*, 2000) and the Stx2a phages of STEC O157 strain EDL933 (933W) (Plunkett *et al.*, 1999) and Sakai (VT2-Sakai) (Makino *et al.*, 1999) revealed that the genomes of Stx phages have a similar organisation to that of phage lambda (Figure 1.10). This is the reason why Stx phages are referred to as lambdoid phages (Plunkett *et al.*, 1999). The *stx* operon is located downstream of the late anti-terminator *q* gene and upstream of the lysis module (Plunkett *et al.*, 1999). The location of the *stx* operon, along with the functional module order of the phage genome (Figure 1.10), has been found to be conserved in all subsequently sequenced Stx phages to date (Kuzminov, 1999; Recktenwald & Schmidt, 2002; Beutin *et al.*, 2012). Each Stx phage only carries one *stx* operon and the detection of multiple *stx* genes in a single STEC strain is due to the carriage of multiple Stx prophages (Hayashi *et al.*, 2001; Perna *et al.*, 2001; Kulasekara *et al.*, 2009).

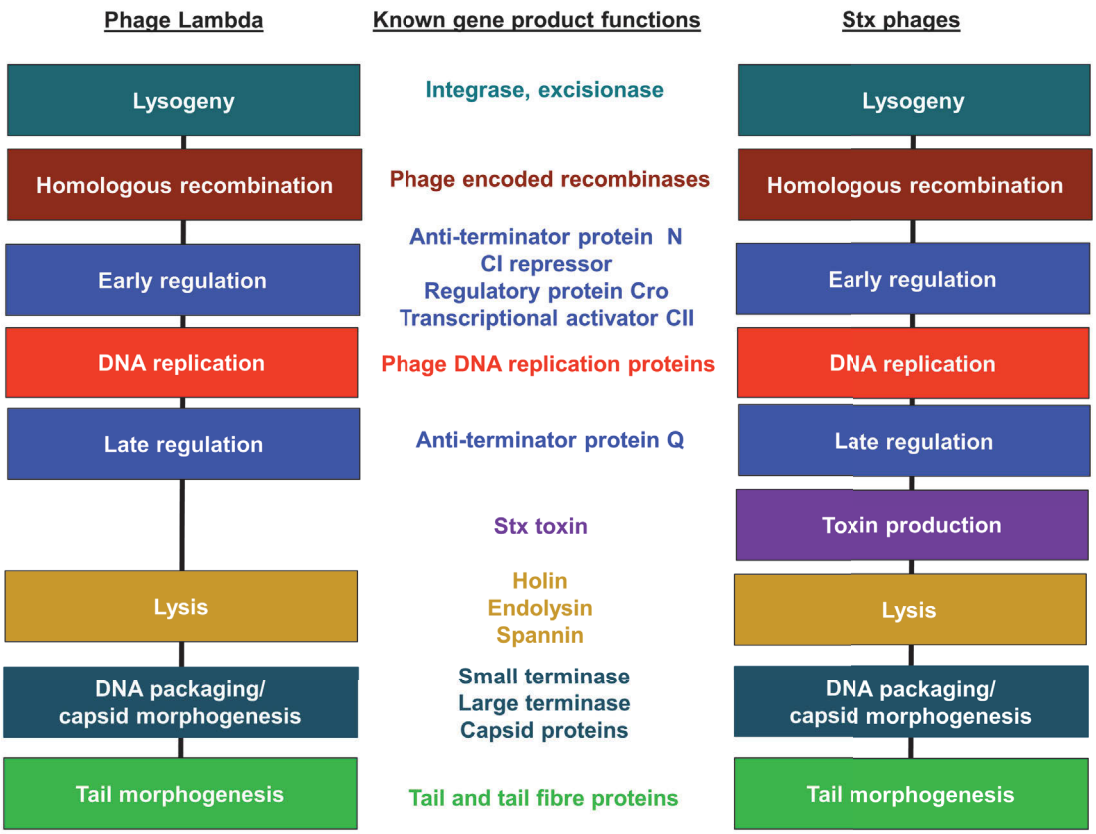


Figure 1.10: Organisation of the functional modules in the genomes of phage lambda and Stx phages. The order of modules shown "top-to-bottom" corresponds to "left-to-right" in the prophage genome. Figure adapted from Poteete (2001)

In phage lambda, the phage repressor maintains the state of lysogeny, but under inducing conditions, repression is alleviated and a regulatory cascade commences (Ptashne, 2004). This alleviation of repression allows the expression of the first anti-terminator protein N to be expressed and subsequently the second anti-terminator protein Q (Ptashne, 2004). The expression of the anti-terminator protein Q, in both phage lambda and Stx phages, allows downstream genes, which include the genes in the lysis and morphogenesis modules to be transcribed as the prophage switches to its lytic life cycle. In Stx phages, the expression of Q will also lead to Stx production (Neely & Friedman, 1998b; Neely & Friedman, 1998a).

To date, 17 Stx phage genomes have been completely sequenced ranging from 42 kb to 69 kb in size (Table 1.1). While all Stx phages share a common functional modular organisation, a multi locus characterisation scheme and comparative genomic analysis,

has shown variation of the genes within the modules and that no two Stx phages are identical to each other (Smith *et al.*, 2007b).

Table 1.1: Completely sequenced Stx phage genomes and Stx prophages (in complete bacterial genomes) available in the public databases (GenBank/EMBL/DDBJ).

Phages	Stx genotype	Size (kbp)	Virion morphology	Integration site	Accession number	Reference
Isolated Stx phages						
BP-4795	*stx*$_1$	57.9	Not determined	*yehV*	AJ556162	(Creuzburg *et al.*, 2005b)
CP-1639	*stx*$_1$	50.6	Not determined	Not determined	AJ304858	(Creuzburg *et al.*, 2005a)
Stx1CPI[a]	*stx*$_1$	59.9	Not determined	Not determined	AP005153	(Sato *et al.*, 2003b)
VT1-Sakai[a]	*stx*$_1$	47.9	Not determined	*yehV*	AP000400	(Yokoyama *et al.*, 2000)
HUN2013	*stx*$_1$	61.1	*Podoviridae*	*wrbA*	KJ909655	(Sváb *et al.*, 2015)
933W[a]	*stx*$_{2a}$	61.7	*Podoviridae*	*wrbA*	AF125520	(Plunkett *et al.*, 1999)
P13374	*stx*$_{2a}$	60.9	*Podoviridae*	*wrbA*	HE664024	(Beutin *et al.*, 2012)
Phi-191	*stx*$_{2a}$	61	Not determined	Not determined	KF971864	(Grande *et al.*, 2014)
24$_B$[a]	Δ*stx*$_{2a}$::*cat*	57.7	*Podoviridae*	Not determined	HM208303	(Smith *et al.*, 2012)
Min27[a]	*stx*$_{2a}$	63.4	Not determined	Not determined	EU311208	(Su *et al.*, 2009)
stx2CPI[a]	*stx*$_{2a}$	61.8	Not determined	Not determined	AP004402	(Sato *et al.*, 2003a)
stx2CPII[a]	*stx*$_{2a}$	62.7	Not determined	Not determined	AP005154	(Sato *et al.*, 2003b)
TL2011c	*stx*$_{2a}$	60.5	Not determined	*wrbA*	JQ011318	(L'abee-Lund *et al.*, 2012)
VT2-Sakai[a]	*stx*$_{2a}$	62.7	*Podovirdae*	*wrbA*	AP000422	(Makino *et al.*, 1999)
VT2-SA[a]	*stx*$_{2a}$	60.9	Not determined	Not determined	AP000363	(Miyamoto *et al.*, 1999)
2851[a]	*stx*$_{2c}$	57.2	*Siphoviridae*	*sbcB*	FM180578	(Strauch *et al.*, 2008)
P27	*stx*$_{2e}$	42.6	*Myoviridae*	*yecE*	AJ298298	(Recktenwald & Schmidt, 2002)
Stx prophages (in STEC strain)						
Unnamed (O157:H7 EC4115)	*stx*$_{2a}$	63.3	Not determined	tRNA-*argW*	CP001164	(Eppinger *et al.*, 2011)
Unnamed (O157:H7 EC4115)	*stx*$_{2c}$	57.2	Not determined	*sbcB*	CP001164	(Eppinger *et al.*, 2011)
Unnamed (O157 SS17)	*stx*$_{2a}$	61.9	Not determined	tRNA-*argW*	CP008805	(Cote *et al.*, 2015)
Unnamed(O157 SS17)	*stx*$_{2c}$	57.1	Not determined	*sbcB*	CP008805	(Cote *et al.*, 2015)
Unnamed (O157:H7 TW14359)	*stx*$_{2a}$	62.3	Not determined	tRNA-*argW*	CP001368	(Kulasekara *et al.*, 2009)
Unnamed (O157:H7 TW14359)	*stx*$_{2c}$	57.2	Not determined	*sbcB*	CP001368	(Kulasekara *et al.*, 2009)

Phages	Stx genotype	Size (kbp)	Virion morphology	Integration site	Accession number	Reference
Unnamed (O157:H7 Xuzhou21)	*stx*$_1$	47.9	Not determined	*yehV*	CP001925	(Xiong *et al.*, 2012)
Unnamed (O157:H7 Xuzhou21)	*stx*$_{2c}$	63.6	Not determined	*wrbA*	CP001925	(Xiong *et al.*, 2012)
ECO26_P06 (O26:H11 11368)	*stx*$_1$	55.5	Not determined	*wrbA*	AP010953	(Ogura *et al.*, 2009)
ECO103_P15 (O103:H2 12009)	*stx*$_1$	53.9	Not determined	*prfC*	AP010958	(Ogura *et al.*, 2009)
ECO103_P12 (O103:H2 12009)	*stx*$_{2a}$	62.6	Not determined	tRNA-*argW*	AP010958	(Ogura *et al.*, 2009)
ECO111_P11 (O111:H- 11128)	*stx*$_{2a}$	48.1	Not determined	*yecE*	AP010960	(Ogura *et al.*, 2009)
ECO111_P16 (O111:H- 11128)	*stx*$_1$	29.7[b]	Not determined	tRNA-*ssrA*	AP010960	(Ogura *et al.*, 2009)
Unnamed (O104:H4 2011C-3493)	*stx*$_{2a}$	68.7	Not determined	*wrbA*	CP003289	(Ahmed *et al.*, 2012)
Unnamed (O104:H4 2009EL-2050)	*stx*$_{2a}$	68.5	Not determined	*wrbA*	CP003297	(Ahmed *et al.*, 2012)
Unnamed (O104:H4 2009EL-2071)	*stx*$_{2a}$	68.5	Not determined	*wrbA*	CP003301	(Ahmed *et al.*, 2012)
Unnamed (O145:H28 RM13514)	*stx*$_{2a}$	62.5	Not determined	tRNA-*argW*	CP006027	(Cooper *et al.*, 2014a)
Unnamed (O145:H28 RM13516)	*stx*$_{2a}$	47.1	Not determined	*yecD-yecE*	CP006262	(Cooper *et al.*, 2014a)
Unnamed (O145:H28 RM12761)	*stx*$_{2a}$	47.1	Not determined	Not determined	CP007136	(Cooper *et al.*, 2014b)
Unnamed (O145:H28 RM12581	*stx*$_{2a}$	62.5	Not determined	Not determined	CP007133	(Cooper *et al.*, 2014b)

[a]: Stx phages originating from STEC O157 strains.
[b]: Cryptic prophage

1.4.2. Horizontal gene transfer and Stx phages

It has been hypothesised that the Stx2a podovirus 933W, isolated from STEC O157 strain EDL933, is a headful packaging phage based on the lack of a definite phage termini and a overrepresentation of sequencing reads over a region of several thousand base pairs in the phage genome (Plunkett *et al.*, 1999) as well as clustering together with other headful packaging phages in a phylogenetic analysis of the amino acid sequence of the large terminase (Casjens *et al.*, 2005). Recent experiments using phage lysates induced from a laboratory *E. coli* K-12 strain with a 933W prophage, to infect another laboratory K-12 strain have indicated that 933W may be capable of generalised transduction (Marinus & Poteete, 2013). However, these experiments lacked proper controls and this conclusion was based on the mobilisation of only one genetic marker. In order to confirm generalised transduction, additional experiments are required, focused on the transduction of several, well-defined genomic markers located in different parts of the bacterial genome, as performed for the classification of other generalised transducing phages (Budzik *et al.*, 2004; Petty *et al.*, 2006; Petty *et al.*, 2007; Battaglioli *et al.*, 2011). Only one Stx myovirus (Stx2e phage P27) has been sequence thus far and while the structural genes shared no similarity to the Stx podoviruses, the location of the stx_{2e} operon is conserved and that the Stx2e myovirus packages its genome via *cos* packaging (Recktenwald & Schmidt, 2002). There is also only one completely sequenced Stx siphovirus (Stx2c phage 2851) however its packaging strategy was not investigated (Strauch *et al.*, 2008).

In addition to the carriage of the *stx* operon, all Stx phages studied to date also carry other cargo genes. Some Stx phages carry homologs of the genes *bor* and *lom* harboured by phage lambda, which confer serum resistance and enhanced adhesion to mammalian cells respectively upon lambda lysogens (Barondess & Beckwith, 1995; Vica Pacheco *et al.*, 1997; Plunkett *et al.*, 1999). Another cargo gene carried by some Stx phages is *stk* which was predicted to encode a serine/threonine protein kinase although its function has yet to be elucidated (Tyler & Friedman, 2004). Other cargo genes detected within the genomes of Stx phages include a gene encoding a T3SS effector protein (Strauch *et al.*, 2008) and a gene encoding a functional esterase which was suggested to be required for substrate utilisation by STEC strains during human colon colonisation (Nubling *et al.*, 2014).

1.4.3. Induction of Stx phages

As for all temperate phages, the induction of Stx prophages are RecA-dependent, which was demonstrated in a study showing that the induction of Stx phages was not detected in *recA*- bacterial strains (Mühldorfer *et al.*, 1996; Fuchs *et al.*, 1999). However, a RecA independent mechanism was recently discovered whereby treating *recA* negative STEC with EDTA, which permeablises the bacterial outer membrane, led to the induction of resident Stx prophages (Imamovic & Muniesa, 2012). Stx phages are more readily induced when compared to other lambdoid phages with a slightly more than one fold increase in the rate of induction (Livny & Friedman, 2004). Inducing agents that causes stress and DNA damage which can lead to the induction of Stx phages include antibiotics like Mitomycin C, ^{60}Co irradiation (Yamamoto *et al.*, 2003), colicins (Toshima *et al.*, 2007), high temperature coupled with UV treatment (Yue *et al.*, 2012), hydrogen peroxide (Łoś *et al.*, 2010) and hydrostatic pressure (Aertsen *et al.*, 2005). It has also been observed that different Stx prophages have different induction frequencies, which suggests there are other genetic factors that could influence the induction of the Stx phages (Muniesa *et al.*, 2004a).

1.4.4. Persistence in the environment

Free, extracellular Stx phages have been detected in cattle faeces, rivers and wastewater indicating that the Stx phages are able to persist in the environment (Muniesa *et al.*, 2004a; Muniesa *et al.*, 2004b; Garcia-Aljaro *et al.*, 2009). The ability of Stx phages to persist in the environment is also demonstrated in its ability to remain infective, despite an observable drop in phage titre, after treatments with disinfectants as well as food processing conditions (Kajiura *et al.*, 2001; Rode *et al.*, 2011). The ability to persist in the environment also has implications of the transfer of the *stx* operon to other susceptible bacterial strains, which has been demonstrated in biofilms, food and water (Imamovic *et al.*, 2009; Solheim *et al.*, 2013).

1.4.5. Stx phage host range

Phages have been noted to typically possess a narrow host range (Ackermann *et al.*, 1978). However, this does not appear to hold true for Stx phages as, in addition to *E. coli*, Stx phages have been isolated from *Shigella sonnei* (Strauch *et al*., 2001) and *stx* genes have been detected in *Shigella dysenteriae* (Neisser & Shiga, 1903),

Citrobacter freundii (Schmidt *et al.*, 1993; Tschape *et al.*, 1995), *Enterobacter cloacae* (Paton & Paton, 1996), *Acinetobacter haemolyticus* (Grotiuz *et al.*, 2006) and *Enterococcus* spp (Casas *et al.*, 2011). Although, the association of *stx* genes with phage sequences was only shown in *Shigella dysentariae* (Yang *et al.*, 2005) and *Acinetobacter haemolyticus* (Grotiuz *et al.*, 2006), *stx* genes have never been found outside of phage genomes before, so it is likely that the *stx* genes detected in the aforementioned bacterial strains are all associated with Stx prophages. Previous host range studies of Stx phages have utilised a variety of techniques including either the detection of lysis on a lawn of bacteria, lysogeny or by absorption studies (Table 1.2). It is important to note, however, that host range studies based solely on lysis may give false negative results as it has been shown that Stx phages tend to preferentially enter the lysogenic life cycle (Islam *et al.*, 2012). Nonetheless, published host range studies have shown that host range varies from phage to phage and not only can Stx phages infect *E. coli* strains, some phages can infect other *Enterobactericeae* strains confirming that Stx phages have a broad host range.

Table 1.2: Host range studies of Stx phages

Stx phage	Isolated from	Host range	Reference
7888	*Shigella sonnei*	Able to infect *Shigella dysenteriae*, *Shigella boydii*, *Shigella. flexneri* and *E. coli* strains lacking the O antigen	(Strauch *et al.*, 2001) [a]
Multiple (n=30)	Multiple STEC (n=30)	Tested on laboratory *E. coli* strains C600 and DH5α, *Shigella sonnei* strain 866 and *E. coli* O157:H7 ATCC 43888 (Stx negative). Different phages showed different host range profile but 11 Stx phages were able to infect all strains tested.	(Muniesa *et al.*, 2004a) [b]
Multiple (n=11)	Free Stx phages from sewage	Tested on a range of bacterial strains inclusive of *Shigella dysenteriae* strain 500H, *Shigella boydii* strains 316 and 351, *Shigella sonnei* strains 866 and 635, *Shigella flexneri* strains 805 and 668, *E. coli* O111 strain 209, *E. coli* O26 strains 216 and 224, *E. coli* O157 strain ATCC 43888 and laboratory strains C600, DH5α and WG5. Different phage showed different host range profiles but one Stx phage was able to infect all strains tested.	(Muniesa *et al.*, 2004b) [b]

Stx phage	Isolated from	Host range	Reference
933W	STEC O157 strain EDL933	Able to infect majority of the ECOR collection of *E. coli* strains as well as commensal *E. coli* isolates from volunteers and patients in from the Cincinnati Children's Hospital Medical Center. Unable to infect *Citrobacter freundii*, *Klebsiella pneumonia*, *Klebsiella oxytoca*, *Klebsiella ozaenae*, *Enterobacter amnigenes*, *Enterobacter aerogenes*, *Serratia marcesscens*, *Acinetobacter Iwoffii* and *Edwardsiella* spp.	(Gamage *et al.*, 2004)[a]
24_B	STEC O157	Able to infect tested *Shigella flexneri*, *Shigella sonnei,* laboratory *E. coli* K-12 strains*, E. coli strains* isolated from animal and human sources, *Salmonella enterica* serovar Choleraesius strain SC-B67 and *C. rodentium* strain ICC168	(James *et al.*, 2001)[a], (Smith *et al.*, 2007a)[c]
Multiple Stx phages (n=79)	Multiple STEC strains (n=79)	Tested over a range of *E. coil* (pathogenic and laboratory K-12 strains), *Shigella dysenteriae*, *Shigella boydii*, *Shigella. Flexneri*, *Shigella sonnei*, *Enterobacter aerogenes* and *Citrobacter freundii*. Different phages showed different host range profile but three Stx phages were able to infect all strains tested	(Garcia-Aljaro *et al.*, 2009)[b]
Multiple Stx phages (n=5)	Multiple STEC strains (n=5)	Tested over a range of *E. coil* (pathogenic and laboratory K-12 strains) Different phages showed different host range profile but all five phages could infect all laboratory K-12 strains tested	(Tozzoli *et al.*, 2014)[b]

[a]: Host range tested by lysogeny
[b]: Host range tested by lysis
[c]: Host range tested by adsorption assay

Watarai *et al*., (1998) defined the bacterial surface receptor for one Stx2a phage (stx2CPI; morphotype unknown), which they showed utilised FadL, a beta barrel outer membrane ligand gated channel which uptakes long chain fatty acids. They also demonstrated a different Stx2a phage (stx2CPII; morphotype unknown) could utilise both FadL and LamB, another beta barrel outer membrane protein which facilitates the diffusion of maltose across the outer membrane, by the observation of lack of

plaque formation over a lawn of *E. coli* with either *fadL*, *lamB*, or both gene knocked out. A later study by Smith *et al*., (2007a) defined the surface receptor of a Stx2a podovirus, $24_{B,}$ to be BamA, an essential outer membrane protein and part of the BAM complex which is responsible for the assembly and insertion of outer membrane beta barrel proteins. The utilisation of BamA was identified through complementation of *bamA* into 24_B resistant *E. coli* and confirmed by phage binding assays with and without an anti-BamA antiserum Smith *et al.* (2007a). The authors also suggested that all podoviral Stx2a phages utilise the same surface receptor due to the carriage of a conserved tail spike protein gene Smith *et al.* (2007a). A later study conducted by Islam *et al*., (2012) utilising a different Stx2a podovirus (VT2-Sakai) concurred that BamA, not FadL or LamB, is the surface receptor for short tailed phages, based on the observation of lysis on a lawn of *E. coli* mutants (gene knockouts for *fadL* and *lamB* and overexpression plasmid for BamA as knockouts are lethal mutations). Based on these three studies conducted on surface receptors, it appears that different Stx phages utilises different receptors.

1.4.6. The impact of Stx prophages on their bacterial host

Stx prophages are also able to influence the expression of genes within the bacterial chromosome. A microarray study showed that lysogenisation of an Stx2 phage has an effect on gene expression where genes related to host energy production were suppressed and genes related to acid resistance and motility were up regulated (Su *et al.*, 2010). Stx phages have also been documented to have a role in the repression of the LEE, preventing the expressing of T3SS (Xu *et al.*, 2012). Xu and colleagues suggested that the repression of the T3SS by the Stx phage is a form of control whereby the T3SS-mediated bacterial colonisation of the host is somewhat dependent upon the Stx prophage resident within the bacterium. Carriage of multiple, different Stx prophages can also lead to the formation of novel combinations of Stx phages by recombination. This has been observed in STEC O157 strain Sakai where the stx_1 operon and its flanking region of the Stx1 prophage replaced the stx_2 operon and its flanking region of a Stx2 phage resulting in the formation of a chimeric Stx2 phage which carries the stx_1 operon (Asadulghani *et al.*, 2009; Sváb *et al.*, 2015). While prophage lambda is able to prevent infection by additional lambda phages via superinfection immunity (Ptashne, 2004), multiple isogenic Stx phages have been

shown to integrate into a single bacterial chromosome at different locations with each lysogenic event occurring at a higher frequency (Fogg *et al.*, 2007).

Stx prophages can also provide an anti-predatory defence to their host bacterium as the production of Stx has been shown to kill grazing protozoa. It has been documented that a Stx produced from *E. coli*, induced by the production of H_2O_2 by *Tetrahymena thermophila*, can kill this predatory protozoa via an altruistic suicide where the death of a few STEC cells by Stx phage induction produces enough Stx to kill off the protozoa while ensuring the survival of other STEC strains (Meltz Steinberg & Levin, 2007; Stolfa & Koudelka, 2012). The predation of bacterial strains by grazing protozoa is similar to how immune cells (phagocytocysis followed by production of H_2O_2 by neutrophils) attack bacteria in humans and it has been postulated that the release of Stx, which causes disease in humans, maybe due to this altruistic behaviour (Łoś *et al.*, 2012).

1.4.7. Insertion sites of Stx phages

The first identified insertion sites of Stx phages were the chromosomal genes *yehV* for Stx1 (Yokoyama *et al.*, 2000) and *wrbA* for Stx2a prophages (Makino *et al.*, 1999; Plunkett *et al.*, 1999) in STEC O157 strains EDL933 and Sakai. Subsequent studies have identified Stx phages in other genomic loci (Table 1.1). In addition, a study conducted by Serra-Morena *et al*., (2007) showed that Stx2a phages preferentially integrate into a genomic loci but if there is already another isogenic Stx phage integrated at that location, the Stx phage can integrate into a secondary insertion site. The *attP* sequence has been determined for Stx phages that insert into *yehV* (Yokoyama *et al.*, 2000), *wrbA* (Makino *et al*., 1999), *yecE* (Recktenwald & Schmidt, 2002), and *sbcB* (Strauch *et al*., 2008). In all sequenced STEC O157 strains to date, Stx phages have only been found integrated at *wrbA*, *yehV*, *sbcB* and tRNA-*argW* (Table 1.1).

The identification of Stx phages integrated at multiple genomic locations in STEC O157; has implications upon the currently accepted evolutionary model (see section 1.3.3). This model only accounted for Stx1 and Stx2a prophage integrated in *yehV* and *wrbA* respectively, but not for the Stx prophages that were integrated into tRNA-

argW and *sbcB*. Recently Kyle *et al*., (2012) argued that despite the clonal nature of STEC O157, phage insertion sites should not be used as genetic markers for evolution due to the mobile nature of phages. The influx of genomic data on Stx phages, particularly on insertion sites, shows that Stx phages are a driving force for evolution, contributing to diversity and generating different subpopulations of STEC O157.1.5. Australian STEC O157 and Stx prophages

1.5. Australian STEC O157 and Stx phages

As mentioned in section 1.3.3., STEC O157 are not a homogenous group but a heterogenic one. The earliest study showing the distinctiveness of Australian STEC O157 was published in 2002 showing that, cattle and clinical STEC O157 strains are non-motile and carry either the stx_{2c} gene, the stx_1 gene or both stx_1/stx_{2c} genes in combination (Fegan & Desmarchelier, 2002). The genomic heterogeneity and geographical distinction of Australian STEC O157 was demonstrated using PFGE on 73 clinical isolates spanning the southern hemisphere countries Argentina, Australia and New Zealand. Each country showed distinct restriction patterns with no overlap between countries, providing evidence that STEC O157 strains from Australia are genetically distinct from other international strains, including its closest geographical neighbour, New Zealand (Leotta *et al.*, 2008).

The idea of geographically distinct STEC O157 populations in Australia was further enforced from comparisons based on a panel of subtyping methodologies inclusive of lineage specific polymorphism assay, single gene polymorphism and MLVA, between STEC O157 strains from Australia, Argentina and the United States of America, showing that STEC O157 strains isolated from each country possess a distinct geographical genotypic profile, indicative that Australian STEC O157 strains are distinct from Argentina and the United States of America (Mellor *et al.*, 2012; Mellor *et al.*, 2013). Recently, a 48-plex SNP assay was performed on 148 STEC O157 strains from clinical and cattle isolates spanning Australia, Argentina and the United states and similar to the aforementioned studies, this SNP analysis showed that both cattle and clinical STEC O157 isolates fall into distinct SNP lineages with the predominant lineages not showing overlaps with other countries (Mellor *et al.*, 2015). In addition, the predominant lineages in Australia were shown to produce less toxin

than the predominant lineages of both the United States and Argentina which was attributed to the non-carriage of *stx*$_{2a}$ in the Australian strains (Mellor *et al.*, 2015).

Prophage content, not limited to the Stx prophages, is one factor that contributes to the genomic diversity of STEC O157 (Ogura *et al.*, 2006) and it is possible that Australian STEC O157 strains may possess different prophages to strains from other countries. In addition, it was also determined that the Stx1 prophages in Australian STEC O157 strains are integrated adjacent to tRNA-*argW* and the Stx2c prophages are integrated adjacent to *sbcB* (Mellor *et al.*, 2012). Prior to the study of Mellor and colleagues, Stx1 prophages had never been previously found integrated adjacent tRNA-*argW*, which suggested that there might be a unique population of Stx1 phage circulating in Australia (Mellor *et al.*, 2012). As genomic mosciasm in phages is believed to be caused by modular exchange of genes within a "pool" of phage genes (Botstein, 1980), this difference in insertion site utilisation suggested that the Stx prophages of Australian STEC O157 might be a geographically biased and genetically distinct population and could have evolved based on regional selective pressures. Therefore, an investigation into the Stx phages of Australian STEC O157 isolates may lead to insights that could possibly account for the lower incidence of clinical epidemiology in Australia.

1.6. Objectives of this study

As Stx phages provide the main virulence factor (Stx production) to STEC O157, investigations into the Stx phages could provide insights into the decreased virulence of Australian STEC O157 strains. Moreover, the interest in Stx phages is predominantly due to the ability to confer a toxin-producing phenotype, with Stx phage biology usually relegated to the sideline. The overall aim of this thesis is to investigate and gain insights into both the biology of of Stx phages and their role in phage-mediated HGT, as well as seeking to understand the role of Stx phages in differentially influencing pathogenicity of STEC O157 strains. To this end, investigations of Stx production, phage morphology, and transduction of genomic markers, coupled with comparative genomic analyses of the Stx phages from STEC O157 were performed to understand the role that Stx phages play in influencing pathogenicity and in turn, these findings could be utilised to implement more effective control measures against this pathogen in the future.

2. Materials and methods

2.1. Media, supplements and solutions

Constituents of the media and solutions used in this study are listed in Table 2.1. All were made with deionised water, unless otherwise stated. Where appropriate, solutions were sterilised by either autoclaving at 121^{o}C for 20 minutes or filtering through a 0.22 μm sterile filter (Satorius Stedium: 16534K). Chloroform used in this study was saturated with sodium hydrogen carbonate.

Table 2.1: Media and solutions

Media/solutions	Constituents
Media for bacterial culture	
Luria Broth (per litre)	25 g dehydrated Luria Bertani (Millers) Broth (Oxoid: CM0996B)
Sorbitol MacConkey agar (per litre)	51.5 g dehydrated Sorbitol MacConkey agar (Oxoid: CM0813B)
Solutions for phage work	
SM buffer	100 mM NaCl, 8mM $MgSO_4.7H_2O$ (Sigma: M2773) , 50 mM Tris.Cl pH 7.4, 0.01% (w/v) gelatin (Sigma: G9391)
Equilibrium CsCl buffer	0.75 g ml^{-1} CsCl (Sigma: 289329) in SM buffer
Lambda diluent	10 mM Tris.Cl pH 7.4 (ThermoFisher Scientific: 10977-015) , 8mM $MgSO_4$ (Sigma: S7899).
Solutions for DNA work	
100X DNase/RNase buffer	300 mM NaCl (Sigma: S9888), 100 mM Tris.Cl pH 7.5 (Quantum: A10861000), 10 mM EDTA pH 8.0 (Sigma: E5134), 0.2 mg ml^{-1} Bovine Serum Albumin (Sigma: A-7030), 20 mg ml^{-1} RNaseA (Sigma: R6513), 12 mg ml^{-1} DNaseI (Sigma: D5025).
20X TAE (per litre)	96.8 g Tris (Quantum: A10861000), 22.84 ml Glacial acetic Acid (Sigma: 320099), 40 ml 0.5M EDTA pH 8.0 (Sigma: E5134).
1 X TBE (per litre)	54 g Tris (Quantum: A10861000), 27.5 g Boric acid (Sigma: B0252), 20 ml 0.5M EDTA (pH 8.0).
Agarose gel mix (ethidium bromide)	2.0% (w/v) agarose, 1μg ml^{-1} ethidium bromide (Sigma: E-7637), 1X TAE

Media/solutions	Constituents
Agarose gel mix (Gelred)	0.7% (w/v) agarose, 0.125X Gelred (Biotium: 41003), 1X TBE
Gel loading buffer	18% (v/v) gel loading buffer (Sigma: G2526) in water

2.2. Bacterial strains, phages and culture conditions

Bacterial strains used in this study are listed in Table 2.2. Unless otherwise stated, all bacterial strains were grown at 37ºC in Luria broth (LB). Overnight cultures were routinely grown in 5 ml LB in 15 ml tubes in an orbital shaker (200 rpm). For solid media, 1.5% (w/v) agar (Oxoid: LP0011B) was used and for soft medium overlay (top agar) 0.35% (w/v) agar was added into the media. When required for selection, media was supplemented with the appropriate antibiotic as listed in Table 2.3. All bacterial cultures were stored in either 15% (v/v) glycerol or in protect preservers (Oxoid, Basingstoke, United Kingdom) at -80ºC, for long term storage.

Table 2.2: Bacterial strains used in this study

Strain	Relevant information	Reference/ source
Laboratory *Escherichia coli* strains		
Q358sm^R	Non-pathogenic K-12 derivative. (F-, *supF*, *hsdR*, *mcrB*, attphi80, lambda-) spontaneous streptomycin resistant mutant (mutation unknown)	(Karn *et al.*, 1980)
MG1655	Non-pathogenic K-12 strain. (F-, lambda-, *ilvG*-, *rfb*-50, *rph*-1). Sequenced *E. coli* K-12 reference strain	(Blattner *et al.*, 1997)
DH5α	Non-pathogenic K-12 strain. (F-, *endA1*, *glnV*44 *thi*-1, *recA*1, *relA*1, *gyrA*96, *deoR, nupG,* Φ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)U169, *hsdR*17(r_K^- m_K^+), λ–)	(Grant *et al.*, 1990)
Clinical STEC O157 strains		
Sakai	STEC O157:H7 strain associated with the 1996 Sakai outbreak (*stx*$_1$, *stx*$_2$, *eae*, *ehx*)	(Hayashi *et al.*, 2001)
EC3185	STEC O157:H- clinical strain, isolated from Brisbane, Australia. (*stx*$_1$, *eae*, *ehx*)	Queensland Health, Brisbane

Strain	Relevant information	Reference/ source
EC3206	STEC O157:H- clinical isolate, isolated from Brisbane, Australia. (*stx*$_{1,}$ *eae*, *ehx*)	Queensland Health, Brisbane
EC1810	STEC O157:H-, clinical strain isolated from New Zealand (*stx*$_{2a,}$ *eae*, *ehx*)	CSIRO, Coopers plains, Brisbane
EC3192	STEC O157:H-, clinical isolate (*stx*$_{2c,}$ *eae*, *ehx*)	Queensland Health, Brisbane
Non-clinical STEC O157 strains		
EC14	STEC O157:H-, isolated from lamb chops in Australia. (*stx*$_{1,}$ *stx*$_{2c,}$ *eae*, *ehx*)	CSIRO, Coopers plains, Brisbane
EC571	STEC O157:H-, isolated from calf faeces in Australia. (*stx*$_{1,}$ *stx*$_{2c,}$ *eae*, *ehx*)	CSIRO, Coopers plains, Brisbane
EC2441	STEC O157:H-, isolated from cattle faeces in Australia. (*stx*$_{2c,}$ *eae*, *ehx*)	CSIRO, Coopers plains, Brisbane
Derivatives of clinical STEC O157 strains and lysogens		
FNS8	STEC O157:H- EC1810 derivative, attenuated pathogen, (*stx*$_{2a}$, Δ*eae*::Tn9 *cat*, *ehx*)	CSIRO, Coopers plains, Brisbane
FNS9	STEC O157:H- EC1810 derivative, attenuated pathogen, (Δ*stx*$_{2a}$::Tn5 *kan*, *eae*, *ehx*)	CSIRO, Coopers plains, Brisbane
FNS65	STEC O157:H- EC1810 derivative, attenuated pathogen, (Δ*stx*$_{2a}$::Tn5 *kan* , Δ*eae*::Tn9 *cat*, *ehx*	This study
FNS45	STEC O157:H- EC3192 derivative lysogenised with phage 1810Stx2K, (*stx*$_{2c}$, 1810Δ*stx*$_{2a}$::Tn5 *kan*, *eae*, *ehx*)	CSIRO, Coopers plains, Brisbane
FNS72	STEC O157:H-, EC3192 derivative, EC3192 lysogenised with phage 1810Stx2K (*stx*$_{2c}$, 1810Δ*stx*$_{2a}$::Tn5 *kan*, Δ*eae*::Tn9 *cat*, *ehx*	This study
FNS 94	STEC O157:H-, attenuated pathogen, parent: EC3192 (*stx*$_{2c}$, Δ*eae*::Tn9 *cat*, *ehx*)	This study
FNS 105	STEC O157:H-, attenuated pathogen, parent: EC3192 (*stx*$_{2c}$, 1810Δ*stx*$_{2a}$::Tn5 *kan*, *eae*, *ehx*). Spontaneous rifampicin resistant mutant	This study

Strain	Relevant information	Reference/ source
EC3285	Q358sm^R lysogenised with Stx1 phage AU5Stx1 from EC3185	(Mellor *et al.*, 2012)
EC3286	Q358sm^R lysogenised with Stx1 phage AU6Stx1 from EC3206	(Mellor *et al.*, 2012)
EC3001	Q358sm^R lysogenised with phage 1810Stx2	CSIRO, Coopers plains, Brisbane
FNS76	MG1655 lysogenised with phage 1810Stx2K	This study
FNS78	DH5α lysogenised with phage 1810Stx2K	This study
Stx phages		
AU5Stx1	Stx1 phage from EC3185 (induced from K-12 lysogen EC3285)	This study
AU6Stx1	Stx1 phage from EC3206 (induced from K-12 lysogen EC3286)	This study
1810Stx2a	Stx2a phage induced from STEC O157 strain EC1810	This study
1810Stx2K	Stx2a phage, Δ*stx*$_2$::Tn5 *kan*, induced from STEC O157 strain FNS9	This study

Table 2.3: Antibiotics

Antibiotic[a]	Abbreviation	Solvent	Stock concentration	Working concentration
Chloramphenicol (C0378)	Cm	95% ethanol	30 mg ml^{-1}	30 μg ml^{-1}
Kanamycin (K1377)	Km	Water	30 mg ml^{-1}	30 μg ml^{-1}
Streptomycin (S-6501)	Sm	Water	50 mg ml^{-1}	50 μg ml^{-1}
Mitomycin C (M4287)	MMC	Water	0.5 mg ml^{-1}	1.0 μg ml^{-1}
Rifampicin (R-3501)	Rif	100% methanol	100 mg ml^{-1}	100 μg ml^{-1}

[a]: Sigma-Aldrich catalogue number in parenthesis

2.3. Bacterial phenotype test

2.3.1. Differentiation of STEC O157 from other bacteria

STEC O157 strains were differentiated from laboratory *E. coli* strains as described in March & Ratnam (1986). Briefly, a single bacterial colony was picked and streaked on Sorbitol MacConkey agar and incubated overnight at 37^{o}C, together with two control plates streaked with either STEC O157 strain Sakai or MG1655. After overnight incubation, STEC O157 were identified by colourless colonies, compared to the pink colonies of non STEC O157 strains.

2.3.2. Stx production

Total Stx production was assessed as described by Shringi *et al*., (2012) with modifications. Briefly, overnight bacterial culture was diluted 1:10 in 10 ml fresh LB and grown for 18 hours (37^{o}C) with shaking (200 rpm). If required, mitomycin C was added to a final concentration of 1 μg ml^{-1} after the first two hours of incubation then returned for the remaining 16 hours of incubation (37^{o}C, 200 rpm). Cultures were then centrifuged at 4000 x*g* for 10 minutes at 4^{o}C to sediment the bacterial cells and a 100μl aliquot of the supernatant was subsequently diluted 1:100 in sterile LB and followed by 1:2 dilution in sample diluent (Premier EHEC, Meridian Biosciences: 608096). Stx production was measured using a commercial ELISA kit (Premier EHEC, Meridian Biosciences: 608096) according to manufacturer's instructions and ELISA scores were obtained at dual wavelengths (450/630 nm) using a Victor X plate reader (PerkinElmer).

2.4. Phage methods

2.4.1. Isolation of Stx phages

Stx phages were isolated from their bacterial host according to the method described by Raya & Hébert (2008) with modifications. Briefly, 1 ml of an overnight bacterial culture was added to 9 ml fresh LB and grown for 2 hours with shaking (37^{o}C, 200 rpm). Mitomycin C was added to a final concentration of 1 μg ml^{-1} to induce the prophages and incubated for another 5 hours with shaking (37^{o}C, 200 rpm). Cultures were centrifuged at 4000 x*g* for 10 minutes at 4^{o}C. The resulting supernatant was syringe filtered first with a 0.45 μm filter (Satorius Stedium: 16533K) followed by a

0.2 μm filter (Satorius Stedium: 16534K) to remove bacterial debris. Lysates were stored at 4^{o}C until use.

2.4.2. Phage concentration and purification

2.4.2.1. Isopycnic ultracentrifugation

Stx phages were induced from their bacterial host as described in 2.4.1 with modifications to the induction volume where instead of 10 ml, 1L of culture was induced with mitomycin C. Following the five hour incubation, 100 ml aliquots of the culture were each dispensed into 250 ml polypropylene centrifuge bottles (Nalgene®; style 3120). Four millilitres of chloroform (Sigma: C2432) was added to each bottle and vortexed vigourously for 2 minutes. The culture was left to stand at room temperature for 30 minutes before being centrifuged at 4000 x*g* for 10 min at 4^{o}C and the resulting supernatant was filtered with a 0.45 μm pore filter (Satorius Stedium: 16533K). Solid NaCl (Sigma: S9888) and Polyethylene glycol 8000 (Sigma: P5413) were added to the filtrate, to a final concentration of 1M and 10% w/v respectively. Phages were left to precipitate overnight at 4^{o}C. The precipitated phages were then pelleted by centrifugation at 11,000 x*g* for 20 minutes at 4^{o}C and the supernatant was discarded. The resulting precipitate was resuspended in 1 ml SM buffer using a wide bore pipette. Isopycnic centrifugation of the phage lysate through CsCl (Sigma: C4036) equilibrium gradients and the subsequent dialysis of phage suspensions were performed according to Sambrook & Russell (2001).

2.4.2.1. Ultrafiltration

Phage lysates were prepared as described in 2.4.1. After syringe filtration, 8 ml of phage lysates were ultrafiltered using a filter unit with a 100 KDa membrane (Merck Millipore: 156145) at 5000 x*g* for 20 minutes at room temperature. The retentate was resuspended up to 500 μl with lambda diluent and treated with DNaseI (Final concentration: 100 units; Sigma: 4536282001) and lysozyme (Final concentration: 1 mg ml^{-1}; ThermoFisher Scientific: PIE90082) for 30 minutes at 37^{o}C. Two rounds of buffer exchange using lambda diluent was subsequently performed at 5000 x*g* for 20 minutes at room temperature. Ultrafiltered lysates were stored at 4^{o}C until use.

2.4.3. Phage quantification

2.4.3.1. Double agar overlay plaque assay

Phage lysates were quantified according to Kropinski *et al*.,(2008) with modifications. Briefly, phage lysates were serially diluted 1:10 in LB to 10^{-5}. For each dilution, 100 µl of indicator overnight bacterial culture and 10 µl of phage suspension was added to 4 ml of molten LB soft overlay media, mixed gently and poured onto a LB agar plate. The overlay media was left to set at room temperature for 30 minutes before incubating in a 37^{o}C incubator overnight. Phages were counted on those plates showing between 30 and 300 plaques, and phage titres were reported as plaque forming units (pfu) ml^{-1}.

2.4.3.1. Semi quantitative spot test

Spot tests were performed according to Mazzocco *et al*.,(2008) with modifications. Briefly, serial dilutions of phage lysates, up to 10^{-5} were prepared. One hundred microliters of overnight indicator bacteria culture was added to 4 ml of molten LB soft overlay media, mixed gently and poured onto a LB agar plate. The overlay media was left to set at room temperature for 30 minutes. Ten microliters of each phage dilution were subsequently dropped onto the set overlay media and the plate was incubated at room temperature for 30 minutes then overnight in a 37^{o}C incubator. Plaques were counted in spots showing between 3 and 30 plaques, and phage titres were reported as pfu ml^{-1}.

2.4.4 Test for lysogeny

2.4.4.1. Lysogenic conversion of bacterial strains on solid media

For phages capable of forming plaques on a lawn of recipient bacteria, tests for lysogeny were performed using the protocol described by Strauch *et al*., (2001) with modifications. Briefly, bacterial colonies from the centre of spot tests (as described in 2.4.3.1, but incubated up to a maximum of 48 hours) were picked and streaked to single colonies. Isolated colonies were restreaked three times on LB agar (supplemented with antibiotic selection where appropriate). The putative lysogens were validated by polymerase chain reaction (PCR) (as described in section 2.5.2) using primers targeting both phage and bacterial markers (Table 2.4).

2.4.4.2. Lysogenic conversion of bacterial strains in liquid culture

Tests for lysogeny of phages that could not form plaques on a lawn of recipient bacteria and were performed using an adaptation of the transduction protocol described by Petty *et al*., (2006). Briefly, 3 ml of phage lysate (prepared as per 2.4.1) was added to 10 ml of overnight recipient culture and incubated static at 37^{o}C for three days. The culture was then centrifuged at 4000 xg for 10 min at 4^{o}C. The supernatant was discarded and the bacterial pellet was resuspended in 300 µl of LB. For phages with a selectable antibiotic marker, the resuspended bacterial pellet was spread onto two LBA plates, containing appropriate antibiotic selection. Plates were incubated for up to 48 hours at 37^{o}C together with controls for spontaneous mutation (overnight recipient without addition of phage) and lysate contamination (phage only, no recipient strain). Antibiotic resistant single colonies were picked and streaked to single colonies three times on LB agar supplemented with appropriate antibiotics. For phages without a selectable antibiotic marker, serial dilutions of the resuspended pellet were performed and each dilution spread onto separate LBA plates and incubated overnight at 37^{o}C. Ten colonies on each plate showing single colonies were picked and each streaked to single colonies twice on LB agar. Putative lysogens were confirmed by PCR (as described in section 2.5.2)

2.4.5. Transduction assay

Phages were tested for their ability to transduce an antibiotic resistant genetic marker from donor to recipient strain (bacterial strains either harbouring wildtype *eae* or do not possess *eae*). Transduction was measured by production of Cm^{R} colonies using the basic transduction assay described by Petty *et al*., (2006) with modifications. Briefly, 3 ml of phage lysate induced from the appropriate donor strain (prepared as per 2.4.1), was added to a 10 ml overnight culture of recipient bacteria, mixed for 5 seconds and left static at 37^{o}C for three days. This was followed by centrifugation at 4000 x*g* for 10 minutes at 4^{o}C. The supernatant was discarded and pellet resuspended in 300 µl LB. One hundred and fifty microlitres of the suspension was spread onto each of the two LB agar plates supplemented with the appropriate antibiotics. These plates, along with controls for spontaneous mutation to antibiotic resistance and lysate contamination were incubated at 37^{o}C for up to 48 hours. Transduction was confirmed by screening antibiotic resistant colonies for the relevant genetic markers

by PCR. This basic transduction assay was further optimised by investigating best parameters for multiplicity of infection (M.O.I), addition of divalent ions prior to static incubation and length of static incubation as described in Chapter 5.

2.4.6. Phage electron microscopy

Squared mesh (200 squares per inch) copper grids (Ted Pella inc: G200HS) were coated with 1.1% formvar solution and dried overnight at room temperature. Grids were glow discharged under the UV lamp, in a biosafety class II cabinet for 25 minutes. Five microlitres of phage lysate was added onto the copper grid and allowed to adsorb onto the grids for one minute and washed three times with 5 µl of autoclaved deionised water and blotted with filter paper to remove excess water. The grids were stained with 5 µl of 2% (w/v) uranyl acetate for one minute and excess stain was removed by blotting with filter paper. Grids were air dried before storing in grid boxes. Grids with negatively stained phages were examined either with a Phillips JEOL JEM-1011 transmission electron microscope operating at 80 kV and captured with a Veleta 2k by 2k wide angle, side mounted CCD digital camera (Olympus) at the Centre of Microscopy and Microanalysis at the Queensland Bioscience Precinct, University of Queensland, Brisbane or examined with a technical staff using a FEI Tecnai G2 20 TEM operating at 200 kV at the Mark Wainwright Analytical Centre, University of New South Wales, Sydney. Grids were viewed between 80,000x and 150,000x magnification to visualise the phages. Measurements of phage heads and tails were done physically via printouts of the digital electron micrographs.

2.5. DNA work

2.5.1. Primers

Oligonucleotide primers used in this study are listed in Table 2.4. Custom primers for this study were designed using Primer3 (Untergasser *et al.*, 2012) and evaluated using oligoevaluator from Sigma Aldrich (http://www.oligoevaluator.com/) to reduce the chance of secondary structures being formed. Primers were ordered from either Invitrogen or Geneworks.

Table 2.4: Primers used in this study

Name	5'-3' sequence	Target	Reference
stx1F	ATAAATCGCCATT CGTTGACTAC	Forward primer used for detection of *stx*$_1$ in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
stx1R	AGAACGCCCACTG AGATCATC	Reverse primer used for detection of *stx*$_1$ in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
stx2F	GGCACTGTCTGAA ACTGATCC	Forward primer used for detection of *stx*$_{2a}$/*stx*$_{2c}$ in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
stx2R	TCGCCAGTTATCT GACATTCTG	Reverse primer used for detection of *stx*$_{2a}$/*stx*$_{2c}$ in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
eaeAF	GACCCGGCACAAG CATAAGC	Forward primer used for detection of *eae* in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
eaeAR	CCACCTGCAGCAA CAAGAGG	Reverse primer used for detection of LEE encoded *eae* in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
hlyAF	GCATCATCAAGCG TACGTTCC	Forward primer used for detection of pO157 plasmid encoded *ehx* in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
hlyAR	AATGAGCCAAGCT GGTTAAGCT	Reverse primer used for detection of pO157 plasmid encoded *ehx* in a STEC O157 multiplex PCR	(Paton & Paton, 1998)
int4045f	ACCATCGAGTAGG CGGTATG	Forward primer used for detection of *attL* of a tRNA-*argW* integrated Stx prophage	(Mellor *et al.*, 2012)
int1810R	ATTTCAGCAGGGC CAGAGTA	Reverse primer used for detection of *attL* of a tRNA-*argW* integrated Stx prophage	(Mellor *et al.*, 2012)
yfdCF	ACTGGAGCGATTT CATCTGG	Forward primer used for detection of *attR* of a tRNA-*argW* integrated Stx prophage	(Mellor *et al.*, 2012)

Name	5'-3' sequence	Target	Reference
phi1810stx2F	GGTTGAGCGGGATATGAAAA	Reverse primer used for detection of *attR* of a tRNA-*argW* integrated Stx prophage	(Mellor *et al.*, 2012)
sbcBF	ATTGTCGCGCTAAAGCTGAT	Forward primer used for detection of *attL* of a *sbcB* integrated Stx prophage	(Mellor *et al.*, 2012)
stx2cphiB	CAACGATGCTCGTTATGGTG	Reverse primer used for detection of *attL* of a *sbcB* integrated Stx prophage	(Mellor *et al.*, 2012)
stx2cphiA	GGACAACAGCGCACAGTAAA	Forward primer used for detection of *attR* of a *sbcB* integrated Stx prophage	(Mellor *et al.*, 2012)
sbcBR1	CGGGCTTGCAGTAAAAGACT	Reverse primer used for detection of *attR* of a *sbcB* integrated Stx prophage	(Mellor *et al.*, 2012)
stk_3206_F	GCATGGCCTCATTTCTGTTT	Forward primer used for resolving assembly error in the *stk* gene in AU6Stx1	This study
stk_3206_R	CGGAGGTAAGGGGATGCTA	Reverse primer used for resolving assembly error in the *stk* gene in AU6Stx1	This study
o_3206_F	AAACGGTGCAATGAAGCCAA	Forward primer used for resolving assembly error in the *O* gene in AU6Stx1	This study
o_3206_R	ATTCGGTTTTCTGGCTGATG	Reverse primer used for resolving assembly error in the *O* gene in AU6Stx1	This study
tail_3206_F	CAGGATGTTAAGGGGCTTGA	Forward primer used for resolving assembly error in a tail gene in AU6Stx1	This study
tail_3206_R	GCTATTTCATCCCGCCAATA	Reverse primer used for resolving assembly error in a tail gene in AU6Stx1	This study
760_3206_F	GAAAATGCGGAGGATGGTAA	Forward primer used for resolving assembly error in the	This study

Name	5'-3' sequence	Target	Reference
		last CDS of AU6Stx1 1	
760_3206_R	CGTTGCCACATTTTCAGAGA	Reverse primer used for resolving assembly error in the last CDS of 3 AU6Stx1	This study
Estx1-F	TGTTGCAGGGATCAGTCGTA	Forward primer used for detection of *stx*$_1$ by qPCR	(Gobert *et al*., 2007)
Estx1-R	TGAGGTTCCACTATGCGACA	Reverse primer used for detection of *stx*$_1$ by qPCR	(Gobert *et al*., 2007)
Estx2-F	CACATTTACAGTGAAGGTTGA	Forward primer used for detection of *stx*$_{2a}$/*stx*$_{2c}$ by qPCR	(Gobert *et al*., 2007)
Estx2-R	TTCAGCAAATCCGGAGCCTG	Reverse primer used for detection of *stx*$_{2a}$/*stx*$_{2c}$ by qPCR	(Gobert *et al*., 2007)
EtufA-F	TGGTTGATGACGAAGAGCTG	Forward primer used for detection of bacterial DNA control *tufA* by qPCR	(Gobert *et al*., 2007)
EtufA-R	GCTCTGGTTCCGGAATGTAA	Reverse primer used for detection of bacterial DNA control *tufA* by qPCR	(Gobert *et al*., 2007)
eae_del_testF	ATGATTACTCATGGTTGTTATACCC	Forward primer used to detect Δ*eae*::*cat*	CSIRO
eae_del_testR	TTATTCTACACAAACCGCATAGAC A	Reverse primer used to detect Δ*eae*::*cat*	CSIRO
stx2_del_testF	ATGAAGTGTATATTATTTAAATGGGTACTG	Forward primer used to detect Δ*stx*$_{2a}$::*kan*	CSIRO
stx2_del_testR	CAGTCATTATTAAACTGCACTTCAGC	Reverse primer used to detect Δ*stx*$_{2a}$::*kan*	CSIRO

2.5.2. PCR

Routine PCR was performed in a 25 µl reaction volume utilising 2 µl of DNA template with 2.5 µl DreamTaq^{TM} reaction buffer (ThermoFisher Scientific: EP0703), 0.2 mM (Final concentration) each of dNTPs (ThermoFisher Scientific: FMTR0181), 1 µM each (final concentration) of forward and reverse primers and 1 Unit of

DreamTaq™ DNA polymerase (ThermoFisher Scientific EP0703). PCR running conditions were 1 cycle of initial denaturation at 95°C for 3 minutes and followed by 30 cycles, each consisting of denaturation at 95°C for 30 seconds, annealing at an appropriate temperature (5°C below the Tm of primer pair) for an appropriate time (scaled according to 1 min per 1 kb product size), and extension at 72°C for 1 minute. After 30 cycles, a final extension at 72°C for 7 minutes was performed and the PCR product was held at 4°C.

For PCR products to be sequenced, reactions were performed in 25 µl reaction volumes with 2.5 µl platinum Taq DNA Polymerase High Fidelity reaction buffer (Invitrogen: 11304), 2 mM $MgSO_4$ (final concentration), 0.2 mM (Final concentration) of dNTPs (ThermoFisher Scientific: FMTR0181), 0.2 µM each (final concentration) of forward and reverse primers and 1 Unit of Platinum Taq DNA Polymerase High Fidelity (Invitrogen catalogue #: 11304) and 2 µl of DNA template.

To screen for STEC markers *stx*$_1$, , *eae* and *ehx* , 2 µl of extracted gDNA as described in section 2.5.4 was used as a template for a multiplex PCR, performed as described in Paton & Paton (1998) using the primers listed in Table 2.4

2.5.2.1. Colony PCR

A boiled cell lysate was prepared by adding a single bacterial colony into 100 µl of autoclaved deionised water and vortexed before being heated to 99°C for 10 minutes (Desmarchelier *et al.*, 1998). The lysed bacterial suspension was then centrifuged at 13,000 x*g* at room temperature for 3 minutes. Two microlitres of the resulting supernatant was used as DNA template for PCR (as described in 2.5.2).

2.5.2.2. Quantitative PCR

Measurement of Stx phage production using quantitative PCR (qPCR) was performed using a method adapted from Refardt (2012). Briefly, 200 µl of phage lysate (prepared as described in section 2.4.1) was treated with 2 µl 100x DNaseI/RNase buffer, then incubated at 100°C for 10 mins. Two microlitres of this DNA sample was used as a template for qPCR performed on a Viia™ 7 Real-Time PCR system (Applied Biosystems) with a 20 µl reaction mixture consisting of 10 µl SYBR Select Master Mix (Applied Biosystems: 4472908) and 200 nM (final concentration) of

forward and reverse primers. To generate a standard curve, the concentration of STEC O157 strain Sakai gDNA was measured using Nanodrop (Implen GmbH), then serially diluted to 1:100000 in nuclease free water (Ambion: AM9932). The standard curve, along with a no DNA template control was included in each qPCR run. The following PCR cycling conditions were used: 50°C for 2 minutes, 95°C for 2 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Gene copy number of either *stx*$_1$ or *stx*$_2$ was extrapolated automatically by the Viia™ 7 Real-Time PCR system from the cycle threshold in relation to the standard curve, using the following formula: Number of DNA copies = ((STEC O157 strain Sakai gDNA concentration ng μl^{-1})) x (6.022×10^{23})) / ((length of Sakai genome = 5,594,477bp) x (1×10^9) x (average weight of bp = 650 dalton).

2.5.2.4. Purification of PCR products

When required for Sanger sequencing of PCR products, following gel electrophoresis (section 2.5.3), PCR products were excised from the gel and processed using the QIAquick Gel extraction kit (Qiagen: 28704) according to manufacturer's instructions with an additional wash step with buffer PE.

2.5.3. Gel electrophoresis

One microlitre of DNA was added to 6 µl of PCR loading dye and loaded into either a 2% w/v agarose gel (in 1X TAE buffer) prestained with ethidium bromide or a 0.7% w/v agarose gel (in 1X TBE buffer) prestained with 0.25X v/v Gelred. For PCR products a 100 bp DNA ladder (ThermoFisher Scientific: 15628-019) was used and for gDNA a lambda DNA hinDIII digest ladder (New England Biolabs: N3012S) was used to aid size determination.. Gel electrophoresis was run at 75 volts for 90 minutes and gels were visualised in a UV transilluminator.

2.5.4. DNA extraction

Bacterial genomic DNA (gDNA) was extracted either using the DNeasy Blood & Tissue kit (Qiagen: 69504) for PCR and Sanger sequencing of PCR products, or Wizard® Genomic DNA Purification Kit (Promega: A1120) for creation of standard curves for qPCR, or UltraClean® Microbial DNA isolation kit (MO BIO Laboratories: 12224) for PacBio sequencing. DNA extractions were performed according to

manufacturers' instructions. The quality and quantity of extracted DNA was assessed using Nanodrop (Implen GmbH). Additionally, gel electrophoresis (see section 2.5.3) was performed on gDNA to be sequenced, to control for DNA degradation.

2.5.5. DNA sequencing

Sanger sequencing of PCR products was performed by the Australian Genome Research Facility (AGRF) using an AB3730 sequencer. PacBio sequencing using P2/C4 chemistry was performed by The McGill University and Génome Québec Innovation Centre, Montréal, Québec, Canada while PacBio sequencing using P4/C6 chemistry was performed by the Ramaciotii Centre for Genomics, University of New South Wales, Sydney, Australia.

2.6. Bioinformatics methods

2.6.1. Sequence quality control

The quality of sequences from Sanger sequencing were examined using 4 peaks (http://nucleobytes.com/4peaks/). Each chromatograph was manually inspected and low quality bases were removed using a text editor. Once free of mis-calls and poor quality sequences, the sequence data was used for downstream analysis (e.g. resolving of pseudogenes).

2.6.2. Genome assembly

454 sequencing was previously performed by the Australian Animal Health Laboratory and assembled using Newbler (Version 2.3) (Margulies *et al.*, 2005). STEC O157 strain TW14359 (accession number: CP001368), was used as a reference against which the assembled contigs were ordered using Mauve (Darling *et al.*, 2004). To check the assemblies, the 454 sequencing reads were mapped onto the ordered contigs using BWA mem (Li, 2013) with default settings and visualised using BamView (Carver *et al.*, 2013). Where mapped reads were found to span two adjacent contigs without any gaps and at least 90% of the reads mapping to the ends of the contigs spanned over the contig breaks, these contigs were joined together manually. Where required, pseudogenes which resulted from sequencing errors, were resolved using the data obtained from Sanger sequencing. Sequencing reads were

manually inspected to ensure that > 75% of individual base calls from the 454 reads agree with the consensus sequence generated by Newbler.

Pacbio sequencing reads were *de-novo* assembled using HGAP 3 (Chin *et al.*, 2013) with default settings. The assembled contigs were reordered against the reference strain STEC O157 strain Sakai chromosome and plasmids (accession numbers: BA000007, AB011549, AB011548) using Mauve (Darling *et al.*, 2004). Reads were mapped to the ordered contigs using BWA mem (Li, 2013) using 'pacbio' settings and visualised using Tablet (Milne *et al.*, 2013) to confirm the genome assembly.

2.6.4. Genome annotation

Automated genome annotation was performed using RAST (Aziz *et al.*, 2008). Artemis (Rutherford *et al.*, 2000) was used to collate the data from RAST and facilitate the manual curation of the automated annotation. BLASTp was performed on all predicted coding sequences (CDS) against the UniProt database UniProtKB and start codons were verified manually. Conserved protein motifs were predicted using the Conserved Domain Database (CDD) and Interpro (Hunter *et al.*, 2012).

2.6.5. Comparative genomic analyses

Published Stx phage genomes used for comparative genomic analysis were downloaded from GenBank are listed together with accession numbers in Table 1.3. Pairwise comparisons of the genomes were performed using BLASTn or tBLASTx and visualised using Artemis Comparison Tool (Carver *et al.*, 2005) Easyfig (Sullivan *et al.*, 2011) and BRIG (Alikhan *et al.*, 2011).

3. Characterisation of Stx1 phages from Australian *E. coli* O157 clinical isolates

3.1. Introduction

Australian STEC O157 strains are known to have reduced virulence and Stx1 prophages integrated at unique sites in their genomes, compared to STEC O157 strains from other countries (Mellor *et al.*, 2012). These findings led to the hypothesis that Australian Stx1 phages may have a unique genetic makeup that could influence the virulence of their bacterial host. In a previous study, two *E. coli* K-12 strains were lysogenised with Stx1 phages originating from Australian clinical STEC O157 strains and it was observed that these Stx1 phages integrated adjacent to tRNA-*argW* in both *E. coli* K-12 and their parent STEC O157 (Sim, 2010; Mellor *et al.*, 2012). DNA was also previously extracted from these Stx1 phages induced from their *E. coli* K-12 lysogens, sequenced and assembled into draft genomes (Sim, 2010). The aim of this study was to investigate these Stx1 phages further and characterise their ability to confer a toxin producing phenotype, phage morphology, attachment site, and their relationship to published Stx phage genomes to determine if there are any genetic factors encoded within the Stx1 phages that could shed insight to the reduced virulence of Australian STEC O157 strains.

3.2. Results

3.2.1. Host genetic background affects Shiga toxin production

The STEC O157 strains EC3185 and EC3206 along with their K-12 lysogens EC3285 and EC3286 (Table 2.4) (Mellor *et al.*, 2012) were tested for their ability to produce Stx and form infective phage particles (Figure 3.1). Toxin production and phage quantification were tested by ELISA and plaque assays (using Q358smR as an indicator) respectively. Overnight cultures of EC3185, EC3206, EC3285 and EC3286 not induced with MMC did not produce detectable toxin (Figure 3.1) but spontaneous phage production from the same overnight cultures, was detected by plaque assay ranging from 10^3 pfu ml^{-1} to 10^5 pfu ml^{-1} for each of the four strains (Figure 3.1).

When the four bacterial strains were induced with MMC, Stx production was detected from EC3185, EC3285 and EC3286 but not EC3206 (Figure 3.1). When the same overnight cultures were tested for phage production, phages were detected at titres ranging from 10^6 pfu ml^{-1} to 10^8 pfu ml^{-1} for each of the four strains, which is higher than that of cultures that were not induced with MMC (Figure 3.1) and consist with other published results showing that phage induction is enhanced with the addition of an inducing agent (Muniesa *et al.*, 2004a). Taken together, these results indicated that strains producing less than 2.33 x 10^6 pfu ml^{-1} of phages do not produce Stx within the limits of detection of the assay.

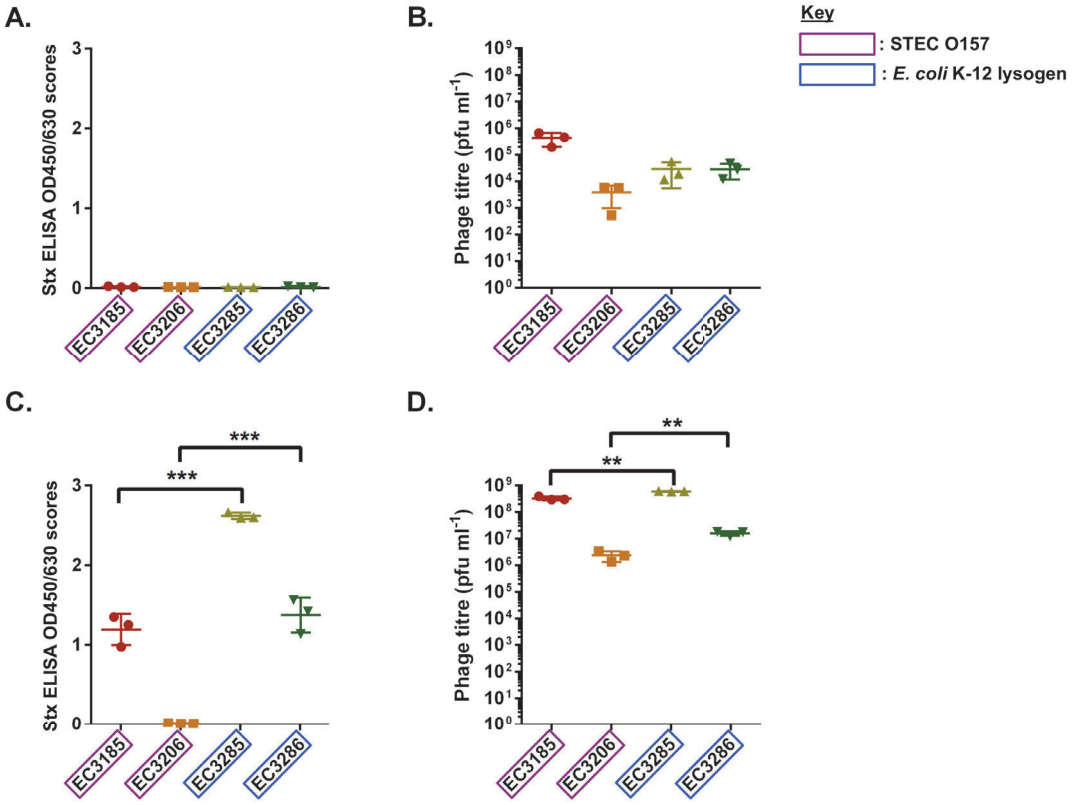


Figure 3.1: Comparison of Stx phage induction and Stx production in STEC O157 and K-12 lysogens. All individual data points (n=3) are shown together with the mean and standard deviations of the mean (bar and whiskers respectively). Significance is calculated by unpaired t-test demarcated by asterisks where ** represents $P \leq 0.01$ and *** represents $P \leq 0.001$. (A) Production of Stx from uninduced overnight cultures. (B) Phage titres from uninduced overnight cultures. (C) Production of Stx from MMC induced overnight cultures. (D) Phage titres from MMC induced overnight cultures.

The K-12 lysogens EC3285 and EC3206 were shown to produced significantly higher toxin and phage titres than their parent O157 strains EC3185 and EC3206 (Figure 3.1) This showed that the bacterial host genetic background can influence Stx production and that the increase in Stx production from the K-12 lysogen is due to higher phage induction.

3.2.2. Stx1 phages belong to the *Podoviridae* family

Stx1 phages were induced from the STEC O157 strains EC3185 and EC3285, and when observed using transmission electron microscopy (TEM) phages induced from EC3185 and EC3206, phages AU5Stx1 and AU6Stx1 respectively, both showed a single morphology, an icosahedral capsid with a short tail (Figure 3.2). No base plates or tail fibres were observed. In addition to the similarity in morphology, both Stx1 phages were also found to be the same size with both phages having a capsid diameter of 70 ± 3 nm and a mean tail length of 21 ± 3 nm (measurements are the mean of n=10 for each phage). The morphology observed allowed the classification of the two Stx1 phages into the order *Caudovirales* and family *Podoviridae.* When compared to the morphology of other, well-characterised podoviruses (T7, P22, phi29) (Casjens & Molineux, 2012), the virions of both AU5Stx1 and AU6Stx1 are highly similar to the virions of phage T7 (icosahedral head with a single short tail with no visible emanating tail fibres) (Steven & Trus, 1986) in contrast to the morphology of P22 and phi29, which both have thick tail spikes emanating from the side of the short tail tube (Casjens & Molineux, 2012). The only previous study published on Stx1 phage morphology has shown that Stx1 phages (H-19B) are siphoviruses (Smith *et al.*, 1983). However the morphology of both AU5Stx1 and AU6Stx1 were found to be comparable, in both morphology and size to the Stx2a phages 933W (Plunkett *et al.*, 1999), 24_B (Allison *et al.*, 2003),VT2-Sakai (Asadulghani *et al.*, 2009) and P13374 (Beutin *et al.*, 2012).

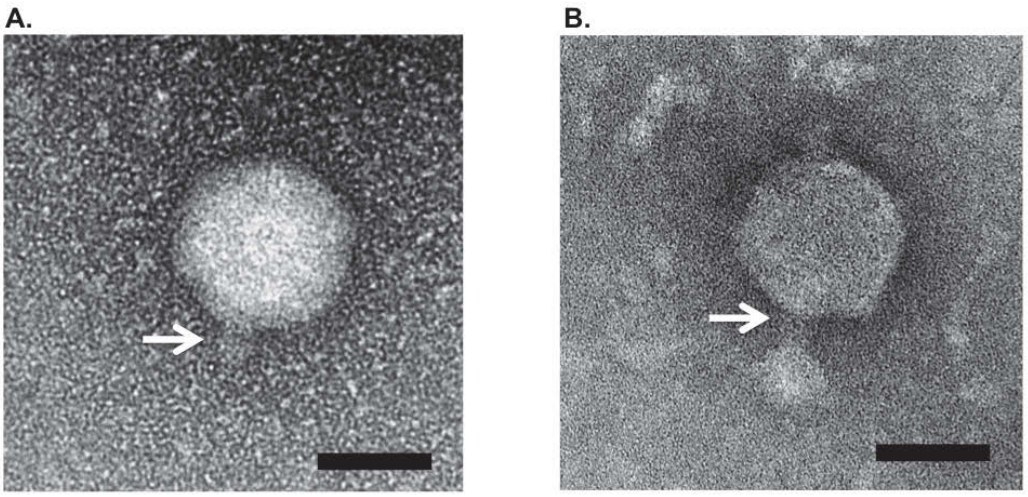


Figure 3.2: Morphology of the two Stx1 phages. Representative phage electron micrographs of (A) AU5Stx1 and (B) AU6Stx1 (phage tail appears to be adhered to some cellular debris) stained with uranyl acetate. The black bar represents 50 nm and the white arrows are pointing to the phage tail in each electron micrograph. TEM was performed using a FEI Tecnai G2 20.

3.2.3. Assembly of the Stx1 phage genomes

In a previous study, DNA was extracted from the Stx1 phages, AU5Stx1 and AU6Stx1 induced from the K-12 lysogens EC3285 and EC3286 respectively and DNA was sequenced using 454 single-end pyrosequencing and assembled (Newbler version 2.3) (Sim, 2010). The sequencing and assembly statistics of the two DNA samples are listed in Table 3.1.

Table 3.1: Sequencing and assembly statistics for phages AU5Stx1 and AU6Stx1.

	DNA Samples	
	AU5Stx1	AU6Stx1
Number of reads	15,859	5,046
Mean read length	388	271
Number of contigs	64	24
Total size of assembled contigs (bp)	129,948	65,749
Mean read depth[a]	45.67	16.26

[a]: Sequencing reads were mapped to the concatenated assembled contigs and the read depth was calculated using the following formula= (number of mapped reads x mean read length)/assembly size.

The assembled contigs showed that both samples had assemblies larger than the expected size of a Stx phage genome of approximately 60 kb (Table 3.1). The contigs

were reordered against the genome of STEC O157 strain TW14359 [(Kulasekara *et al.*, 2009); accession number: CP001368], due to it having an Stx2a prophage integrated adjacent to tRNA-*argW*, and the reads mapped to the reordered contigs (as described in section 2.6.2). The read mapping profiles showed that the assemblies consisted of both high read coverage contigs and low coverage contigs (Figure 3.3). The 454 reads were then mapped onto the genome of STEC O157 strain TW14359 which revealed that the high read coverage contigs correlated to the Stx phage and the low coverage contigs correlate to either a 80 kb region downstream of the Stx2a prophage or regions elsewhere in the genome of TW14359 (Figure 3.3). It is possible that the contigs correlating to the bacterial genomic region downstream of the Stx1 prophage could be indicative of specialised transduction whereby upon phage excision, this adjacent region is packaged into phage capsids. However, proper controls for bacterial DNA contamination during phage DNA extraction were not performed (Sim, 2010), so it is also possible that these low coverage contigs could be due to bacterial DNA contamination.

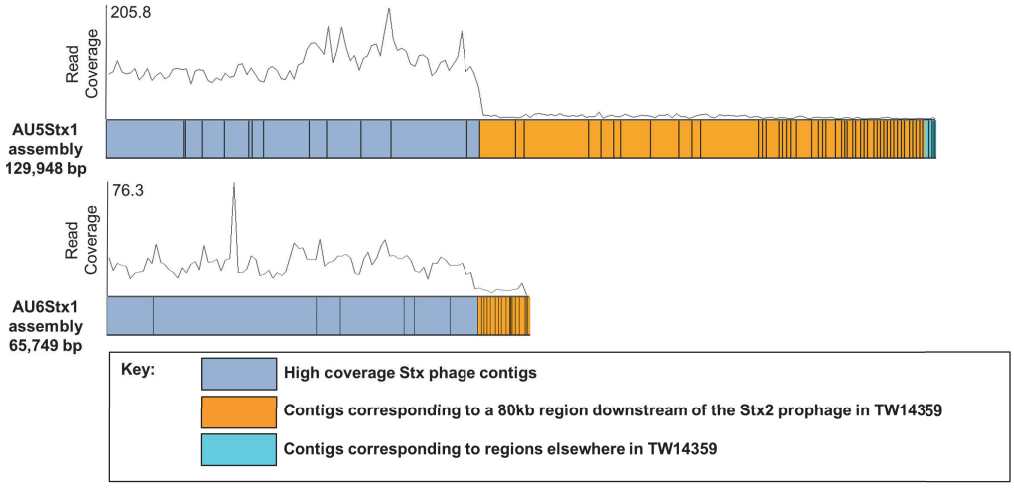


Figure 3.3: Read coverage over assembly. Each individual point on the line graph above the assembly represents the read coverage per base.

While the Stx1 phage genomes had assembled into contigs (21 for AU5Stx1 and six for AU6Stx1), mapped reads were observed to span across all of the contig boundaries of the ordered contigs. Cross-referencing of the reordered contigs with the "454ContigGraph.txt" Newbler output file confirmed that the contig breaks within each of the two Stx1 phages possessed no gaps between the contigs as more than 90%

of the reads at the contig boundaries spanned across the contig breakpoint into the adjacent contig (Figure 3.4). When interrogated further, the fragmentation of the Stx1 phage genomes into contigs was due to identical repeat sequences within the phages and from the low coverage contigs (Figure 3.4). Moreover, this analysis also revealed that the Stx1 phage genomes were circular and enabled the genomes of the 2 phages to be closed. The reads were mapped back onto the final complete phage assemblies and each individual base was checked to ensure that the consensus sequence agreed with the mapped reads (> 80% reads).

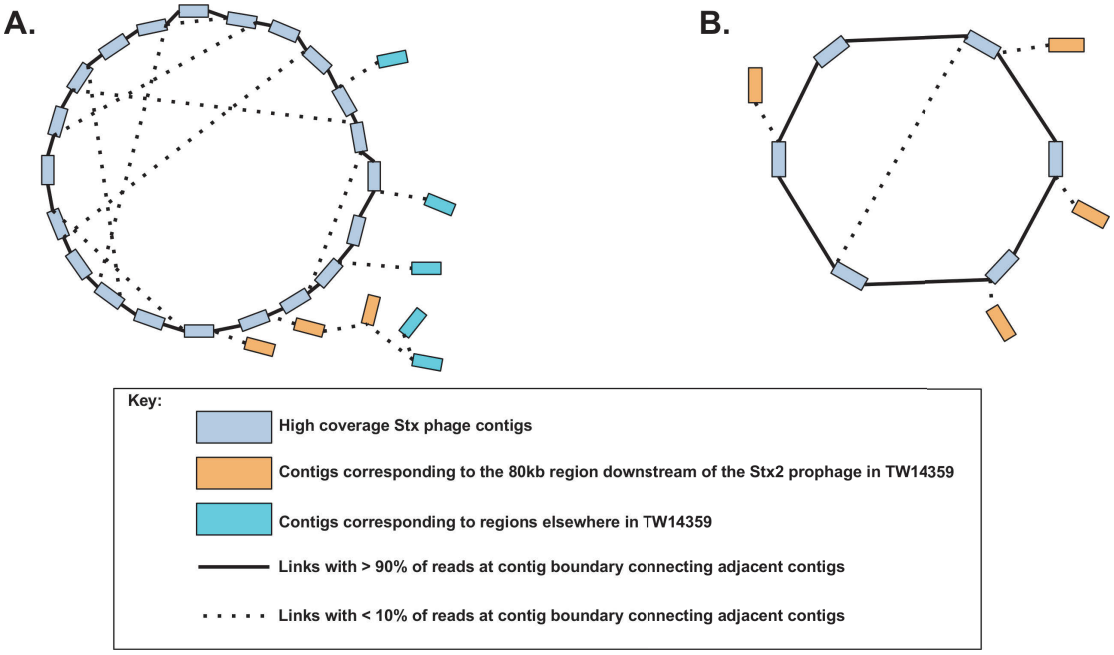


Figure 3.4: Graphical visualisation of the 454ContigGraph.txt output file generated by Newbler. Repeats within the phage genome and similar sequences in the low coverage contigs prevented the phage genomes to be assembled into a single contig for both (A) AU5Stx1 and (B) AU6Stx1.

The finalised assemblies were annotated automatically using RAST (Aziz *et al.*, 2008) followed by manual curation to ensure all putative coding sequences (CDSs) were predicted correctly. During manual curation, no pseudogenes were identified in AU5Stx1 but six pseudogenes each with a frameshift mutation were found in AU6Stx1. Frameshifts in four of these pseudogenes were found to be, within homopolymeric tracts and the mapped sequencing reads showed low read coverage and ambiguities. Homopolymeric tracts are common 454 sequencing errors (Margulies *et al.*, 2005), so PCR and Sanger sequencing was performed using

primers, designed 200 bp either side of all 4 frameshifts (Table 2.4). All four pseudogenes at homopolymeric tracts were confirmed to be 454 sequencing errors and following correction with the sequences obtained from Sanger sequencing, resolved into intact CDSs. The other two pseudogenes in AU6Stx1 had no evidence for 454 sequencing error and following examination of the reads mapped, were determined to be true pseudogenes.

3.2.4. The genomes of phages AU5Stx1 and AU6Stx1

The genome sizes of the two Stx1 phages were found to be within the expected range of an Stx phage which is between 40 kb to 60 kb (Table 1.1) and their G+C content is highly similar to that of their *E. coli* host at 50% (Blattner *et al.*, 1997; Hayashi *et al.*, 2001). A summary of the features of the two Stx phage genomes is listed in Table 3.2. and the predicted functional annotations of the Stx1 phages are provided in Appendix I and II for AU5Stx1 and AU6Stx1 respectively.

Table 3.2: Summary of the annotated phage genomes of AU5Stx1 and AU6Stx1.

	Phages	
	AU5Stx1	**AU6Stx1**
Total size (bp)	58,371	57,123
G+C content	49.77%	48.73%
Number of CDS	85	75
Mean read depth[a]	102.17	22.8

[a]: Sequencing reads were mapped to the final phage assembly and the read depth was calculated using the following formula= (number of mapped reads x mean read length)/size of phage assembly.

When the sequencing reads were mapped backed onto the annotated phage genomes, multiple regions of increased coverage were observed (Figure 3.5). Each of these peaks were interrogated and most of these peaks were determined to be artefacts of 454 sequencing where multiple reads start at the same position in the genome but have different lengths due to artificial in-*vitro* amplification in the emulsion PCR step prior to sequencing (Gomez-Alvarez *et al.*, 2009). After accounting for sequencing artefacts, a trend of increased coverage was distributed over a 5 kb genomic region at the terminase genes in both Stx1 phages (Figure 3.5). This increase in coverage over this region was 2.1x and 1.4x higher than the mean coverage of the total genome for

both AU5Stx1 and AU6Stx1 respectively. This observation of increased read coverage over the terminase genes was previously observed when the Stx2a podovirus 933W was sequenced and it is likely an indication of terminal redundancy at the phage genome termini (Plunkett *et al.*, 1999). In addition, the amino acid sequence of the large terminase, which can be used to predict packaging mechanism (Casjens *et al.*, 2005), of AU5Stx1 and AU6Stx1 has 99% amino acid sequence identity over its entire length to the large terminase of 933W [which has been showed to cluster with other headful packaging phages (Casjens *et al.*, 2005)]. The high level of amino acid sequence identity of the large terminase to that of phage 933W, coupled with the increased read coverage over the terminase genes, suggest that the genomes of both AU5Stx1 and AU6Stx1 are circularly permuted and terminally redundant, and that they may be headful packaging phages. As the exact phage termini sequence of AU5Stx1 and AU6Stx1 could not be ascertained in-*silico*, these two phages are represented as a linear genome in its prophage orientation (Figure 3.5), as is standard for other published Stx phages including 933W (Plunkett *et al.*, 1999), VT2-Sakai (Makino *et al.*, 1999) and phage 24_B (Smith *et al.*, 2012).

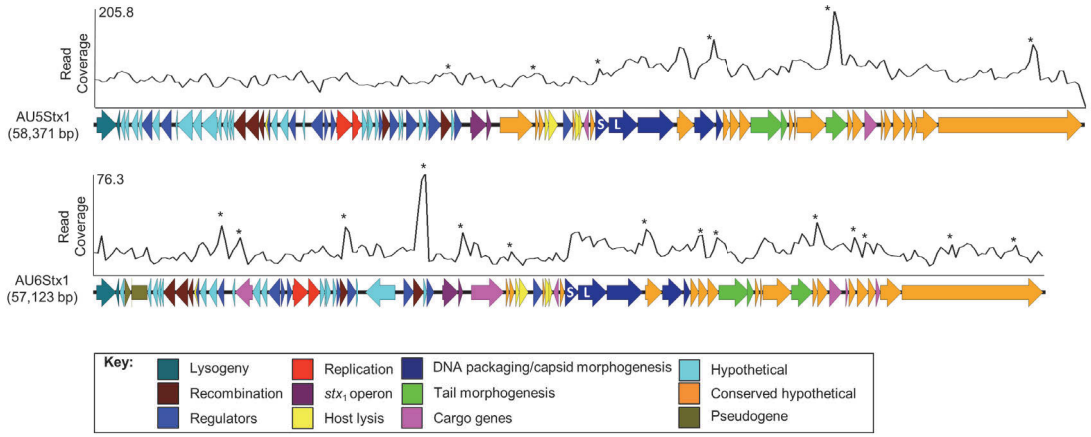


Figure 3.5: Stx1 phage genomes represented as linear genome in its prophage orientation. Phage coding sequences are colour coded as shown in the figure key. Sequencing coverage is shown as a line graph with the x-axis being the individual bases in the phage genome and the y-axis representing coverage. The terminase genes are marked in the figure (S: Small terminase subunit; L: Large terminase subunit) for both phages. Image generated using Easyfig (Sullivan *et al.*, 2011)

3.2.5. The genomes of AU5Stx1 and AU6Stx1 are highly similar to each other and to Stx2a phages

When compared pairwise, AU5Stx1 and AU6Stx1 were found to be highly similar to each other at the nucleotide level, especially in the region encoding the structural genes (Figure 3.6). Within this highly similar region, the stx_1 operons in both Stx1 phages were located, downstream of the *q* gene and upstream of the lysis modules (Figure 3.6), as found in other Stx phage genomes.

The genomic region consisting of the genes responsible for lysogeny and phage regulation were more varied in the two phages (Figure 3.6) and notably AU5Stx1 and AU6Stx1 were observed to have different repressor genes (Figure 3.6). In addition to maintenance of the lysogenic state, the repressor protein CI also confers immunity against superinfecting isogenic phages carrying identical *cI* genes (Ptashne, 2004). This indicated that AU5Stx1 and AU6Stx1 would each confer a different immunity profile to its host, meaning that their host is susceptible to infection from phages carrying different *cI* genes.

Apart from the difference in the *cI* genes, there are three other observable regions of differences. One region of difference is at the genomic location that harbours a gene that putatively encodes an adenine methylase and a phage anti-repressor (Figure 3.6). Modification of phage DNA is a countermeasure against bacterial defences, allowing the phage DNA to evade degradation once injected into the bacterial cell (Drozdz *et al.*, 2012) and the adenine methylase identified in AU5Stx1 is likely to also serve that function. The putative anti-repressor is likely to interact with the repressor and alleviate repression, switching the prophage into its lytic life cycle. In the synonymous region in the AU6Stx1 lie two pseudogenes, a truncated transposase and a truncated gene encoding a hypothetical protein.

Downstream of the gene encoding the anti-terminator N protein in both AU5Stx1 and AU6Stx1 lies another region of difference; AU6Stx1 harbours a gene (*stk*) which encodes a serine/threonine protein kinase. The *stk* gene has been previously shown to be harboured in the genomes of some Stx phages (Plunkett *et al.*, 1999; Smith *et al.*, 2012) but its function has yet to be elucidated (Tyler & Friedman, 2004)

Another notable difference between the two phages is the region encoding the hypothetical protein AU6STX1_350 in AU6Stx1 (Figure 3.6; Appendix I and II). No conserved domains were identified in the predicted amino acid sequence of AU6STX1_350. BLASTn searches against the NCBI nr/nt database revealed that this gene shared 99% nucleotide similarity over the entire length to a gene (locus tag: ECO26_3192) encoded in prophage P17 of *E. coli* O26:H11 strain 11368 (Ogura *et al.*, 2009) and 99% nucleotide similarity over 95% of a gene (locus tag: EAKF1_ch3950c) encoded in a cryptic phage remnant in *Escherchia albertii* KF1 (Fiedoruk *et al.*, 2014). In addition, the nucleotide sequence of AU6STX1_350 is also found in the genome (99% nucleotide similarity over entire length) of *E. coli* O104:H28 strains RM12581, RM12761, RM13514 and RM13516, located within a prophage, although it was not annotated in these four *E. coli* O104:H28 strains. Taken together, these evidences suggested that AU6STX1_350 is not a cargo gene but likely a gene having a phage function. In the synonymous region in the AU5Stx1 genome lies genes encoding a putative anti-repressor and a phage regulatory protein. The phage regulatory gene in phage AU5Stx1 may affect transcription of the downstream gene *ninG*, which encodes a protein that participates in homologous recombination (Tarkowski *et al.*, 2002). The presence of this anti-repressor, along with the other anti-repressor in another region of difference may be an explanation for the higher phage titre observed for AU5Stx1 over AU6Stx1, both in the K-12 lysogen and its parent STEC O157 strain, resulting in a higher Stx production (Figure 3.1).

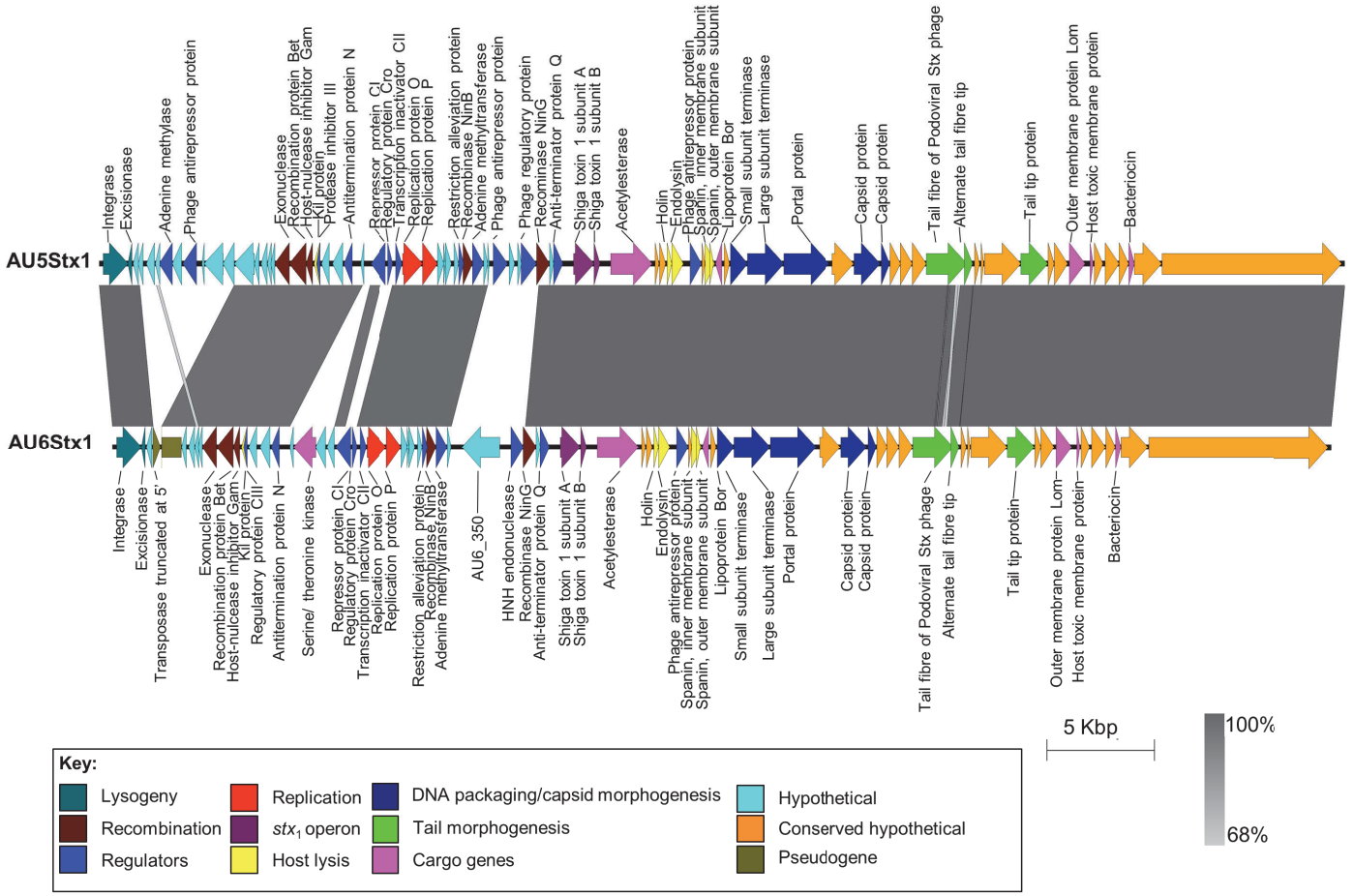


Figure 3.6: Pairwise comparison between AU5Stx1 and AU6Stx1. Genomic regions of nucleotide similarity between the genomes are highlighted in grey. Phage coding sequences are colour coded according to the figure key. Image was generated using Easyfig (Sullivan *et al.*, 2011)

The genomes of both AU5Stx1 and AU6Stx1 were found to be highly similar to all published genomes of Stx2a phages and two Stx1 phages (stx1CPI and HUN2013) as well as the non-Stx encoding prophages P04, of *E. coli* O111:H- strain 11128, and a prophage integrated adjacent to *ynfG* in the genomes *E. coli* O104:H4 strains 2011C-3493, 2009EL-2050 and 2009EL-2071. The prophage in the *E. coli* O104:H4 strains is identical in all three strains and the prophage from strain 2011C-3493 (henceforth known as prophage *ynfG*) is selected as a representative. High levels of nucleotide similarity were observed between the Australian Stx1 phages and the aforementioned phages especially over the virion morphogenesis modules, with varying levels of similarity in the region encoding genes for lysogeny and phage regulation (Figure 3.7). This comparative genomic analysis shows that there are no published genomes identical to either AU5Stx1 or AU6Stx1 and that regions of similarities are localised to either individual genes or functional modules, supporting the idea that all Stx phages are genetic mosaics of one another (Allison, 2007). The Stx phage with the highest level of nucleotide similarity to both AU5Stx1 and AU6Stx1 is the Stx2a phage 24_B (Figure 3.7).

In the genomic region encoding phage structural genes, there was no similarity to the Stx2c phage 2851 (siphovirus), the Stx2e phage P27 (myovirus) or phage lambda (siphovirus) (Hershey & Dove, 1983; Muniesa *et al.*, 2000; Strauch *et al.*, 2004). The difference in phage morphology between these three phages to both AU5Stx1 and AU6Stx1 is the likely reason for the lack of similarity. No morphological data have been previously documented for the Stx1 phages in the comparison group except for HUN2013 which is a podovirus (Sváb *et al*., 2015). Similarly, not all Stx2a phages in this comparison group have morphological data with only 933W (Plunkett *et al.*, 1999), 24_B (Allison *et al.*, 2003) and P133474 (Beutin *et al.*, 2012) classified as podoviruses. This similarity in morphology corroborates with the high level of genomic sequence similarity and it is highly likely that the other genetically highly similar Stx2a phages with no morphological data are also podoviruses (Figure 3.7).

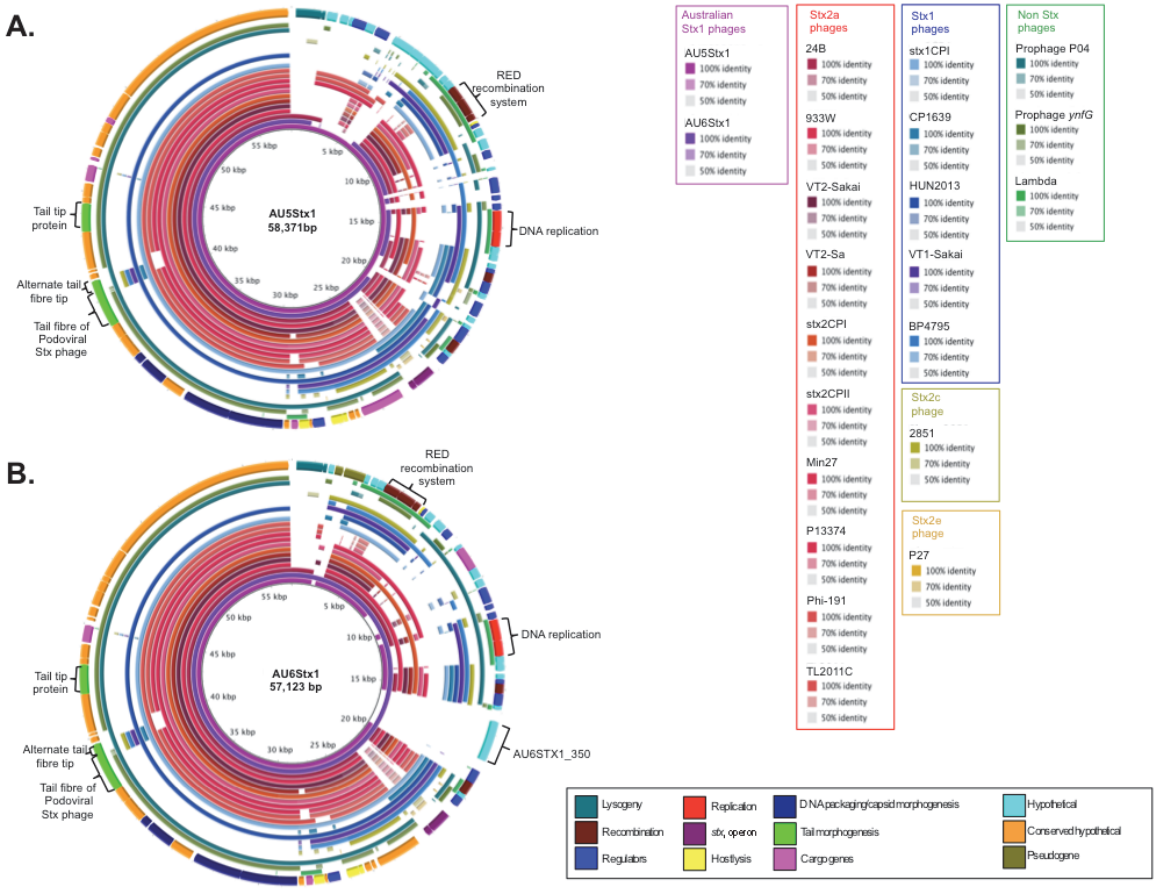


Figure 3.7: Comparison of Stx1 phages to published phages. The Stx1 phages (A) AU5Stx1 and (B) AU6Stx1 are presented as a reference genome in the centre with a circular representation of its genetic map. The outer most ring are the phage annotation, colour coded as per the key. Each of the coloured rings represents a different query phage genome, and the coloured regions represent BLASTn matches to the reference. Query genomes and their percentage identity are colour coded as listed in the figure key. Image was generated using BRIG (Alikhan *et al.*, 2011).

Included in the multi-phage genome comparison is the well-characterised phage lambda (Hershey & Dove, 1983). While no nucleotide similarity was observed in the structural genes, the lambda RED recombination system and DNA replication module both showed high similarity to the two Australian Stx1 phages (Figure 3.7). In phage lambda, the module making up the RED recombination system can promote homologous recombination of dsDNA between any regions of similar identity with as little as 40 bp sequence similarity (Poteete, 2001). This module consists of three genes, *exo*, *bet* and *gam* which encode a 5' – 3' exonuclease, a single stranded DNA annealing protein and a protein which inhibits host exonuclease activity respectively (Kuzminov, 1999). Homologs of these genes are also found in the majority of other Stx phages (Figure 3.7) (Makino *et al.*, 1999; Plunkett *et al.*, 1999; Smith *et al.*, 2012). In-*vitro* experimentation on phage lambda have previously demonstrated that phage genomic mosaic formation is driven by phage encoded recombinases acquiring new DNA from other prophages in its host genome (De Paepe *et al.*, 2014). A similar speculation has been made for the RED recombination system encoded by the Stx phages to have a role in mediating recombination with other Stx and lambdoid phages, leading to mosaic genomes and the formation of novel combinations of Stx phages (Allison, 2007).

The initiation of DNA replication is encoded by the genes *o* and *p* in phage lambda which are required for phage DNA synthesis and hijacking host DNA polymerase for replication (Furth & Wickner, 1983). At the nucleotide level, the *o* gene of AU5Stx1 differed from AU6Stx1 by a single synonymous SNP. The *o* gene of phage lambda is 36 bp shorter than that of the Australian Stx1 phages and has 98% nucleotide identity to AU5Stx1 and AU6Stx1 respectively. Comparison between the Australian Stx1 phages and lambda also attributed this size difference to the difference in number of the origin of replication (*ori* iterons) which the replication protein O interacts with during the initiation of replication (Scherer, 1978). Phage lambda has four *ori* iterons (Tsurimoto & Matsubara, 1981) while both AU5Stx1 and AU6Stx1 each have six identifiable *ori* iterons identical to the *ori* iterons previously observed in phage 933W (Plunkett *et al.*, 1999). The *p* genes in both AU5Stx1 and AU6Stx1 are identical to each other and share 96% nucleotide identity to the *p* gene in phage lambda, over its entire length. These high levels of nucleotide similarity suggest that the replication machinery of the two Stx1 phages is similar to phage lambda where replication starts

via a circle to circle replication and transitions to a rolling circle replication (Furth & Wickner, 1983).

3.2.6. Integration of the Stx1 phages occurs at the 5' end of tRNA-*argW*

Integrases are responsible for the site-specific recombination of the infecting phage genome into the bacterial chromosome (Campbell, 1994) and the integrase (*int*) genes carried by AU5Stx1 and AU6Stx1 are identical to each other in nucleotide sequence and share 99% nucleotide similarity to the *int* carried by 24_B. BLASTn searches of the *int* carried by AU5Stx1 (as a representative) against the NCBI nr/nt database showed highly similar integrases of identical size were found in other *E. coli* strains with high nucleotide identity: O157:H7 strains TW14359 (95%), EC4115 (95%), SS17 (95%) SS52 (95%), *E. coli* O103:H2 strain 12009 (98%), *E. coli* O145:H28 strain RM12581 (98%) and RM13514 (98%). When the genomes of these seven strains were interrogated, the *int* genes in all seven strains were also found located adjacent to tRNA-*argW* and associated with their Stx2a prophage. This suggests that these seven prophages, together with AU5Stx1 and AU6Stx1 have the same insertion site.

Through both conserved protein domain searches and multiple sequence alignment to amino acid sequences of known tyrosine integrases (phages lambda, HK022, P22, HP1, L5) and serine integrases (phages phiC31, R4 and TP901) (Groth & Calos, 2004), it was revealed that the integrase of AU5Stx1 (as a representative) possesses a conserved tyrosine catalytic residue. These results indicate that the integrase carried by both AU5Stx1 and AU6Stx1 belong to the tyrosine family of integrases.

PCR analysis confirmed that both AU5Stx1 and AU6Stx1 are integrated adjacent to tRNA-*argW* in both the parent STEC O157 strain and in the K-12 lysogen (Figure 3.8) (Mellor *et al.*, 2012). To determine the attachment sites of AU5Stx1 and AU6Stx1, PCR products derived from using primers targeting the tRNA-*argW* left boundary and tRNA-*argW* right boundary (Table 2.4, Figure 3.8) on the genomic DNA of EC3185, EC3206, EC3285 and EC3286 from both the right boundary and the left boundary were sequenced. When the sequences were aligned, an identical 24 bp sequence 5'-GTCCTCTTAGTTAAATGGATATAA-3' was found in both the left

and right boundary PCR products, indicative of the *attL* and *attR* sites. An identical sequence was also detected once within the genomes of both AU5Stx1 and AU6Stx1 corresponding to the *attP*. In the bacterial genome, the 24 bp *attB* sequence was found to form the 5' end of tRNA-*argW*, and was identified as spanning the acceptor stem and the D-loop of the tRNA secondary structure (Figure 3.8). This same 24 bp sequence was previously documented as the *attB* for the Stx2a phage 24_B (Fogg *et al.*, 2007) but its identity being the 5' end of tRNA-*argW* was not previously described. The result shown here expands on the localisation of the insertion site and is the first description of integration into the 5' end of a tRNA gene by Stx phages.

The anti-codon carried by tRNA-*argW* is CCU and when checked against all *E. coli* strains in the Genomic tRNA database (Chan & Lowe, 2009), it was observed that there was only one copy of this tRNA in each *E. coli* genome. It has been previously experimentally shown that the DNA binding protein Fis (factor for inversion stimulation) binds to a site (Fis binding site) 81 bases upstream of tRNA-*argW* and it acts as a positive transcriptional activator for the expression of tRNA-*argW* to meet the demand for protein synthesis (Nilsson & Emilsson, 1994; Bradley *et al.*, 2007). The insertion of the Stx1 phage, although reconstituting the 5' end of tRNA-*argW*, disassociates the Fis binding site from tRNA-*argW*. In addition, no Fis binding site was located within the genomes of AU5Stx1 and AU6Stx1. Taken together, these results suggest that insertion of the Stx phages may prevent expression of tRNA-*argW*.

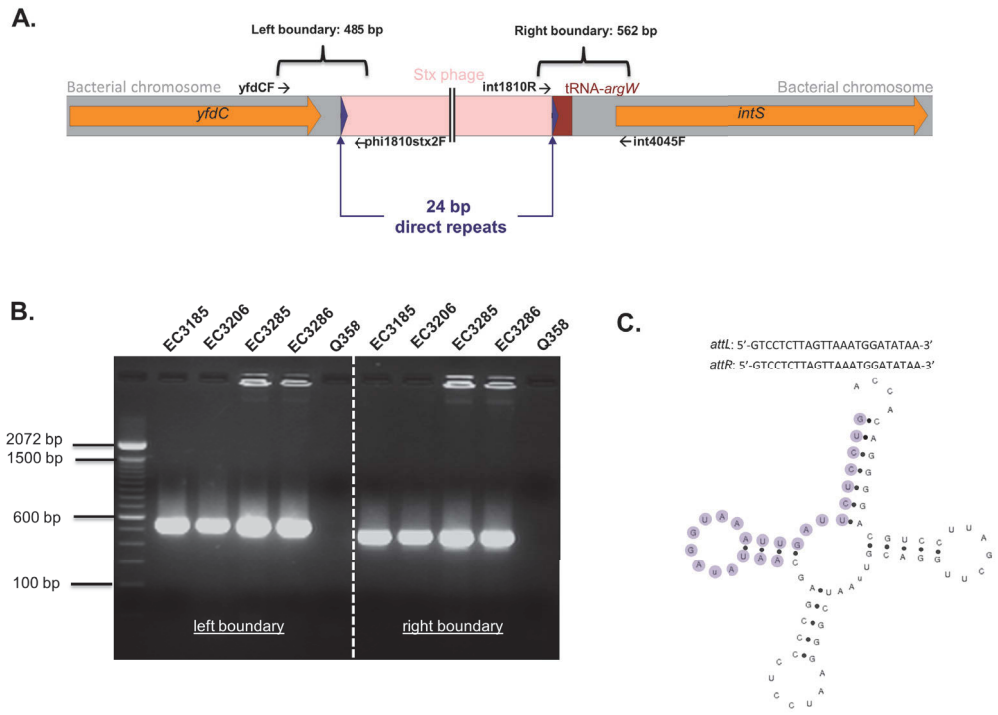


Figure 3.8: *att* sequence of the Stx1 phages AU5Stx1 and AU6Stx1. (A) The boundary primers to detect for insertion into tRNA-*argW* consist of two primer-pair set. Each primer pair has one primer within the integrated phage (pink rectangle) and one primer at the bacterial chromosome (grey rectangle). Expected product size are shown in the image (B) Confirmation of location of the Stx1 phages AU5Stx1 and AU6Stx1 adjacent to tRNA-*argW* by PCR using primers described in (Table 2.4). *E. coli* K-12 strain Q358, without a phage integrated at tRNA-*argW* was used as a negative control. (C) Location of 24 bp *att* within tRNA-*argW*. Image generated using tRNAscan-SE (Lowe & Eddy, 1997).

3.2.7. Host recognition by the Stx1 phages

Both AU5Stx1 and AU6Stx1 possessed an identical gene predicted to encode the tail tip protein (Figure 3.6) following a conserved domain search using the amino acid sequence acid of this gene which showed that this protein is predicted to have a conserved domain found in various central tail fibres of siphoviruses (Figure 3.9). This gene is present in all known Stx2a phage, stx1CPI and HUN2013 with 99% nucleotide similarity with nucleotide difference being synonymous SNPs, where the nucleotide change did not change the amino acid sequence (Figure 3.7). However, this CDSs in the aforementioned Stx phages were not marked up as encoding a tail tip protein.

This protein has weak similarity to the C-terminus of gp21 of phage N15 (Ravin *et al.*, 2000), which was previously observed to be similar to the tail tip protein of phage lambda (Ravin *et al.*, 2000) (Figure 3.9). The C- terminus of the central tail fibre in phage lambda is responsible for the attachment to the host receptor LamB (Roessner & Ihler, 1984) which suggested that the tail tip protein identified in both AU5Stx1 and AU6Stx1 is involved in host recognition. Host specificity of the tail tip protein is dictated by the protein C termini and since there is no amino acid sequence similarity at the C termini, between the tail tip protein of the Australian Stx1 phages and phage lambda, it is unlikely that LamB is the host receptor for AU5Stx1 and AU6Stx1

This central tail fibre was also predicted to have an immunoglobulin-like Fibronectin type III fold domain following an amino acid sequence search of Interpro (Figure 3.9). Such domains were previously observed to be primarily localised within phage tail proteins from an in-*silico* study (Fraser *et al.*, 2006). The function of these immunoglobulin-like domains has yet to be elucidated but a weak, non-specific binding between these immunoglobulin-like domains and carbohydrate molecules of the bacterial host cell is postulated to allow the phage to hold the bacteria in close proximity until it encounters its receptor (Fraser *et al.*, 2006). This could be a plausible strategy by which the Stx podoviruses may establish contact with the bacterial host cell prior to binding onto its target receptor with its tail fibre protein.

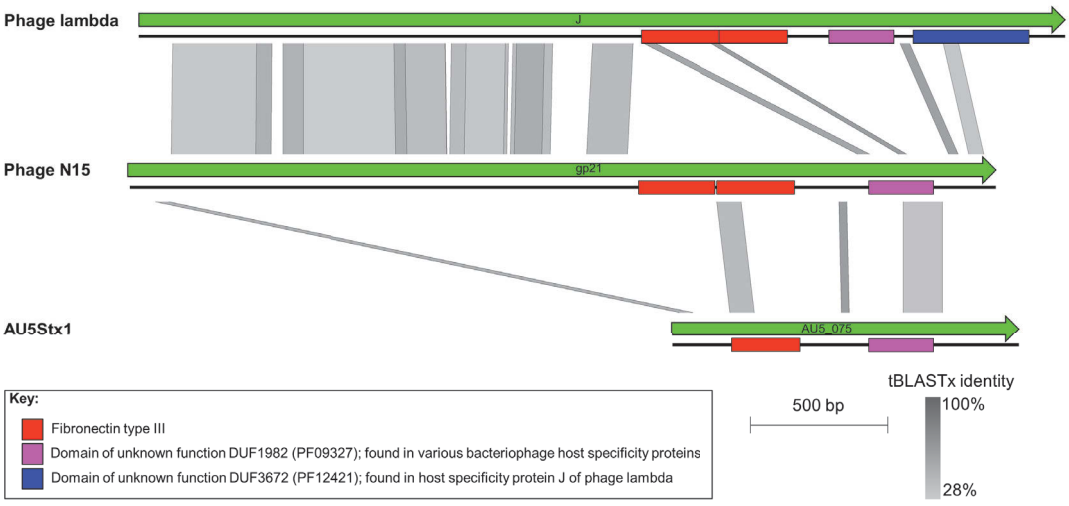


Figure 3.9: Comparison of the tail tip protein of phage lambda, phage N15 and AU5Stx1. tBLASTx comparison of the tail tip protein of the three phages scaled according to the grey bar. Image generated using Easyfig (Sullivan *et al.*, 2011).

Conserved domain searches of a protein encoded gene carried by AU5Stx1 (locus tag: AU5STX1_700) and AU6Stx1 (locus tag: AU6STX1_710), showed that it possessed a conserved side tail fibre domain. When compared (tBLASTx) to the characterised phages Lambda and Mu, a region of amino acid sequence similarity was observed at this conserved side tail fibre domain (Figure 3.10A). No similarity was detected at the N- or C-termini but this was not unexpected as the N-terminus of the tail fibre is attached specifically to its associated virion while the C-terminus contains the host recognition domain, which varies according to host range (Xue & Egan, 1995; Mitraki *et al.*, 2002). Apart from the conserved side tail fibre domain, a carboxypeptidase- like regulatory N-terminal domain, collagen triple helix repeats, and a host recognition C-terminal domain (cterm1) was also predicted (Figure 3.10). This gene in both AU5Stx1 and AU6Stx1shares 99% nucleotide identity to a gene in STEC O157 strain Sakai (locus tag: ECs1228), which has been previously shown to form a 94 nm homotrimeric protein due to the collagen triple helix repeats (Ghosh *et al.*, 2012), a structural characteristic of tail fibres in characterised phages (Smith *et al.*, 1998; Weigele *et al.*, 2003). No tail fibres were observed in the TEM study of these two Stx1 phages (as described in section 3.2.2) but this phenomenon could be similar to that of the podovirus T7 as the long thin tail fibres of T7 were also not visible under the electron microscope (Steven & Trus, 1986). A homolog of this gene was also previously identified in phage 24_B, which shared 99% nucleotide identity to the gene encoded by AU5Stx1 and AU6Stx1 and 42 bases longer (Figure 3.10). Despite this gene in AU5Stx1 and AU6Stx1 being 42 bases shorter, the conserved domains are intact when compared to phage 24_B, inclusive of the conserved side tail fibre domain (Figure 3.10.) The authors of the study suggested that this gene encode the phage tail spike protein (gene name: *tspS*) (Smith *et al.*, 2007a). However, no short stubby tail spike like appendages has been observed in the TEM micrographs of phage 24_B. In addition, all TEM micrographs of podoviral Stx phages to date, do not possess short stubby tail spike like appendages. This carriage of a conserved tail fibre domain coupled with the non-observation of short stubby appendages emanating from the tail tube (Figure 3.2) and the possibility of AU5Stx1 and AU6Stx1 to form a thin 94 nm tail structure based on a previous study on a homologous tail fibre protein (Ghosh *et al.*, 2012) suggested that the Stx podoviruses are likely to possess phage tail

fibres as opposed to tail spikes. As such this gene in both AU5Stx1 and AU6Stx1 is annotated as the tail fibres of podoviral Stx phages (*tfpS*)

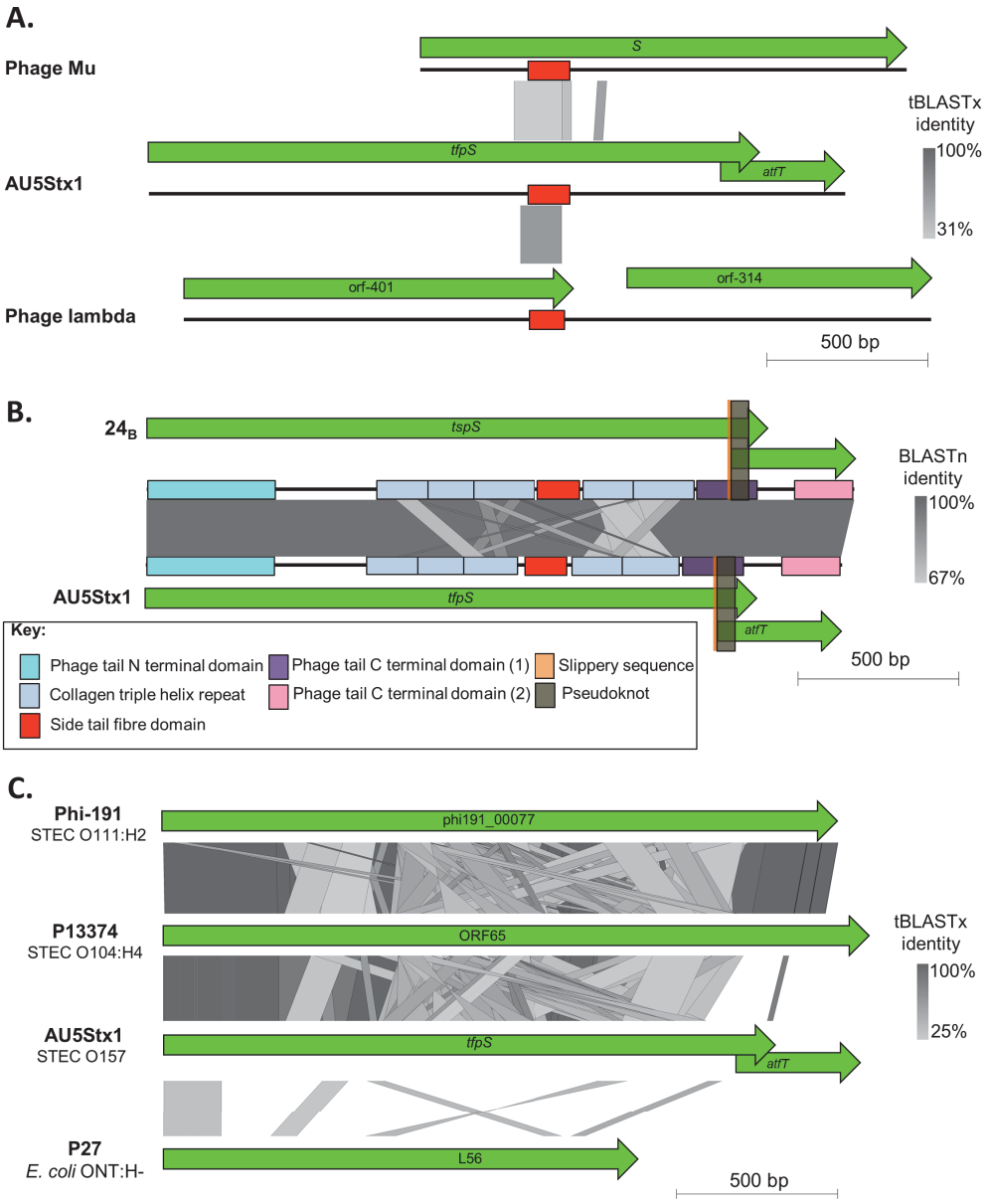


Figure 3.10: Signatures for +1 translational frameshift in tail fibre formation in the majority of known Stx phages. In all diagrams, regions of grey represent BLAST matches scaled according to the respective gradient marker (A) Comparison of tail fibres between AU5Stx1 and characterised phages Mu and lambda. (B) Pairwise comparison of the nucleotide sequence of the tail spike protein gene, and the overlapping downstream ORF between 24_B and the Stx1 phages (represented by AU5Stx1); protein domains along with signatures for +1 translational frame shift are marked as per the key in the figure. In phage 24_B, the overlapping CDS was not designated with either a gene name or a locus tag (C) Three Stx phages in the comparison group possessed a tail fibre with a different C termini to AU5Stx1. The *E. coli* host from which these Stx phages were isolated from are shown in the figure. Image generated using Easyfig (Sullivan *et al.*, 2011).

Downstream of *tfpS* in both AU5Stx1 and AU6Stx1 lies a CDS (AU5Stx1 Locus tag: AU5STX1_710; AU6 locus tag: AU6STX1_720) in the +1 frame, which encodes a protein that possesses another host recognition C-terminal domain (cterm2) with 37% amino acid similarity to cterm1. The nucleotide sequences encoding both cterm1 and cterm2 were also identified in the majority of known Stx phage genomes with the exception of P13374, phi-191 and P27, which carried variant tail fibre proteins with different host recognition domains (Figure 3.7, Figure 3.10.C). The genomes of the previously published Stx phages that possess both cterm1 and cterm2 were interrogated and cterm1 was found to be located within the tail fibre and the cterm2 located within a downstream CDS (annotated as encoding a hypothetical protein in all Stx phages possessing cterm2) in the +1 frame. Further analysis showed that there were amino acid sequence differences between cterm1 and cterm2 in various phages and they were classified into different subtypes with all members within each subtype identical to each other in amino acid sequence (Figure 3.11, Table 3.3). Differences in amino acid sequence among the different cterm1/2 subtypes may indicate they bind different bacterial surface receptors which could lead to different host ranges (Table 3.3) (Weigele *et al.*, 2003).

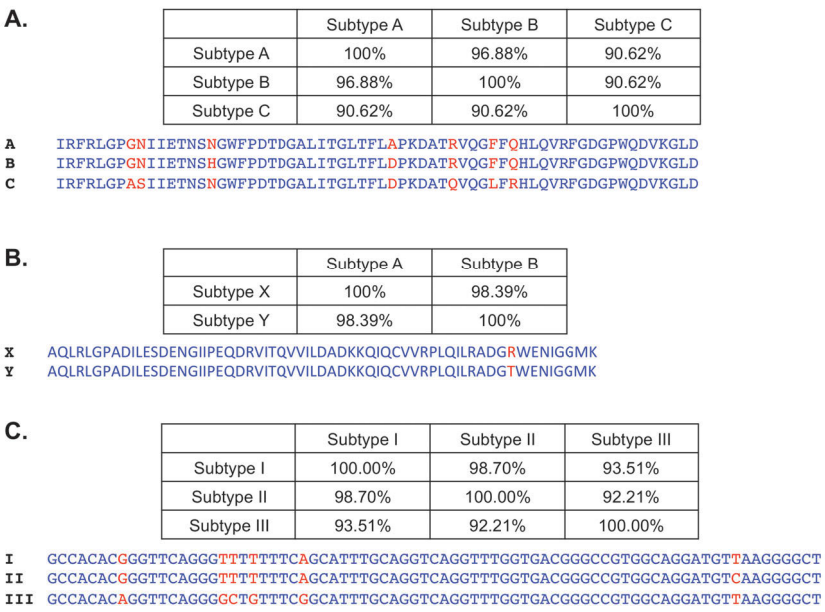


Figure 3.11: Subtypes of (A) cterm1, (B) cterm2 and (C) Pseudoknot sequences Subtypes were classified based on amino acid sequence difference for cterm1 and cterm2 and nucleotide difference for pseudoknot. Percentage identities between the different subtypes are shown in the tables and the sequence alignments displayed below the tables.

Table 3.3: Subtypes of Stx phage tail fibre C terminal domains and pseudoknots.

Phage[a]	Stx genotype	Phage family	cterm1 subtype	cterm2 subtype	Pseudoknot subtype
AU5Stx1	*stx*$_1$	*Podoviridae*	A	X	I
AU6Stx1	*stx*$_1$	*Podoviridae*	A	X	I
BP-4795	*stx*$_1$	ND	B	X	I
stx1CPI	*stx*$_1$	ND	B	Y	II
CP1639	*stx*$_1$	ND	B	X	I
VT1-Sakai	*stx*$_1$	ND	C	Y	III
Hun2013	*stx*$_1$	*Podoviridae*	B	Y	II
933W	*stx*$_{2a}$	*Podoviridae*	B	Y	II
stx2CPI	*stx*$_{2a}$	ND	B	Y	II
stx2CPII	*stx*$_{2a}$	ND	B	Y	II
Min27	*stx*$_{2a}$	ND	B	Y	II
24$_B$	Δ*stx*$_{2a}$::*cat*	*Podoviridae*	B	Y	II
VT2-Sakai	*stx*$_{2a}$	*Podoviridae*	B	Y	II
VT2-SA	*stx*$_{2a}$	ND	B	Y	II
TL2011c	*stx*$_{2a}$	ND	B	Y	II
Phage 2851	*stx*$_{2c}$	*Siphoviridae*	C	X	III
P13374	*stx*$_{2a}$	ND	D	-	-
Phi-191	*stx*$_{2a}$	ND	E	-	-
Phage P27	*stx*$_{2e}$	*Myoviridae*	F	-	-

[a] Further information pertaining the Stx phages can be found in Table 1.1.

As cterm1 and cterm2 are located in separate CDSs in separate frames, genomic signatures for frameshifting at the region of overlap were interrogated. Poly (A) and poly (T) sequences, which are genomic signatures for transcriptional slippage (Wagner *et al.*, 1990), were not detected within the region of overlap. This suggests that the mechanism of frameshift was unlikely to occur by transcriptional slippage. A slippery heptanucleotide sequence (5'- CCCAAAG -3') adhering to the X-XXY-YY-Z motif (Farabaugh, 1996) was identified in a region of overlap within cterm1A and a pseudoknot fold was identified using DotKnot (Figure 3.10B) (Sperschneider *et al.*, 2011) in the nucleotide sequence 80 bp downstream of the slippery sequence. Both slippery sequences and pseudoknots are genomic signatures of translational frameshift where the pseudoknot pauses the ribosome and the slippery sequence causes the slip to another reading frame (Farabaugh, 1996). Identical slippery heptanucleotide sequences were detected in all Stx phages that had both cterm1 and cterm2. Nucleotide sequence variations were observed in the pseudoknots and were classified into three subtypes based on nucleotide similarity to the pseudoknot sequence of both AU5Stx1 and AU6Stx1 (Table 3.3); all members within each subtype are 100% identical to each other in nucleotide sequence (Figure 3.11). Due to a possibility of a frameshift forming a protein with a different C terminus, this gene that harbours cterm2 in AU5Stx1 and AU6Stx1 is annotated as the alternate tail fibre tip (*atfT*)

BLASTn searches of the sequence encoding cterm1A and cterm2X, against the nr/nt database of NCBI, showed similar sequences were also found in the cryptic prophage CRPr13 in *Citrobacter rodentium* strain ICC168 (Petty *et al.*, 2010), and in five prophage sequences in *Escherichia albertii* strain KF1 (Fiedoruk *et al.*, 2014) (Table 3.4). These sequence which spans cterm1 to cterm2, all possess an identical slippery sequence (5'- CCCAAAG -3') and a predicted downstream pseudoknot fold. This sequence analysis indicates that translational frameshift may be possible for alternation of the tail fibres in all of these phages, with the exception of one prophage in *Escherichia albertii* where a premature stop codon would prevent translation of the alternative tail fibre tip EAKF1_ch0442c. The detection of these similar cterm1A and cterm2X in prophages in other bacteria suggests that the Stx phages carrying these related alternative tail fibres, have a broad host range, and may be capable not only of infecting *E. coli*, but also the other attaching and effacing enteric pathogens

Escherichia albertii (Fiedoruk *et al.*, 2014) and *Citrobacter rodentium* (Petty et al., 2010)

Table 3.4: Variant C terminal domains and pseudoknots in the genomes of *Citrobacter rodentium* ICC168 and *Escherichia albertii* KF1

Locus_tag	Frame	Nucleotide identity to			
		cterm1A	cterm2X	Slippery sequence	Type I Pseudoknot
***Citrobacter rodentium* ICC168**[a]					
ROD_12051	0	93%	-	100%	93.51%
ROD_12061	+1	-	97%		
***Escherichia albertii* KF1**[b]					
EAKF1_ch0443c	0	93%	-	100%	93.51%
EAKF1_ch0442c	+1	-	98%		
EAKF1_ch1854c	0	92%	-	100%	93.51%
EAKF1_ch1853c	+1	-	98%		
EAKF1_ch3757	0	92%	-	100%	94.81%
EAKF1_ch3758	+1	-	97%		
EAKF1_ch3955	0	92%	-	100%	94.81%
EAKF1_ch3955	+1	-	92%		
EAKF1_ch1771c	0	91%	-	100%	93.51%
EAKF1_ch1770c	+1	-	95%		

[a]: Accession number: FN543502
[b]: Accession number: CP007025

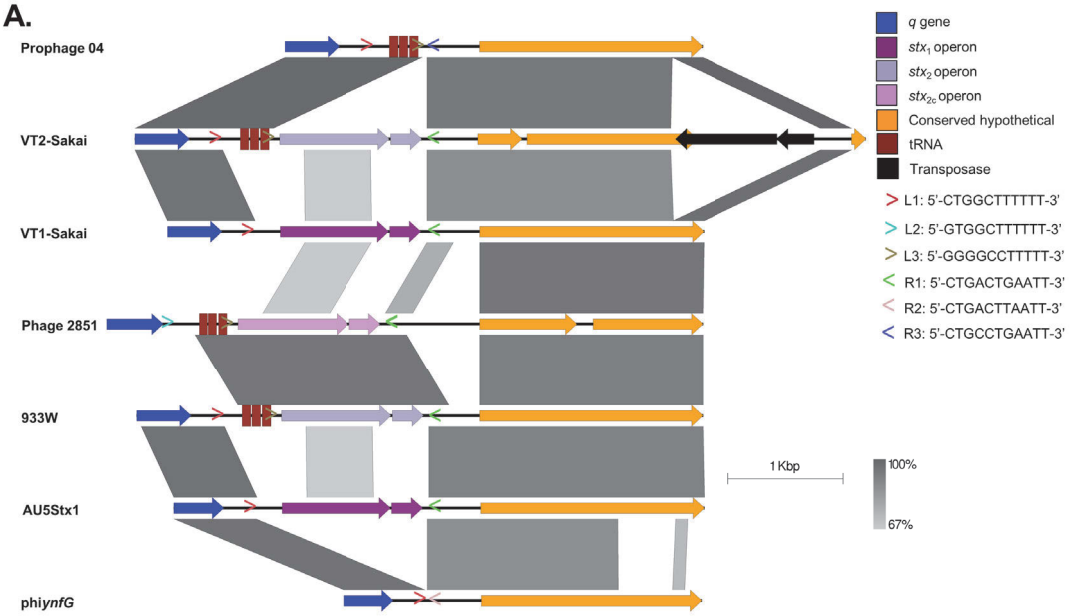
3.2.8. Possible site specific insertion of the *stx* operon into Stx phages

The comparison between published Stx phage genomes revealed a high level of sequence similarity in the genes flanking the *stx1* operon but a clear homology break point was observed surrounding different *stx* operons (Figure 3.7). The same homology breakpoints were also identified in the genomes of highly similar non-Stx prophages, prophage P04 and prophage *yfnG* (Figure 3.7), suggesting that the *stx* operon inserts site specifically into the phage.

Investigation of the sequences surrounding the homology breakpoint flanking the stx_1 operon of AU5Stx1, revealed an 11 bp sequence (5'-CTGGCTTTTTT-3'; L1) upstream of the stx_1 operon, before the homology breakpoint, and an imperfect inverted repeat (5'-CTGACTGAATT-3'; R1) at the homology breakpoint downstream of the stx_1 operon (Figure 3.12). Identical 11 bp imperfect inverted repeats were also detected flanking the *stx* operons in the genomes of all published Stx1 and Stx2a phages. These repeats were also found in the same genomic region of the Stx2c phage 2851 and the non-Stx phages *ynfG* and P04, but one of the repeats within each phage had a single mismatch compared to L1 or R1 ((Figure 3.12). No imperfect inverted repeats could be detected in the Stx2e phage P27, the only Stx phage genome identified that does not have these inverted repeats. When prophage P04 was compared to the Stx2a phages, a homology breakpoint was observed at the stx_{2a} operon, with another 11 bp sequence (5'-CGGGCCTTTTT-3'; L3; 2 mismatch to L1) detected at the upstream breakpoint, downstream of the tRNA genes in P04, as well as in the Stx2a and Stx2c phages (Figure 3.12). In total, six different 11 bp sequences were observed, forming a consensus sequence of 5'-NNGNCNNNTT-3'.

Investigations were subsequently performed to determine if these imperfect inverted repeat flanking the *stx* genes could be part of a Miniature Inverted-repeat Transposition Element (MITE). MITES are non-autonomous mobile genetic elements than are mobilised in-*trans* by a transposase encoded within an IS element (Siguier *et al.*, 2006a). While MITES generally consist of non-coding sequences between terminal inverted repeats, MITES carrying genes that benefit their host bacteria have been previously observed (De Palmenaer *et al.*, 2004; Peters *et al.*, 2004). To identify a potential IS element that could potentially mobilise the *stx* operons in-*trans* as part of MITE, the consensus sequence (5'- NNGNCNNNTT -3') was searched for in the genomes of *E. coli* strains STEC O157 strain Sakai (accession number: BA000007) and the K-12 strain MG1655 (accession number: U00096) as well as *Citrobacter rodentium* strain ICC168 (accession number: FN543502) and *Escherichia albertii* strain KF1 (accession number: CP007025). No IS element candidates were identified in the *Escherichia coli and Escherichia albertii* genomes but a potential IS element candidate, IS*102*, was identified in *Citrobacter rodentium* ICC168. The terminal inverted repeats of IS*102* were previously determined to be an 18 bp sequence (Ohtsubo *et al.*, 1980) and the 11 bp imperfect inverted repeats identified flanking the

stx operons corresponded to the first nine bases of the terminal inverted repeat of IS*102* (Figure 3.12). This suggested that the *stx* operon and flanking inverted repeats may form a MITE that could potentially be mobilised in-*trans* by IS*102*.



B.

```
Terminal inverted repeat of IS102      GGCTTTGTTGAATAAATC
L1                                  CTGGCTTTTTT
L2                                  GTGGCTTTTTT
L3                                  CGGGCCTTTTT
R1                                  CTGACTGAATT
R2                                  CTGACTTAATT
R3                                  CTGCCTGAATT
```

Figure 3.12: Possible site specific insertion of the *stx* gene in the Stx phages. (A) Pair wise comparison of the genetic region encoding the *stx* operon) of representative Stx phages along with prophage 04 and prophage *ynfG*. Imperfect inverted repeats are represented as a coloured arrow representing the orientation of the inverted repeats. Regions of grey between the genomes show BLASTn matches scaled according to the grey bar on the right (B) The 3rd to 11th base of the imperfect inverted repeats identified in the Stx phages showed nucleotide similarity to the first nine bases of the terminal inverted repeat of IS*102*.

3.3. Discussion

In this chapter, two Australian Stx1 phages AU5Stx1 and AU6Stx1 were characterised and the contribution both phage and host to expression of Stx was assessed through lysogeny and toxin production experiments. Electron microscopy and genome sequencing revealed the two Stx1 phages are morphologically and genomically highly related to each other and also to all characterised Stx2a phages.

Genome analysis and comparative genomics revealed new insights into the mechanisms for broad host range and acquisition of the *stx* operons by Stx phages.

3.3.1. Both host and phage genetic background influences Stx1 production.

Results presented in this chapter showed that the bacterial host can influence both Stx1 production and Stx1 prophage induction, as both higher toxin and phage production was observed in the *E. coli* K-12 lysogens when compared to the parent STEC O157 strains (Figure 3.1). While the increase in toxin production in K-12 lysogens harbouring Stx2a phages, compared to their parent STEC was recently demonstrated (Beutin *et al.*, 2012), this is the first demonstration that the same phenomenon is also seen for Stx1 phages. Genetic differences in the STEC O157 isolates from Australia, compared to the rest of the world was first reported in 2001 (Kim *et al.*, 2001) and reinforced by subsequent studies (Leotta *et al.*, 2008; Mellor *et al.*, 2012; Mellor *et al.*, 2013). The demonstration that the same Australian Stx1 phage produces different levels of toxin in different host strains indicates that the genetic differences in Australian STEC strains may contribute to the reduced toxin production and consequently reduced virulence of Australian STEC. In addition, phage borne genes were identified that may also affect both Stx and phage production. Phage AU5Stx1 produced higher levels of phage and toxin compared to AU6Stx1 in both K12 and parent STEC strains (Figure 3.6). An extra anti-repressor was identified in the genome of AU5Stx1 compared to AU6Stx1, and it is plausible that this antirepressor can alleviate repression of the Stx1 prophage leading to higher numbers of phages induced and consequently higher Stx production, . Taken together, this demonstration of different toxin production in both different bacterial hosts harbouring isogenic Stx1 phages, and the same bacterial host harbouring heterogenic Stx1 phages, signifies that not all STEC strains are equally pathogenic and toxin production is influenced by both phage and host genetic factors.

3.3.2. Structure of the Podovirus Stx phage

The most populous members of Stx phages belong to the family *Podoviridae* (Muniesa *et al.*, 2004a; Muniesa *et al.*, 2004b; Garcia-Aljaro *et al.*, 2009) and the tail

appendage responsible for binding these phages onto the bacterial host has been previously named the tail spike protein (Smith *et al.*, 2007a). Results presented in this chapter show that both AU5Stx1 and AU6Stx1 are also podoviruses, however, TEM pictures showed only a single short tail with no evidence of either tail spikes or tail fibres. There is also no evidence of either tail appendage in the literature describing other characterised Stx podoviruses (Plunkett *et al.*, 1999; Allison *et al.*, 2003; Asadulghani *et al.*, 2009; Beutin *et al.*, 2012; Sváb *et al.*, 2015), The name 'tail spikes' is reserved for the short stubby tail appendages that emanate from the central tail tube and are easily visible in electron micrographs of phages that carry them (Casjens & Molineux, 2012) suggesting that the structure for receptor binding in the Stx podoviruses might have been incorrectly named. In addition, the gene predicted to code for the tail fibre has been shown to form a thin, tail fibre-like protein structure that is 94 nm in length when the gene was cloned and expressed in a laboratory *E. coli* strain (Ghosh *et al.*, 2012). In all electron micrographs of podoviral Stx phages, there is no observable tail structure that is 94 nm and the only tail structure observable is the tail tube. This parallels observations made of the podovirus T7 where only the tail tube, but not the tail fibres, are visible in TEM (Steven & Trus, 1986). The results presented in this chapter describing the similarity of morphology of podoviral Stx phages to phage T7, coupled with the discovery that all Stx podoviruses possess a conserved tail fibre gene suggest that, in contrast to previous suggestions, Stx podoviruses uses tail fibres and not tail spikes, for initial reversible binding onto a primary receptor (Figure 3.13).

In addition to the tail fibres, results presented in this chapter also showed the identification of another gene encoding a host recognition protein of the podoviral Stx phages- the tail tip protein. In other characterised phages like phage lambda, the tail tip protein is responsible for the irreversible binding of the phage to a secondary receptor and initiating infection (Roessner & Ihler, 1984). While this gene has been previously identified in other podoviral Stx phages (Plunkett *et al.*, 1999; Beutin *et al.*, 2012; Smith *et al.*, 2012), it has never been annotated as a tail tip protein. The identification of the gene encoding a tail tip protein provided more insights into their biology, suggesting that prior to injecting their DNA into bacteria, the podoviral Stx phages will first undergo reversible binding onto a receptor using the tail fibres,

followed by a second irreversible binding onto a secondary receptor using the tail tip protein prior to committing the phage to inject its DNA into the bacterial cell.

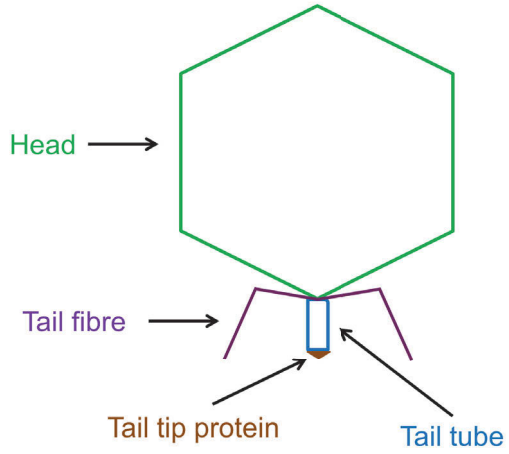


Figure 3.13: Proposed structure of the Podovirus Stx phage. Both phage heads and phage tail tube are visible in TEM (see Figure 3.2). These podoviruses are likely to possess a 94nm tail fibre (coloured purple) emanating from the tail tube but are unable to be visualised in TEM.

3.3.3. First mention of Stx phages integrating into the 5' end of a tRNA

While tRNAs are common insertion sites for phages (Campbell, 2003b) the integration into tRNA genes by Stx phages are rare as they preferentially integrate into chromosomal genes (Makino *et al.*, 1999; Plunkett *et al.*, 1999; Strauch *et al.*, 2008) The *attB* sequence recognised by the integrase of AU5Stx1 and AU6Stx1 was determined to be the first 24 bp at the 5' of tRNA-*argW*. An identical 24 bp attB sequence was also previously determined for 24_B but the authors did not report that the insertion site is tRNA-*argW* (Fogg *et al.*, 2007). In addition, the integrase gene of AU5Stx1 and AU6Stx1 is highly similar to the integrase gene of 24_B. This relationship between the *int* and *attB* suggests that the *attB* of 24_B is tRNA-*argW*

The 5' end of a tRNA is a site rarely used by mobile genetic elements, as integration usually occurs in the 3' end of the tRNA gene (Campbell, 2003b). The reason for this has been suggested to be that the regulatory signals at the 5' end are more crucial, as a transcriptional terminator is easier to provide than an upstream regulatory signal, and

thus displacement of the 5' end after integration creates a pseudogene (Campbell, 1992). In the case of the Australian Stx1 phages, insertion into the 5' of tRNA-*argW* will displace the regulatory signal for the tRNA, however as there is another tRNA in the bacterial chromosome that can replace the function of tRNA-*argW*, this insertion event is tolerated. An integration into the 5' end of a tRNA has only previously been observed once in a cryptic prophage in *Mesorhizobium loti*, a bacterium which forms a nitrogen-fixing symbiosis with leguminous plants, where the insertion site was determined to be the first 39 bp at the 5' end of tRNA-*ser* (Zhao & Williams, 2002). The authors of the study also observed that the prophage provides a regulatory sequence upstream for tRNA-*ser* and suggested that this tRNA is functional. The results presented in this chapter are the first mention of a Stx phage insertion site within the 5' end of a tRNA.

3.3.4. The possibility for alternative tail fibres as an explanation of broad host range

A tail fibre gene (possessing sequences that encode cterm1) and overlapping sequence encoding alternative tips for the tail fibres (possessing sequences that encode cterm2) were identified in the Stx1 phages. A heptanucleotide slippery sequence and pseudoknot was identified. Heptanucleotide slippery sequences and the presence of a downstream pseudoknot are traditionally genomic signatures for a -1 translational frameshift (Farabaugh, 1996). Phages belonging to the *Siphoviridae* family (lambda, HK97 and HK022 infecting *E. coli* ; L5 and TM4 infecting *Mycobacterium*; phiC31 infecting *Streptomyces*) and the *Myoviridae* family (Phage P2 and Mu infecting *E. coli*) have been shown to utilise a -1 translational frameshift to produce proteins required for the assembly of phage tails (Levin *et al.*, 1993; Christie *et al.*, 2002; Xu *et al.*, 2004). However, slippery sequences and a downstream pseudoknot have also been observed to promote a +1 translational frameshift in a temperate phage belonging to the *Siphoviridae* family, phage PSA, which infects *Listeria monocytogenes* (Zimmer *et al.*, 2003). Phage PSA can produce a variant capsid protein and a variant tail protein, with similar N termini but different C termini by +1 translational frameshifting, and both variant products of the capsid and tail genes were found to be required for assembling the virion structure (Zimmer *et al.*, 2003). The proportion of variant capsid proteins produced was 1:6.5 and the proportion of variant

tail proteins was 1:3 (Zimmer *et al.*, 2003). Based on the evidence in Phage PSA, the genomic signatures for translational frameshift observed in Stx phages could allow the formation of one predominant population and subpopulation of phages, expressing tail fibres with different host recognition domains; effectively extending the host range of these Stx phages, accounting for the biological observations that Stx phages have a broad host range. To date, this is the first time that +1 translational frameshift is suggested as a mechanism for phages to infect a broad host range as well as first mention of +1 translational frameshift potentially occurring in *Podoviridae* and Stx phages.

While this in-*silico* analysis revealed the potential for a +1 translational frameshift, experimental evidence is needed to confirm this phenomenon. This could be tested by cloning the tail fibre gene and the downstream gene into an expression vector and expressing it in *E. coli*. The N-terminus of the tail fibre would be tagged and western blot analysis performed using an antibody against the N-terminus tag. Presence of two protein bands of different sizes, corresponding to the different sizes of the predicted alternative tips, would confirm the translational frameshift occurs. Host range studies could then be conducted with a Stx phage with wildtype *tfpS* gene and the downstream gene along with knockout mutations of each the *tfpS* gene and the downstream gene to determine if host range has been affected in the knockouts.

Previously, it has been suggested that short tailed Stx phages only encode one host recognition protein (Plunkett *et al.*, 1999; Smith *et al.*, 2007a). In addition to the tail fibre that can potentially undergo translational frameshift, another gene predicted to encode the central tail fibre was detected in both Australian Stx1 phages as well as other similar Stx phages. The identification of more than one encoded tail fibre, suggests that there could be multiple surface receptors that the Stx phage can interact with. The identification of surface receptors utilised by Stx phages has a history of ambiguity. A study in 1998, using knockout and complementation studies in *E. coli* C600 showed that phages 933W and stx2CPI utilised FadL as a surface receptor and stx2CPII utilised both FadL and LamB (Watarai *et al.*, 1998). In a later study, the surface receptor utilised by a different Stx2a phage, phage 24_B, was determined to be BamA indirectly using antibodies and absorption studies as *bamA* is an essential gene

and knockouts could not be performed (Smith *et al.*, 2007a). The comparative genomics analyses performed on the tail fibres in this chapter, showed that phages 933W, stx2CPI, stx2CPII and 24_B all encode the same alternative tail fibre tip domains cterm1B and cterm2Y as well as pseudoknot subtype II (Table 3.4). While tantalising to suggest that the study by Watarai *et al*., (1998) provided evidence for phages utilising alternative surface proteins, the inability of 933W and stx2CPI to utilise LamB is contradictory, especially when they both harbour the same tail fibres with identical host recognition domains, as well as tail tip proteins with stx2CPII. In addition, a recent study showed that the Stx2a phage VT2-Sakai (cterm1B, cterm2X; Table 3.4) utilised BamA as a surface receptor but not LamB and FadL, through an indirect absorption for BamA and knockout and complementation experiment for FadL and LamB in *E. coli* strain MG1655 (Islam *et al.*, 2012). If Stx phages only utilise a single surface receptor, the study by Islam *et al*., (2012) proved that BamA is the surface receptor. However, if there is a possibility of utilising multiple receptors due to translational frameshift, the study by Islam *et al*., (2012) could also be interpreted as BamA being the dominant receptor recognised by the Stx phages. The way forward to determine the surface receptors of these related podoviral Stx phages would be absorption studies on purified surface proteins as performed for other well-characterised phages (Randall-Hazelbauer & Schwartz, 1973). In addition, it might also be prudent to extend the search of other possible surface receptors by infecting a mutant library, as previously done for other phages (Shin *et al.*, 2012), and selecting mutant strains that can inhibit phage absorption and possibly identify other potential surface receptors.

3.3.5. Site-specific insertion of the *stx* operon

Both genomically and morphologically, the two Australian Stx1 phages in this chapter resemble Stx2a phages. In addition, the stx_1 operon has been previously found in phages with structural genes typical of Stx2a phages. Firstly, genome sequencing of a Stx1 phage, stx1CP1 isolated from *E. coli* O157 strain Morioka V526 (Sato *et al.*, 2003b), showed that the stx_1 operon was carried by a phage with high genomic similarity to Stx2a phages (Sato *et al.*, 2003b). Secondly, Asadulghani *et al*.,(2009) observed that the stx_{2a} operon and the flanking regions from the Stx2a phage were swapped out with the stx_1 operon and its flanking region from the Stx1 phage upon

phage induction from STEC O157 strain Sakai. PCR performed on lysogens made in *E. coli* K-12 from this chimeric phage (stx_1 operon in previously Stx2a phage) induced from Sakai, showed that this chimeric phage inserts into the same insertion site utilised by the Stx2a phage, *wrbA* (Asadulghani *et al.*, 2009). Further PCR analysis using primers designed for Stx1 phages, not Stx2 phages, could only amplify flanking genomic regions of the stx_1 gene, in the chimeric phage, indicating a ~6 kb region, encoding the stx_1 operon and flanking regions had been swapped with the ~6 kb region, encoding the stx_2 operon, transforming the Stx2a phage into a Stx1 phage (Asadulghani *et al.*, 2009). This lead Asadulghani *et al*.,(2009) to postulate that recombination due to sequence homology flanking the *stx* genes is the mechanism of the transfer of the *stx* genes.

While cargo genes (e.g. *stx* operon) have been detected in many phage genomes, the mechanism by which they entered the phage has never been elucidated (Hendrix *et al.*, 2000). Previously, in P2-like phages, different cargo genes were observed inserted at the same genomic region in related phages and sharp breaks in homology were observed at the site of cargo gene insertion (Nilsson *et al.*, 2004). Inverted repeats of 33 bp in size with mismatches ranging from 6 bp to 10 bp was also observed at the boundary near the homology breakpoints in 10 out of the 11 related phages, suggested to the authors that the insertions were site-specific (Nilsson *et al.*, 2004). The location of the *stx* operon, downstream of the anti-terminator *q* gene, is conserved amongst Stx phages (Allison, 2007). When the genomes of Stx phages were compared, genomic regions upstream and downstream of the *stx* operon show high levels of similarity but a distinct homology breakpoint is observed at the *stx* operon. When investigated in this chapter, imperfect inverted repeats of 11 bp were detected flanking the *stx* operon, in the region of nucleotide similarity close to the breakpoints.

MITES are non-autonomous mobile genetic elements, flanked by terminal inverted repeats, that were implicated as IS remnants that lost the transposase gene but can be mobilised in-*trans* by another transposase that recognises the same terminal inverted repeats (Siguier *et al.*, 2006a; Delihas, 2011). MITES carrying cargo genes have been previously characterised. MIC231V in *Bacillus cereus* carries a fosfomycin resistance gene that has been shown to be trans-activated by a IS*231A* transposase in *vivo* (De

Palmenaer *et al.*, 2004) and a MITE in *Pseudomonas putida* has been previously shown to carry an operon that encodes a phenol degrading phenotype (Peters *et al.*, 2004). Combining the site-specific recombination hypothesis of cargo gene integration by Nilsson *et al*.,(2004) with the capability of MITEs carrying passenger genes to be trans-activated by a transposase (De Palmenaer *et al.*, 2004), and the identification of inverted repeats flanking the stx operon with similarity to the inverted repeats of IS*102*, it seems plausible that the insertion of the *stx* operon into Stx phages may have occurred site specifically as part of a MITE, mediated by a transposase in *trans*. This could be tested experimentally by means of a plasmid mating-out assay (De Palmenaer *et al.*, 2004) to confirm the possibility of the transposition of the *stx* operon by IS*102*. Formation of chimeric Stx phages by recombination due to high levels of nucleotide similarity flanking the *stx* operon has been previously experimentally documented (Asadulghani *et al.*, 2009; Sváb *et al.*, 2015). However, the suggestion posited in this chapter is an additional mechanism, by which a site-specific insertion of only the *stx* operon without any other genes, may account for the conservation in location of the *stx* operon as well as a plausible explanation on how the *stx* genes first entered the Stx phage.

3.4. Conclusion

Morphologically, the Stx1 phages characterised in this chapter resemble characterised Stx2a phages, commensurate with the identified high levels of nucleotide similarity to the structural genes of Stx2a phages. The *attB* sequence of both Stx1 phages were defined to be the first 24 bp of tRNA-*argW* and this is the first mention of a Stx phage integrating into the 5' end of a tRNA. Perhaps the most significant finding presented in this chapter is the identification of a possible a site-specific mechanism for insertion of the *stx* operon into the Stx phage as part of a MITE which provided an insight into how cargo genes originally gets incorporated into phage genomes. The differential toxin production observed for both isogenic Stx1 phages in different hosts and heterogenic Stx1 phages in the same host, showed that Stx production is linked to both phage and host genetic factors. In addition, translational frameshift leading to the formation of alternative tail fibres with different host recognition domains was postulated to be a probable mechanism for the known broad host range of Stx phages. In addition to uncovering the possible contributions of the Stx1 phages and their hosts

to the reduced incidence of disease in Australian STEC, new knowledge has been generated about the Stx phages that provide insights to how they could contribute to the emergence of new Stx producing pathogens.

4. Characterisation of Stx prophages in non-clinical Australian isolates of *E. coli* O157

4.1. Introduction

It was previously observed that if Australian STEC O157 strains harbour a Stx2c prophage, it is located adjacent to *sbcB* (Mellor *et al.*, 2012; Mellor *et al.*, 2013). Stx2c phages are not as well characterised as Stx2a phages with only one Stx2c phage studied in detail to date (Strauch *et al.*, 2004; Strauch *et al.*, 2008). A previous attempt to isolate Australian Stx2c phages with the aim of sequencing their genomes, was unsuccessful as the Stx2c phages could not be induced efficiently and isolated from the host (Sim, 2010). It was not possible to investigate the prophage genomes by sequencing the bacterial host using short-read sequencing technology, where genome assembly can be confounded by repetitive sequences. This is particularly problematic for STEC strains as they have multiple, highly related prophages in their genomes. The advent of long-read sequencing technology overcame these shortcomings (Chin *et al.*, 2013), and provided an alternative solution to study the genomes of the Stx prophages within their host STEC O157 strains. The aim of this chapter was to characterise the Stx prophages in the Australian non-clinical STEC O157 strains EC14, EC571 and EC2441 and through comparative genomics, particularly focusing on the Stx2c prophages, determine any genetic factors within the Stx phages that may contribute to the reduced virulence of their hosts.

4.2. Results

4.2.1. Variation in toxin and phage production in the three STEC O157 strains

STEC O157 strains EC14, EC571 and EC2441 are animal and food isolates selected from the *E coli* O157 strain collection at CSIRO (Table 2.2). All three strains possessed *stx* genotypes and insertion sites representative of Australian STEC O157 strains with EC14 and EC571 each carrying stx_1/stx_{2c}, and EC2441 carrying stx_{2c} only

(Table 2.2). The supernatants of Mitomycin C treated overnight cultures of EC14, EC571 and EC2441 were tested for the presence of Stx using a commercial ELISA kit that detects total Stx production regardless of *stx* subtype. Both EC14 and EC2441 produced similar mean (n=3) Stx ELISA $OD_{450/630}$ scores of 2.0±0.5 and 2.2± 0.4 respectively, however, EC571 did not produce toxin within the detectable range of this assay (Fig 4.1).

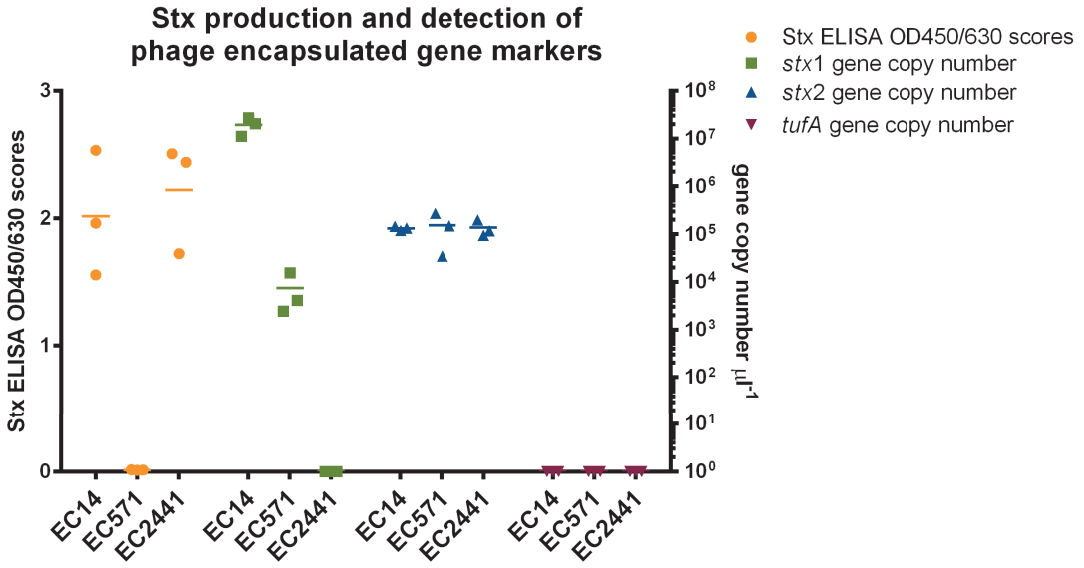


Figure 4.1: Comparison of Stx production and selected gene copy numbers in three Australian STEC O157 strains. Each assay was performed in triplicate, all individual data points are shown and the mean is represented as a horizontal line. Stx ELISA scores are shown on the left axis, with gene copy numbers on the right axis.

To determine the contributions of each Stx prophage to Stx production, phage encapsulated *stx* genes from the same overnight cultures used for the ELISA were quantified by qPCR, as described in section 2.5.2.2. The mean (n =3) phage encapsulated stx_1 was detected at 2.0 x 10^7 gene copies ul^{-1} and 7.3 x 10^3 gene copies ul^{-1} for EC14 and EC571 respectively, with phage encapsulated stx_{2c} detected at approximately the same level of 10^5 gene copies ul^{-1} from all three STEC O157 strains (Figure 4.1). No stx_1 was detected for EC2441, which does not carry a Stx1 prophage, confirming the sensitivity of the assay to differentiate between Stx1 and Stx2. The primers used to detect stx_{2c} are not specific and can also amplify stx_2. However, previous characterisation by colleagues in the CSIRO showed that the only stx_2 subtype carried by EC14, EC571 and EC2441 is stx_{2c} (Table 2.2) and the qPCR

analysis conducted here only amplifies the stx_{2c} gene. The bacterial chromosome control marker *tufA*, was not detected in any of the samples, confirming that the DNA samples were indeed encapsulated within phage virions and were free from bacterial DNA contamination (Figure 4.1). As Stx phages are known to possess only a single set of *stx* genes, it was deduced from the qPCR results that the dominant inducible phage in EC14 was the Stx1 prophage while the Stx2c was the dominant inducible phage in EC571 (Figure 4.1).

4.2.2. Sequencing and assembly of the Stx prophages of EC14, EC571 and EC2441

The three STEC O157 genomes, EC14, EC571 and EC2441 were sequenced using PacBio technology (P4-C2 chemistry; 4 SMRT cells) and assembled as described in Chapter 2. The assembly statistics are listed in Table 4.1.

Table 4.1: Genome assembly statistics for the STEC O157 strains EC14, EC571 and EC2441

	EC14	EC571	EC2441
Total number of contigs	16	12	9
Maximum contig length	4,711,574 bp	4,421,472 bp	4,637,299 bp
Minimum contig length	4,269 bp	4,263 bp	703 bp
Total bases in genome	5,826,350 bp	5,665,564 bp	5,540,805 bp
Mean read coverage[a]	315x	340x	308x

The Stx1 prophages in EC14 and EC571 were each located within the largest single contig in the genome assemblies while the Stx2c prophages of EC14, EC571 and EC2441 were each in two contigs. The Stx1 prophages in EC14 and EC571 were named pp14Stx1 and pp571Stx1 respectively and the Stx2c prophages of EC14, EC571 and EC2441 were named pp14Stx2c, pp571Stx2c and pp2441Stx2c respectively. The prophages were annotated, as described in Chapter 2 (section 2.6.4) and the predicted functional annotations of pp14Stx1, pp571Stx1, pp14Stx2c, pp571Stx2c and pp2441Stx2c are listed in Appendix III, IV, V, VI and VII respectively.

4.2.3. Comparison of the Stx prophages of EC14, EC571 and EC2441

Genome comparison of the Stx1 prophages pp14Stx1 and pp571Stx1 revealed they are highly similar with the notable exception of a 10,215 bp region found in pp14Stx1 but absent from pp571Stx1 (Figure 4.2). This region encodes the genes for lysogeny and phage regulation and its absence, as well as a truncated replication gene (Figure 4.2), suggested that pp571Stx1 does not possess the genes required for either the lytic or lysogenic life cycles, which in turn suggested that it is a defective prophage. At the left boundary of this region of difference, both pp14Stx1 and pp571Stx1 harboured a truncated IS*629* missing the 5' end, an indication of improper excision, a phenomenon previously documented in IS*629* (Kusumoto *et al.*, 2011). IS*629* was previously documented to mediate the deletion of downstream genetic regions up to 20 kb (Ooka *et al.*, 2009) so it is plausible that the deletion of the 10 kb region encoding genes required for the switch between lysis and lysogeny in pp571Stx1 was IS*629*-mediated.

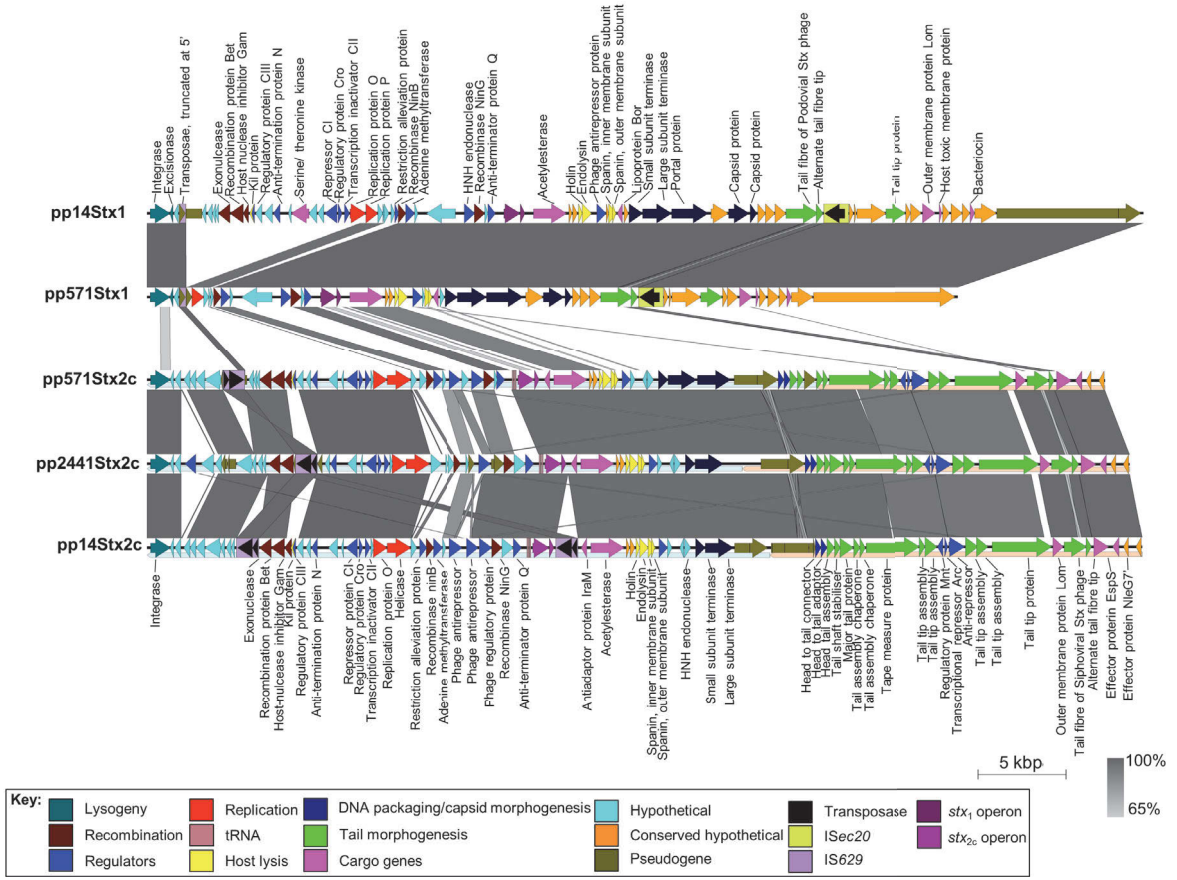


Figure 4.2: Pairwise comparison of the Stx prophages. Contig breaks in the Stx2c prophages are shown in this figure as an alternating light blue and light orange frame below the linear Stx2c prophage genome. Phage functional modules are colour coded as shown in the figure key and regions of nucleotide similarity are shown in grey between the genomes (shaded according to the gradient bar at the bottom right). Image was generated using Easyfig (Sullivan *et al.*, 2011)

Although pp571Stx1 lacked the genes required for the switch from the lysogenic to lytic cycle and DNA replication, the *stx*$_1$ gene was detected in phage encapsulated DNA from EC571 (Figure 4.1). This suggested that it might still be possible for pp571Stx1 to excise from the bacterial genome and be packaged into phage capsids. When phage encapsulated *stx*$_1$ gene copy numbers of both pp571Stx1 and pp14Stx1, under both inducing and non-inducing conditions were compared, it was observed that there was no significant difference in EC571, but a significant difference in EC14 (Table 4.2). This observation further confirmed the inability of pp571Stx1 to switch to the lytic cycle. While the genes required for the switch from lysogeny to lytic cycle and replication are missing and disrupted respectively in pp571Stx1, the genes required for prophage excision (genes encoding the integrase and excisionase) from the bacterial chromosomes were intact (Figure 4.2). In addition, a previous study has shown that a cryptic prophage from *E. coli* K-12 strain BW25113, DLP12, which also possessed a gene encoding the excisionase, but no genes required for the switch from lysogenic to lytic cycle and a truncated replication gene, could be excised from the bacterial chromosome (Wang *et al.*, 2010). Moreover, excised prophages lacking the ability to undergo phage DNA replication have been documented to be packaged into capsids of other phages (Asadulghani *et al.*, 2009) and it is likely that pp571Stx1 is packaged into a phage head from a different phage. These observations paralleled to that of pp571Stx1 and based on these similarities, it provided a likely explanation for the detection of phage encapsulated *stx*$_1$ in EC571 (Figure 4.1).

Table 4.2: Mean *stx*$_1$ gene copy number in inducing and non-inducing conditions.

		mean *stx*$_1$ gene copy number per µl	
	n	EC571	EC14
Spontaneous induction	3	1.94×10^3 (SD:±1.96×10^3)	7.65×10^4 (SD:±8.10×10^4)
Induced with MMC	3	7.38×10^3 (SD:± 7.14×10^3)	1.95×10^7 (SD:±7.92×10^6)

All three Stx2c prophages were highly similar to each other especially within the region encoding structural genes (Figure 4.2). The length of the tape measure gene was 3243 bp, 3081 bp and 3243 bp for pp571Stx2c, pp2441Stx2c and pp14Stx2c respectively. The tape measure gene is responsible for the tail length with the length

of the gene being proportionate to the phage tail length (Katsura & Hendrix, 1984) which suggests that pp2441Stx2c, has a shorter tail than pp571Stx2c and pp14Stx2c. The gene encoding the tail tip protein of pp571Stx2c and pp14Stx2c are identical to each other but showed sequence dissimilarity at the region which encodes the amino acid sequence of the C terminus to pp2441Stx2c (Figure 4.2). The tail tip protein is responsible for the irreversible attachment of the phage onto its surface receptor prior to injection of viral DNA into the cell (Roessner & Ihler, 1984). In phage lambda, point mutations within the C terminus of the central tail fibre altered its host range (Werts *et al.*, 1994), indicating that pp2441Stx2c utilises a different surface receptor from pp571Stx2c and pp14Stx2c.

Intact copies of IS*629* were present in the Stx2c prophages with two copies present in pp14Stx2c and one copy each in pp571Stx2c and pp2441Stx2c (Figure 4.2). In pp14Stx2c, both copies of IS*629* were located within intergenic regions but the IS*629* in both pp571Stx2c and pp2441Stx2c were located within different genes thereby disrupting them. In pp571Stx2c, IS*629* had inserted into a non-conserved gene encoding a hypothetical protein. This insertion did not cause any deletion to the gene, so in an event of precise excision, this gene may be reconstituted. In pp2441Stx2c the IS*629* insertion deleted the 5' end of the *gam* gene, which encodes a host nuclease inhibitor and is one of the genes of the lambda Red recombination system that mediates homologous recombination (Hillyar, 2012). In addition, Gam is also required to inhibit host nucleases that could digest the linear dsDNA produced during phage DNA replication (Enquist & Skalka, 1973). However, in the genome of EC2441, another prophage carried a *gam* homolog and it seems plausible that the Gam, if expressed from this other prophage, could compensate for the disrupted *gam* in pp2441Stx2c.

Both pp14Stx2c and pp2441Stx2c were found to encode a *q* gene that was 100% identical to the nucleotide sequence of the *q* gene of Stx2c phage 2851 (q_{2851}) while the Stx2c prophage pp571Stx2c carried a *q* gene with no nucleotide similarity to q_{2851} but identical to the *q* gene of Stx2a phage 933W (q_{933}). The expression of the anti-terminator protein Q, in Stx phages, allows downstream genes, which include the genes in the lysis and morphogenesis modules to be transcribed as the prophage switches to its lytic life cycle and produce Stx (Neely & Friedman, 1998b; Neely &

Friedman, 1998a). Variation in the *q* gene upstream of the stx_{2c} operon was previously observed in other Australian STEC O157 strains (Mellor *et al.*, 2012; Mellor *et al.*, 2013) but the contribution of either q_{2851} or q_{933} to toxin expression has never been investigated. However, the results presented in Figure 4.1 showing similar quantities of stx_{2c} gene copy number suggested that this variation in the anti-terminator *q* gene has no significant effect on the induction of Stx2c phages.

When the Stx1 prophages and the Stx2c prophages were compared, it was observed that there is no significant nucleotide similarities in the phage morphogenesis region with the only similarity localised to the nucleotide sequences encoding the C-terminus of the tail fibres and the alternative tail tip protein (Figure 4.2). This suggests that both the Stx1 prophage and Stx2c prophages have different phage morphologies. In addition, the tail tip protein is also genetically distinct between the Stx1 and Stx2c prophages which suggests that a different secondary receptor is utilised for both Stx1 and Stx2c prophages

4.2.4. Comparison of the Stx1 prophages of EC14 and EC571 to known Stx phages

The genomes of pp14Stx1 and pp571Stx1 were highly similar (BLASTn) to the genomes of the Australian Stx1 phages AU5Stx1 and AU6Stx1 (Chapter 3) as well as published Stx2 phages and two Stx1 phages, Stx1CPI and HUN2013 (Figure 4.3). Out of all the phages in the comparison, the genomes of both pp14Stx1 and pp571Stx1 were most similar to the genome of AU6Stx1 (Figure 4.3). The Stx1 phage AU6Stx1 is a podovirus and based on the high nucleotide similarity, both pp14Stx1 and pp571Stx1 were inferred to be podoviruses as well.

In both pp14Stx1 and pp571Stx1, an IS element identified as IS*Ec*20 using ISfinder (Siguier *et al.*, 2006b),was integrated downstream of the gene encoding the side tail fibre, in a non-coding region between two phage genes, with no gene disruption. The IS*Ec20* carried by pp14Stx1 and pp571Stx1 are 100% identical to each other in nucleotide sequence, yet none of the other sequenced Stx phages harboured IS*Ec20* (Figure 4.3). The draft genomes of both EC14 and EC571 each only harboured a single copy of this IS was located in the Stx1 prophage which suggested that it did not

come from the host genome. In addition the G+C content of the entire length of IS*Ec*20 was 57%, slightly higher than the G+C content of both *E. coli* (Blattner *et al.*, 1997; Hayashi *et al.*, 2001; Perna *et al.*, 2001) which suggested that this IS element might have been a recent acquisition. To investigate which genomes possess this IS element, a BLASTn search, using the entire IS*Ec20* sequence of pp14Stx1 as a representative against the NCBI nr/nt database was performed. BLASTn results showed that IS*Ec20* was not found in any other genomes of STEC O157 strains, but was found in other strains of *E. coli*, *Shigella boydii*, *Shigella flexneri*, *Shigella dysenteriae* and *Shigella sonnei,* with nucleotide identity ranging from 99%- 100% over the entire IS element length (Table 4.3). This BLASTn result suggest that it is likely that these Stx1 phages may have a host range that extends beyond STEC O157 and most likely acquired this IS element from a host that harbours it.

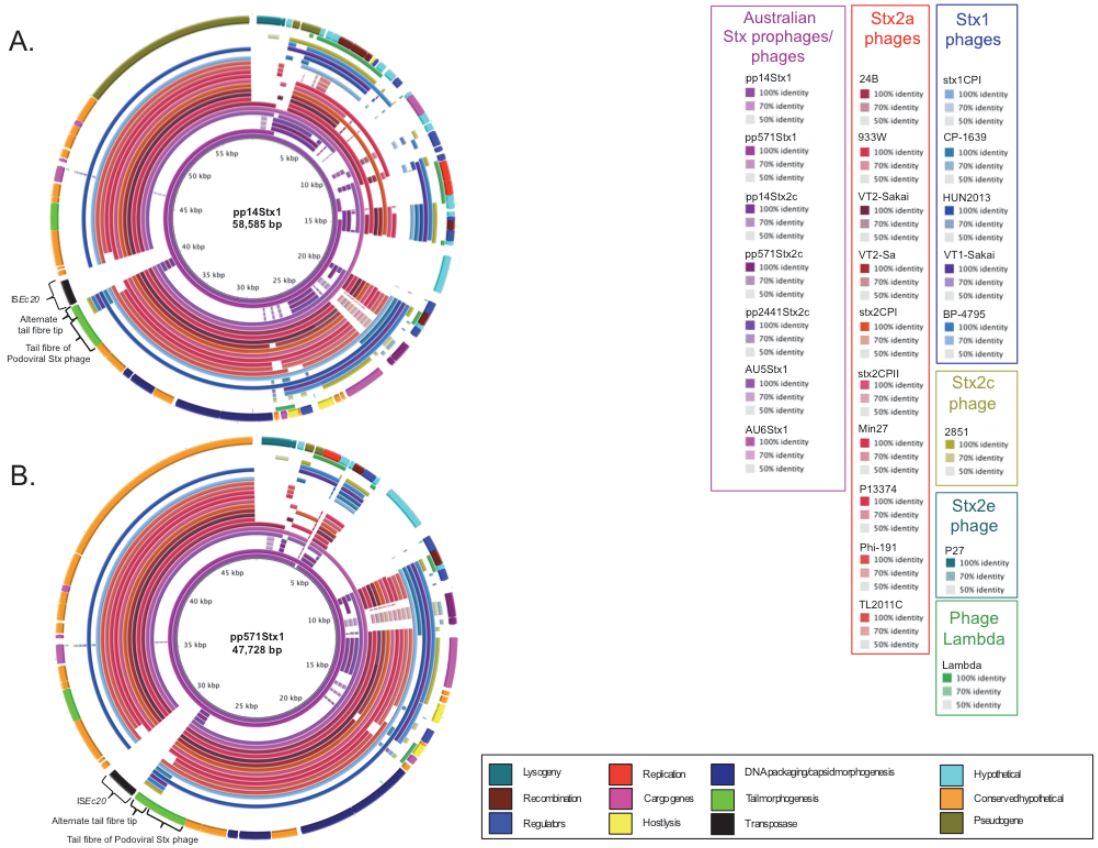


Figure 4.3: Genome comparison of the Stx1 prophages to other Stx phages. The Stx1 prophages from (A) EC14 and (B) EC571 are presented as the reference genome in the centre of each figure in a circular representation of its genetic map. The outer most ring shows the phage annotation, colour coded as per the figure key. Each of the coloured rings represents a different query phage genomes, and the coloured regions represent BLASTn matches to the reference. Query genomes and their percentage identity are colour coded as listed in the figure key. Image generated using BRIG (Alikhan *et al.*, 2011)

Table 4.3: *Enterobactericeae* genomes that harbour IS*Ec20*.

Genomes	Number of copies	Nucleotide similarity to IS*Ec20* of pp14Stx1	Accession number
Escherichia coli			
O145:H28 strain RM12581	6	99%- 100%	CP007136
O145:H28 strain RM13516	6	99%- 100%	CP006262
O145:H28 strain RM13514	6	99%- 100%	CP006027
O145:H28 strain RM12761	6	99%- 100%	CP007133
O111:H- strain 11128	1	100%	AP010960
O103:H2 strain 12009	1	99%	AP010958
O55:H7 strain RM12579	1	99%	CP003109
O55:H7 strain CB9615	1	99%	CP001846
1303	1	99%	CP009166.
RS218	1	99%	CP007149
ATCC 25922	1	99%	CP009072
Nissle 1917	1	99%	CP007799
UM146	1	99%	CP002167
ABU 83972	1	99%	CP001671
IHE3034	1	99%	CP001969
536	2	99%	CP000247
UTI89	1	99%	CP000243
CFT073	1	99%	AE014075
MNCRE44	1	99%	CP010876
ST131 strain EC938	1	99%	HG941718
JJ1886	1	99%	CP006784
PMV-1	1	99%	HG428755
NA114	1	99%	CP002797
‘clone D i14’	1	99%	CP002212
‘clone D i2’	1	99%	CP002211
SE15	1	99%	AP009378
Plasmid pCFSAN029787_01	2	99%	CP011417
Plasmid p53638_226	1	99%	CP001064
Plasmid p948	1	99%	FN649418
Plasmid pSE11	1	99%	AP009243
Shigella boydii			
CDC 3083-94	13	99%	CP001063
Sb277	8	99%	CP000036
ATCC 9210	7	99%	CP011511
Plasmid pBS512_211	1	99%	CP001062
Plasmid pSB4_227	1	99%	CP000037
Shigella dysenteriae			
1617	2	98%	CP006736
Sd197	4	99%	CP000034
Plasmid pSD1_197	1	99%	CP006737
Shigella sonnei			
53G	8	99%	HE616528
Ss046	9	99%	CP000038
CFSAN030807	13	99%	CP011422
Plasmid A	1	99%	HE616529

Genomes	Number of copies	Nucleotide similarity to IS*Ec20* of pp14Stx1	Accession number
Plasmid pSS_046	2	99%	CP000039
Plasmid pCFSAN030807	2	99%	CP011423
Shigella flexneri			
2002017	4	99%	CP001383
2003036	4	99%	CP004056
2457T	5	99%	AE014073
301	6	99%	AE005674
8401	6	99%	CP000266
NCTC1	6	99%	LM651928
Shi06HN006	4	99%	CP004056
G1663	5	99%	CP007037
Plasmid pSFxv_1	3	99%	CP001384
Plasmid pCP301	3	99%	AF386526
Plasmid pWR501	2	99%	AF348706
Plasmid pSF5	1	99%	AY879342
Plasmid pG1663	3	99%	CP007038

Interrogation of the genome sequences of EC14 and EC571 also revealed that the Stx1 prophage in each genome was integrated adjacent to tRNA-*argW*. The integrase genes (*int*) carried by both pp14Stx1 and pp571Stx1 were identical to each other and the *int* genes carried by AU5Stx1 and AU6Stx1 (Chapter 3). In addition, pp14Stx1 and pp571Stx1 were also each flanked by two identical 24 bp repeats (5'-GTCCTCTTAGTTAAATGGATATAA-3') which are identical in each Stx1 prophage and also to AU5Stx1 and AU6Stx1 (Chapter 3), which corresponded to the first 24 bp of tRNA-*argW*. The detection of identical integrase and *att* sequence to AU5Stx1 and AU6Stx1, suggested that the *attB* of both pp14Stx1 and pp571Stx1 was the 24 bp 5' of tRNA-*argW*.

4.2.5. Comparison of the Stx2c prophages of EC14, EC571 and EC2441 to known Stx phage genomes

The genomes of the Stx2c prophages pp14Stx2c, pp571Stx2c and pp2441Stx2c were found to be highly similar (BLASTn) to the genomes of Stx2c phage 2851 (Strauch *et al.*, 2008) and Stx1 phage BP-4795 (Creuzburg *et al.*, 2005b) especially in the phage morphogenesis genes (Figure 4.4). The morphology of BP-4795 is not known, but phage 2851 has been shown to be a siphovirus (Strauch *et al.*, 2004). Based on the high level of nucleotide similarity in the morphogenesis genes, pp14Stx2c,

pp571Stx2c and pp2441Stx2c are inferred to also belong to the *Siphoviridae* family. In addition, a tBLAStx comparison of the tail morphogenesis genes of the Stx2c prophages with the siphoviruses lambda (Sanger *et al.*, 1982) and HK97 (Juhala *et al.*, 2000) showed significant amino acid similarity, furthering the notion that the three Stx2c prophages are siphoviruses (Figure 4.5).

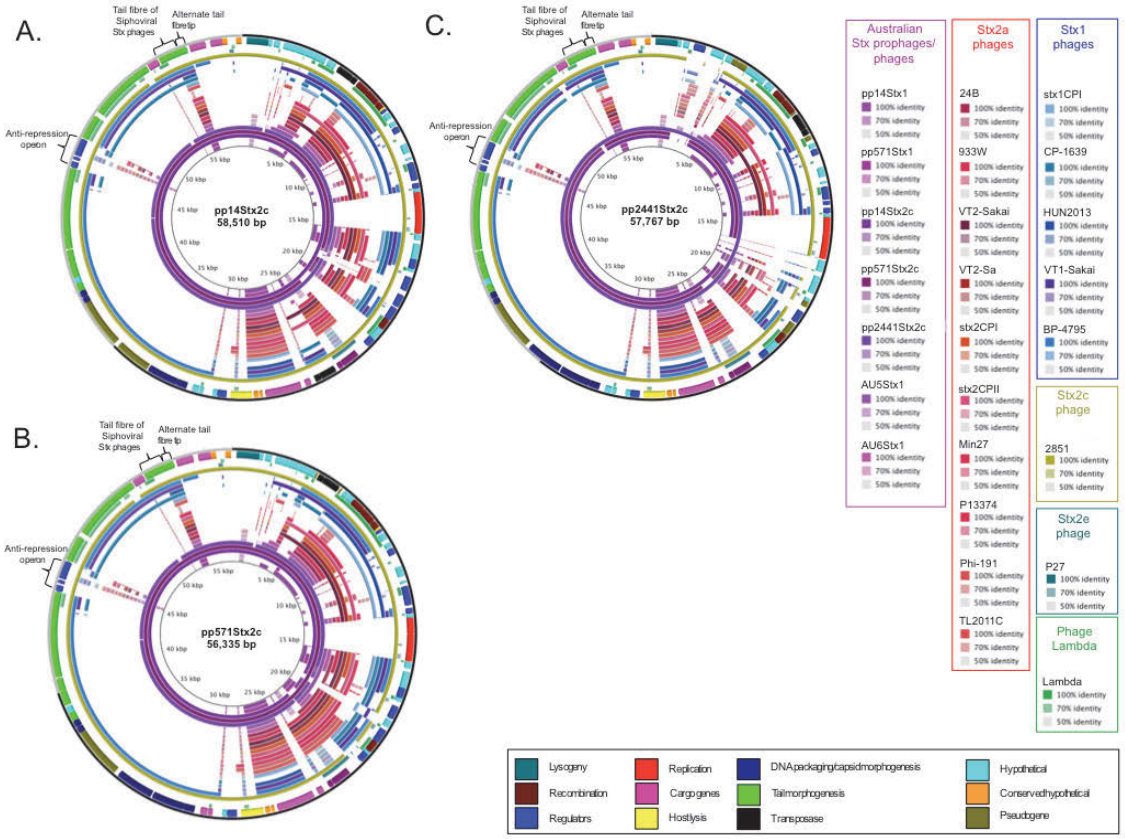


Figure 4.4: Genomic comparison of the Stx2c prophages to other Stx phages. The Stx2c prophages from (A) EC14, (B) EC571 and (C) EC2441 are presented as a reference genome in the centre with a circular representation of its genetic map. The outer most ring are the phage annotation, colour coded as the figure key. Each of the Stx2c prophages are in 2 contigs reflected here by the alternating black-grey arc. Each of the coloured rings represents a different query phage genome, and the coloured regions represent BLASTn matches to the reference. Query genomes and their nucleotide identity are colour coded as listed in the figure key. Image was generated using BRIG (Alikhan *et al.*, 2011).

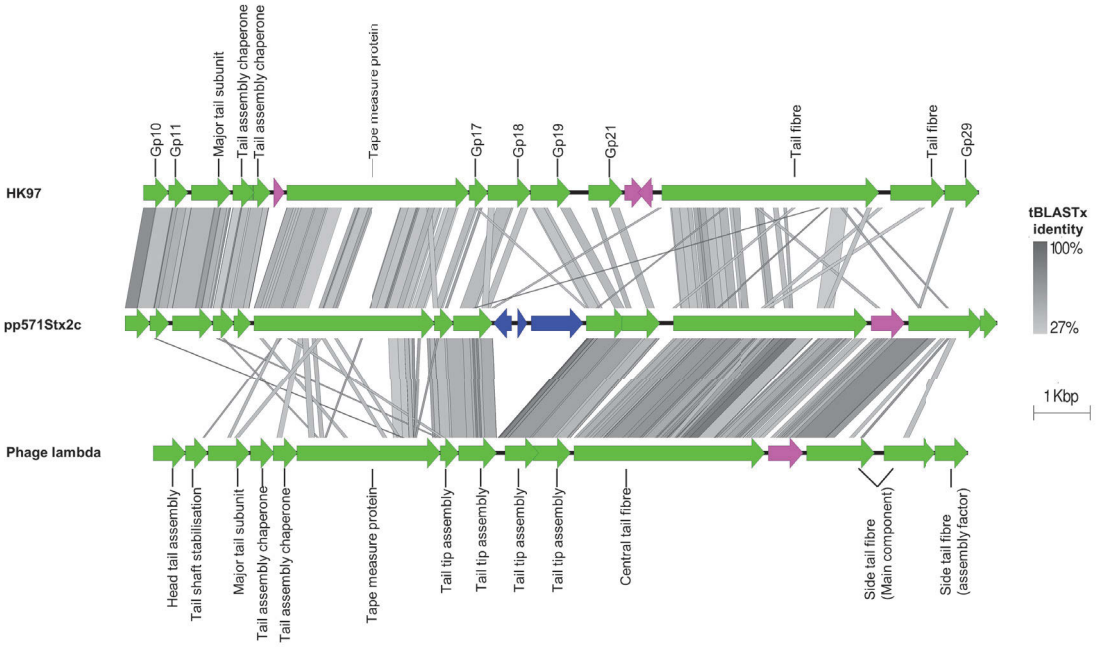


Figure 4.5: Comparison of the tail morphogenesis genes of pp571Stx2c, HK97 and lambda. Pairwise comparison of the tail morphogenesis genes of the three Australian Stx2c prophages (pp571Stx2c as a representative) with those of known siphoviruses HK97 and lambda. Regions of grey between sequences represent tBLASTx matches (shaded according to percentage amino acid identity as indicated in the key). . Image generated with Easyfig (Sullivan *et al.*, 2011)

4.2.6. Translational frameshifting predicted within tail genes

In Chapter 3, two different host recognition C-terminal domains were detected encoded in the majority of known Stx phage genomes with one domain encoded within the tail fibre gene (cterm1) and the other within an overlapping downstream CDS in the +1 frame (cterm2) with a heptanucleotide slippery sequence and pseudoknot, in the overlapping sequence, hypothesised to produce subpopulations of viral progeny with different host recognition domains by translational frameshift, allowing an extended host range (Chapter 3).

All five Stx prophages characterised in this chapter were also found to encode the alternative tail fibre tips with two host recognition C-terminal domains and conserved heptanucleotide slippery sequence and pseudoknot in the overlapping sequence between the two CDSs. The prophages pp14Stx1 and pp571Stx1 possessed identical cterm1, cterm 2 and pseudoknot subtypes to the Stx1 phages AU5Stx1 and AU6Stx1 (Table 4.4; Table 3.3) while pp2441Stx2c possessed an identical C-terminal domains

and pseudoknot subtype to the Stx2c phage 2851 (Table 4.4; Table 3.3). Both pp14Stx2c and pp571Stx2c also possessed an identical cterm2 and pseudoknot subtype to phage 2851, but possessed a cterm1 subtype (cterm1G) that was not observed previously in Stx phages (Table 4.4; Table 3.3).

Table 4.4: Variant C terminal domains and pseudoknots encoded in the tail fibre genes of the five Stx prophages.

Stx prophages	cterm1 subtype	cterm2 subtype	Pseudoknot subtype
pp14Stx1	A	X	I
pp571Stx1	A	X	I
pp2441Stx2c	C	X	III
pp571Stx2c	G	X	III
pp14Stx2c	G	X	III

The cterm1G subtype was almost identical to cterm1C with only a single amino acid substitution, at the last amino acid (Figure 4.6). This substitution was identified to be caused by a SNP, that also resulted in a stop codon (TGA) in the +1 frame where the translational frameshift occurs which would force a premature translational termination after the frameshift (Figure 4.7).

```
Subtype A    IRFRLGPGNIIETNSNGWFPDTDGALITGLTFLAPKDATRVQGFFQHLQVRFGDGPWQDVKGLD
Subtype B    IRFRLGPGNIIETNSHGWFPDTDGALITGLTFLDPKDATRVQGFFQHLQVRFGDGPWQDVKGLD
Subtype C    IRFRLGPASIIETNSNGWFPDTDGALITGLTFLDPKDATQVQGLFRHLQVRFGDGPWQDVKGLD
Subtype G    IRFRLGPASIIETNSNGWFPDTDGALITGLTFLDPKDATQVQGLFRHLQVRFGDGPWQDVKGLN
```

Figure 4.6: Multiple sequence alignment of cterm1 subtypes. Amino acid differences are highlighted in red

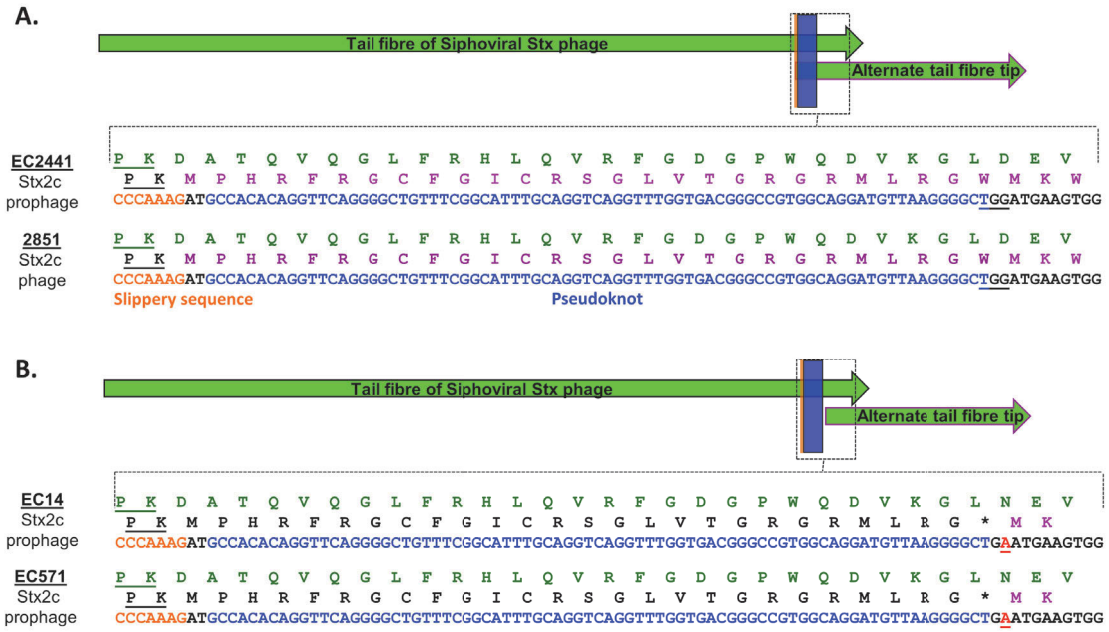


Figure 4.7: Sequences encoding tail fibres with alternative tips in Stx2c prophages. The DNA sequence downstream of the pseudoknot has a SNP difference (highlighted in red) between (A) Stx2c phages encoding cterm1C and (B) Stx2c phages encoding cterm1G, resulting in a single amino acid difference in cterm1 and a premature stop codon in the +1 frame (demarcated by an asterisk) disrupting the CDS encoding the alternative tail fibre tip (coloured purple).

However, the codon TGA not only encodes a stop codon but also the 21st amino acid selenocysteine (Bock *et al.*, 1991). To successfully incorporate selenocysteine and out compete translational termination, four *sel* genes from unlinked operons (*selA*, *selB*, *selC* and *selD*) are required and the disruption of any of these *sel* genes when tested on *E. coli* K-12 will lead to translational termination (Leinfelder *et al.*, 1988). These four *sel* genes were identified in the genomes of EC14 and EC571 and found to be intact and have identical sequences in both strains. The four *sel* genes carried by both EC14 and EC571 were highly similar to *selA*, *selB*, *selC* and *selC* carried by *E. coli* K-12 strain MG1655 with 99%, 98% 100% and 97% nucleotide similarity respectively. This suggests that there could be a possibility that this stop codon could be read through and +1 translational frameshift could still potentially form infectious phage particles with alternative tail fibres, leading to a broad host range.

It has been previously documented that -1 translational frameshifting within overlapping tail assembly chaperone genes located between the major tail protein and the tape measure protein is a conserved mechanism to form proteins required for tail

assembly in long tailed phages (Xu *et al.*, 2004). To determine if this mechanism was conserved in Stx2c phages, the sequences in the region of overlap between the tail assembly chaperone genes were visually inspected in the Stx2c prophages pp14Stx2c, pp571Stx2c, pp2441Stx2c and the Stx2c phage 2851. Each phage possessed an identical heptanucleotide slippery sequence, 5'-AAAAAAG-3' and an identical downstream sequence (5'-CGGAAATTCGCTTTCTGATGCGA-3') that was predicted to fold into a pseudoknot using DotKnot (Sperschneider *et al.*, 2011) (Figure 4.8), both of which are signatures for a -1 translation frameshift (Farabaugh, 1996). This observation confirmed the translational frameshift mechanism for tail assembly in long tailed phages is conserved in these four related Stx2c phages.

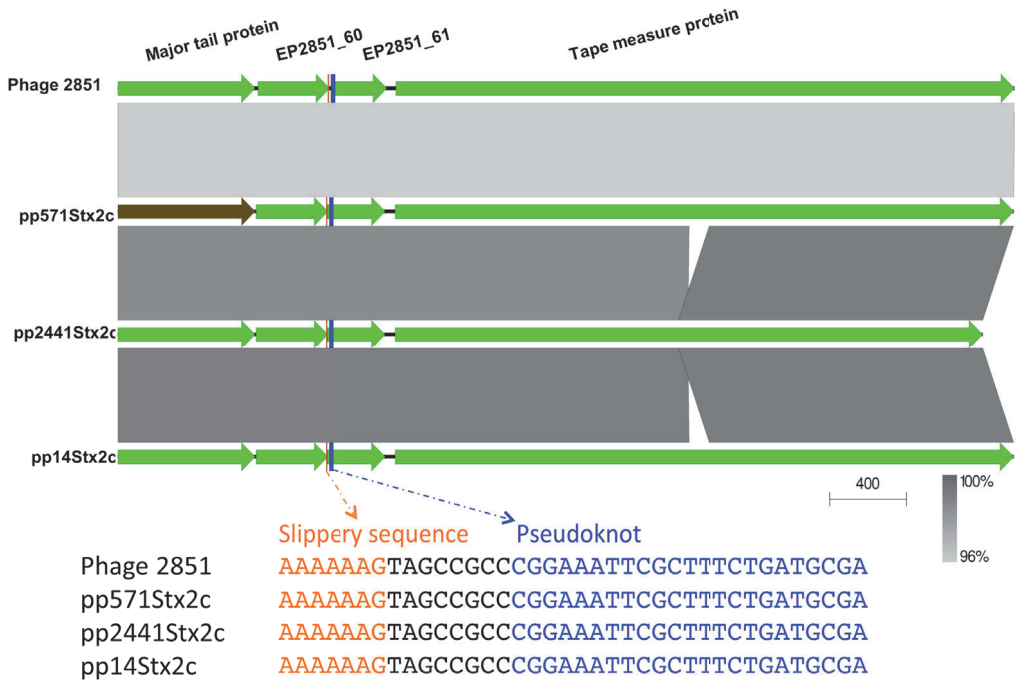


Figure 4.8: The genetic region encoding Stx2c phage tail assembly. Four tail morphogenesis genes from the three Stx2c prophages and the Stx2c phage 2851 are shown in a pairwise comparison. Regions of grey between the different phage sequences represent BLASTn matches, shaded according to percentage identity as shown in the key. The genetic signatures for -1 translational frameshift, slippery sequence and pseudoknot, are highlighted in orange and blue, respectively. Image was generated using Easyfig (Sullivan *et al.*, 2011).

4.2.7. The insertion site of the Stx2c prophages

The *int* genes (1179 bp) carried by pp14Stx2c, pp571Stx2c and pp2441Stx2c shared 99% nucleotide identity amongst each other. Most of the single nucleotide polymorphisms (SNP) were synonymous, where the change in nucleotide sequence did not result in a change to the amino acid sequence. However, a SNP (1073A>C) in pp14Stx2c was non-synonymous, where the nucleotide sequence change resulted in a change of the amino acid sequence. This non-synonymous SNP however was a conserved substitution whereby the altered amino acid possessed similar properties and was unlikely to affect the function of the integrase (Figure 4.9). The integrases of pp14Stx2c, pp571Stx2c and pp2441Stx2c were all predicted to be tyrosine integrases based on conserved Pfam domains.

```
pp571Stx2c          VAISDTKLRTIYGKPYSGPQEVADADGLSVRISPKGVIQFQYRYRWHGKPNRLGLGRYPS
pp2441Stx2c         VAISDTKLRTIYGKPYSGPQEVADADGLSVRISPKGVIQFQYRYRWHGKPNRLGLGRYPS
pp14Stx2c           VAISDTKLRTIYGKPYSGPQEVADADGLSVRISPKGVIQFQYRYRWHGKPNRLGLGRYPS
                    ************************************************************

pp571Stx2c          LSLKDARQITADLRKLYFSGTDPRTYFEEKVENSMTVAQCLDYWFDNYVSTTLREKTQAL
pp2441Stx2c         LSLKDARQITADLRKLYFSGTDPRTYFEEKVENSMTVAQCLDYWFDNYVSTTLREKTQAL
pp14Stx2c           LSLKDARQITADLRKLYFSGTDPRTYFEEKVENSMTVAQCLDYWFDNYVSTTLREKTQAL
                    ************************************************************

pp571Stx2c          YRSTVMKRMHDAFPNRPASSITVKQWVDLLTEEEKDNPRRARQVLSQLRSAISWCMRRQL
pp2441Stx2c         YRSTVMKRMHDAFPNRPASSITVKQWVDLLTEEEKDNPRRARQVLSQLRSAISWCMRRQL
pp14Stx2c           YRSTVMKRMHDAFPNRPASSITVKQWVDLLTEEEKDNPRRARQVLSQLRSAISWCMRRQL
                    ************************************************************

pp571Stx2c          IDSCAIMSIQPRDFGSRAEVGDRVLSYHELAKIWLAIERSRASTSNKLLHQMLMLWGARL
pp2441Stx2c         IDSCAIMSIQPRDFGSRAEVGDRVLSYHELAKIWLAIERSRASTSNKLLHQMLMLWGARL
pp14Stx2c           IDSCAIMSIQPRDFGSRAEVGDRVLSYHELAKIWLAIERSRASTSNKLLHQMLMLWGARL
                    ************************************************************

pp571Stx2c          SELRLATKTEFDLLDNVWTVPKEHSKMGNVIRRPIFEQIKPFLEKAMTTYNDVLFPGEDI
pp2441Stx2c         SELRLATKTEFDLLDNVWTVPKEHSKMGNVIRRPIFEQIKPFLEKAMTTYNDVLFPGEDI
pp14Stx2c           SELRLATKTEFDLLDNVWTVPKEHSKMGNVIRRPIFEQIKPFLEKALTTYNDVLFPGEDI
                    *********************************************:**************

pp571Stx2c          NKPISIAAANRFVNRIRGGMDLGYWRTHDFRRTLVTRLSEMNVEPHVTERMLGHELGGIM
pp2441Stx2c         NKPISIAAANRFVNRIRGGMDLGYWRTHDFRRTLVTRLSEMNVEPHVTERMLGHELGGIM
pp14Stx2c           NKPISIAAANRFVNRIRGGMDLGYWRTHDFRRTLVTRLSEMNVEPHVTERMLGHELGGIM
                    ************************************************************

pp571Stx2c          SVYNKHDWIEAQRKAYELHADKLFWHIRSISD
pp2441Stx2c         SVYNKHDWIEAQRKAYELHADKLFWHIRSISD
pp14Stx2c           SVYNKHDWIEAQRKAYELHADKLFWHIRSISD
                    ********************************
```

Figure 4.9: Multiple amino acid sequence alignment of the integrases of the Stx2c prophages. The conserved substitution is highlighted in red. Amino acid sequence alignment was performed using Clustal Omega (Sievers *et al.*, 2011).

BLASTn searches of the *int* carried by pp571Stx2c (as a representative) against the NCBI nr/nt database showed highly similar integrases of identical size were found in

other *E. coli* strains and phages (nucleotide identity over entire length as shown in parenthesis): *E. coli* O157:H7 strains TW14359 (99%) and EC4115 (99%), *E. coli* O111:H- strain 11128 (99%), *E. coli* O145:H28 strain RM12581 (99%) and RM13514 (99%), *Enterobacteria* phage Sf101 (99%) and the Stx2c phage 2851 (99%). In all these bacterial genomes, the *int* gene was associated with lambdoid prophages located adjacent to *sbcB* and in *E. coli* O157:H7 strains TW14359 and EC4115 these lambdoid phages were the Stx2c prophages, indicating a common insertion site for the Australia Stx2c phages.

In the genomes of EC14, EC571 and EC2441, the Stx2c prophages were also found to be integrated adjacent to *sbcB* and flanked by two identical 13 bp repeats (5'-TTTCACGATTACG-3') (Figure 4.10). This 13 bp sequence was identical to the *attB* of both phage 2851 (Strauch *et al.*, 2008) and *Enterobacteria* phage Sf101 (Jakhetia *et al.*, 2014) and were determined to be the *attL* and *attR* of pp14Stx2c, pp571Stx2c and pp2441Stx2c. This 13 bp sequence was determined to be located within *sbcB,* with the insertion of the Stx2c prophage displacing the first 36 bases of *sbcB* and the upstream promoter from the rest of the gene (Figure 4.10). Despite the displacement, an alternative start site was detected 33 bp upstream of *attR* within the Stx2c prophages, indicating that the insertion of the Stx2c prophage may not cause a loss-of-function mutation in *sbcB*, but may provide an alternative 5' end to *sbcB,* as previously suggested during the characterisation of the Stx2c phage 2851 (Strauch *et al.*, 2008).

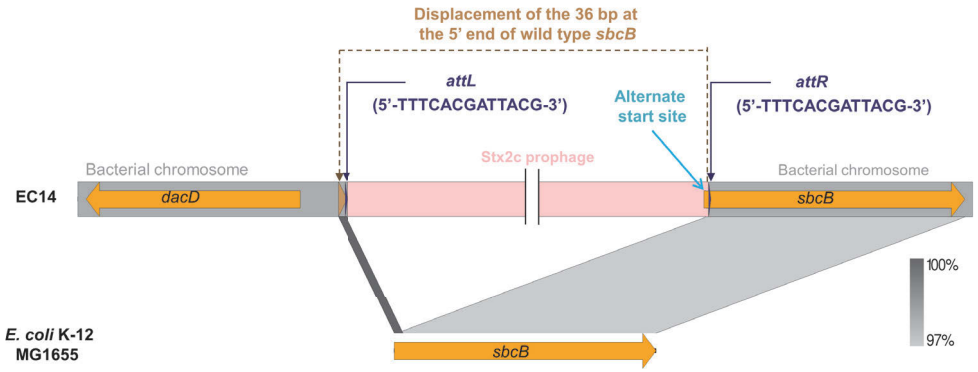


Figure 4.10: Schematic representation of the insertion site of Stx2c prophages. Regions of grey between genome sequences represent BLASTn matches The insertion of the prophage causes displacement of the 5' end of *sbcB*, but a prophage-encoded alternative start site reconstitutes an intact *sbcB* gene with a variant 5' end. Image was generated using Easyfig (Sullivan *et al.*, 2011).

4.2.8. DNA replication systems and predicted DNA packaging of the Stx2c prophages

The three Stx2c prophages, pp14Stx2c, pp571Stx2c and pp2441Stx2c all have genes that encode the replication protein O and a helicase (Figure 4.2). The gene that encodes the replication protein O was identical in all three prophages but the gene encoding the helicase in pp2441Stx2c had a single synonymous SNP when compared to pp14Stx2c and pp571Stx2c. Both replication protein O and helicase of the Stx2c prophages shared 51% (over 93% of query) and 90% (over entire length) amino acid sequence identity to Gp18 and helicase (Gp12) of phage P22 respectively. P22 does not require a host helicase for replication (Backhaus & Petri, 1984; Wickner, 1984a; Wickner, 1984b) and the level of similarity with the replication machinery of pp14Stx2c, pp571Stx2c and pp2441Stx2c indicates they may also not require a host-encoded factor for DNA replication.

Located upstream of the genes encoding the phage terminase, in all three Stx2c prophages, is a 366 bp gene encoding a predicted HNH endonuclease (Figure 4.2). The HNH endonuclease gene encoded by pp571Stx2c and pp14Stx2c are 100% identical to each other while the endonuclease gene of pp2441Stx2c shared 89.34% nucleotide identity to the other two Stx2c prophages. When compared to the endonuclease gene encoded by the Stx2c phage 2851 (Locus tag: EP2851_50), pp571Stx2c and pp14Stx2c shared 95.08% nucleotide identity while pp2441Stx2c shared 86.07% nucleotide identity. An endonuclease gene was also previously detected upstream of the terminase genes in the Stx2e phage P27 (Quiles-Puchalt *et al.*, 2014). When compared to the endonuclease gene of P27 (locus tag: L30), the endonuclease gene harboured by pp571Stx2c, pp14Stx2c and pp2441Stx2c shared no nucleotide similarity (Figure 4.4) but the amino acid sequences of the three Stx2c prophages shared a similar HNH endonuclease domain with P27. Through knockout and complementation analysis, it was previously shown that both the HNH endonuclease and the terminase genes, were required for *cos*-site cleavage in the Stx2e phage P27 (Quiles-Puchalt *et al.*, 2014). In addition, another study showed that a gene encoding a HNH endonuclease was also located upstream of *cos*-packaging phages: *Bacillus* phage γ, *Bacillus* phage 105, *Clostridium* phage phi3626, *Lactobacillus* phage Lrm1 and phage HK97 (Xu & Gupta, 2013). Taken together,

these data suggest that the three Stx2c prophages and the Stx2c Phage 2851 are also *cos*-packaging phages.

4.2.9. The three Stx2c prophages carry a cargo gene which provides adaptation to magnesium starvation

All three Stx2c prophages encoded an identical 324 bp gene (*iraM*) located downstream of the stx_{2c} operon (Figure 4.2). The nucleotide sequence of this gene was also found to be present in the genome of the Stx2c phage 2861 (100% nucleotide identity) but it was not previously identified or annotated. The *iraM* carried by the Stx2c phages has 72% nucleotide identity to the chromosomally encoded *iraM* of *E. coli* K-12 strain MG1655.The gene *iraM* encodes an anti-adaptor protein that was found to be induced in *E. coli* K-12 strain MG1655 during magnesium starvation and stabilise the central stress response regulator RpoS, which allowed the bacterium to adapt to its magnesium-starved environment (Bougdour *et al.*, 2008). BLASTn searches of the *iraM* carried by pp14Stx2c (as a representative) against NCBI's nr/nt database showed genes of identical size with highly similarity (nucleotide identity and genomic location as shown in parenthesis) in *E. coli* O111:H- strain 11128 (Stx2a prophage: 98%), *E. coli* O157:H7 strains TW14359 (Stx2c prophage: 100%) and EC4115 (Stx2c prophage: 100%), *E. coli* O145:H28 strains RM12761 (Stx2 prophage: 98%; Chromosome: 71%) and RM13516 (Stx2a prophage: 98%,; Chromosome: 71%). However, in all these genomes, with the one exception of strain 11128, this CDS was not previously marked. As this gene does not perform a function pertaining to the phage life cycle, but instead provides a benefit to the host, it is a cargo gene.

4.2.10. The Stx2c prophages encode an anti-repression operon

All three Stx2c prophages carried three genes, *mnt, arc* and *ant*, which were predicted to encode a putative regulatory protein, transcriptional repressor, and an anti-repressor respectively (Figure 4.2). Between the three Stx2c prophages, the genes *mnt* and *arc* are 100% identical in nucleotide sequence to each other but for *ant* reduced nucleotide similarity ranging from 96.65% to 99.89% was observed. The gene organisation of

mnt, *arc* and *ant* carried by the three Stx2c prophages were highly similar to the *immI* region (henceforth referred to as the anti-repression operon) in Phage P22 (Sauer *et al.*, 1983). Based on the gene organisation and similarity to the anti-repression operon of Phage P22, the anti-repression operon detected in the Stx2c phages was suggested to have the same function as in Phage P22 (Figure 4.11). The anti-repression operon was also detected in the genome of the similar Stx2c phage 2851 where *arc* (Locus tag: EP2851_65) and *ant* (Locus tag: EP2851_66; incorrect predicted start site) was annotated but the *mnt* gene was not annotated, presumably due to *mnt* being located on the opposite strand. Similar anti-repression operons (*mnt*- *ant*) were also detected at the same location in the Stx2c prophages (nucleotide identity over entire operon), of STEC O157:H7 strain TW14359 (99%) and EC4115 (99%) following a BLASTn search against the nr/nt database in NCBI.

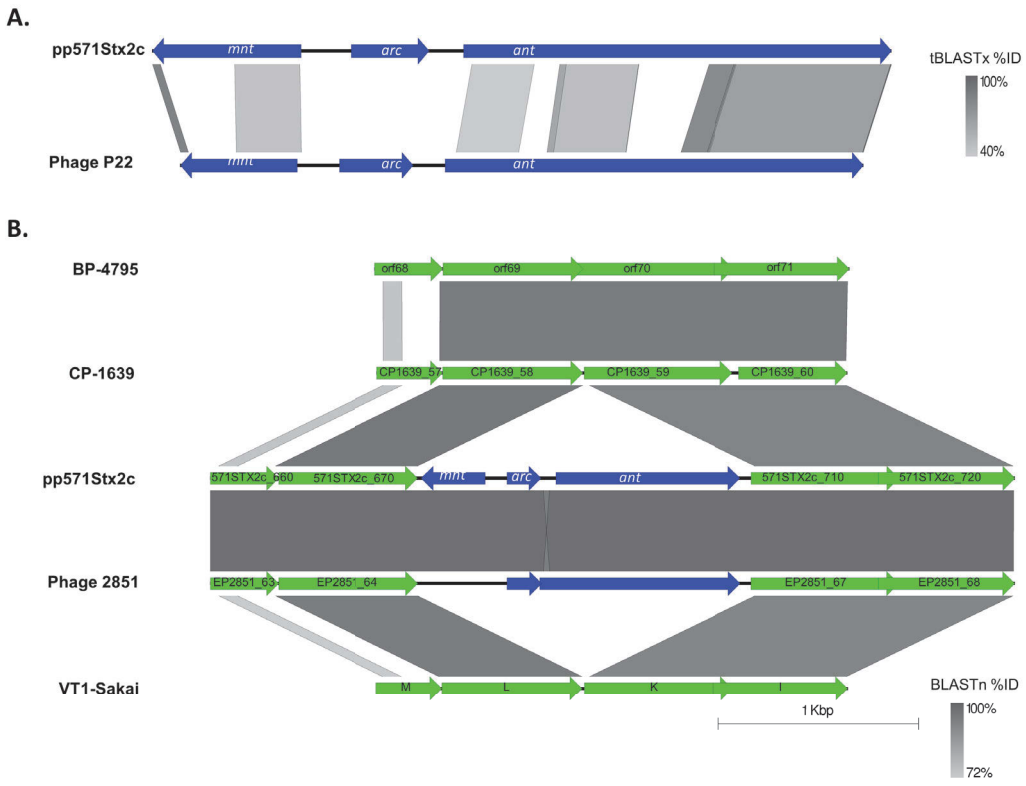


Figure 4.11: Stx2c prophages are the only Stx phages which possess an anti-repression operon. Regions of represent BLAST matches between sequences (A) tBLASTx comparison between the anti-repressor operon of pp571Stx2c (as a representative of the three Stx2c prophages) and P22. (B) Genomic comparison of a genomic region (BLASTn) between the Stx2c prophage (pp571Stx2c as a representative) and other similar type Stx phages.

In both P22 and the Stx2c prophages, the anti-repression operon is located within the tail morphogenesis modules. In addition, when compared to closely related sequences of the Stx1 phages BP-4795, VT1 Sakai, Stx1 prophage CP-1639 and the Stx2c phage 2851, it was observed that this anti-repression operon was only present in the Australian Stx2c prophages, phage 2851 and the Stx2c prophages of STEC O157:H7 strain TW14359 and EC4115 which suggested that this anti-repression operon was horizontally acquired (Figure 4.11).

In Phage P22, the anti-repressor encoded within the anti-repression operon inactivates the phage repressor and induces lytic growth of the phage (Sauer *et al.*, 1983). The Mnt protein prevents the expression of Ant, allowing the lysogen to be maintained and also to prevent expression of Ant from other super infecting phages (Sauer *et al.*, 1983). Arc prevents the high level expression of anti-repressor during late stages of the lytic cycle (Sauer *et al.*, 1983). In addition, the anti-repressor produced by P22 was previously shown to be capable of alleviating repression of other closely related prophages within the same host genome, by acting on their repressors in-*trans* (Susskind & Botstein, 1975). The demonstration of wide specificity of the anti-repressor in the anti-repression operon of phage P22 to closely related phages (Susskind & Botstein, 1975) and its similarity to the anti-repression operon in Stx2c phages may indicate that Stx2c phage-encoded anti-repressors could also act in-*trans* and potentially alleviate repression of other prophages in the host genome.

An experiment to test for alleviation of repression in-*trans* was performed using the STEC O157 strain EC14 and EC3206 (Table 2.2). To investigate this hypothesis further, data on spontaneous Stx1 phage induction from strains EC14and EC3206 were compared. EC14 harbours an intact Stx1 prophage and a Stx2c prophage (with an anti-repression operon) while EC3206 harbours only a Stx1 prophage (Table 2.2), which is identical to the Stx1 prophage in EC14 (with the one exception of IS*Ec20* insertion in EC14 and the last gene in Stx1 prophage marked as a pseudogene due to a frameshift disruption) (Figure 4.12). To investigate if a Stx2c prophage, which possesses an anti-repression operon, could alleviate repression of the other Stx prophage in-*trans*, phage encapsulated stx_1 from the spontaneous induction. The mean (n=3) phage encapsulated stx_1 gene copy number of EC3206 was 9.4×10^3 gene copies ul^{-1} , which is 2 logs lower than for EC14 at 1.5×10^5 gene copies ul^{-1}.

While there is a possibility of the involvement of other host genetic factors, this result does provide tantalising evidence to support the suggestion that the presence of the anti-repression operon in the Stx2c prophage may be able to counteract repression of the Stx1 prophage in-*trans* and thus be responsible for higher induction of the Stx1 prophage in EC14 as compared to EC3206.

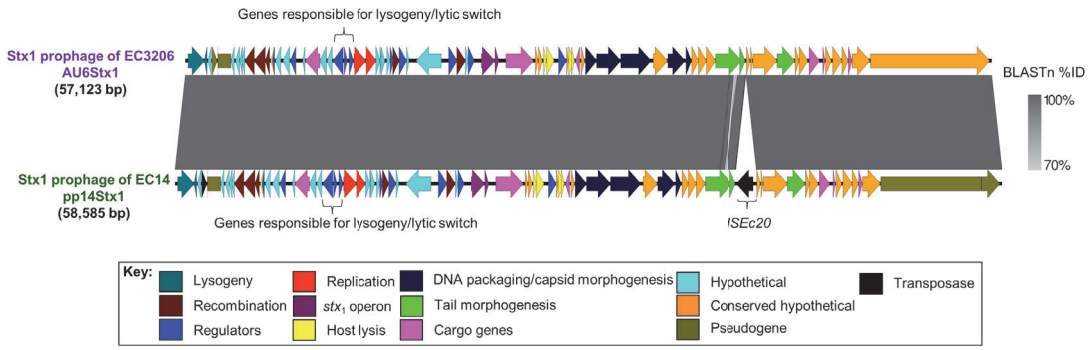


Figure 4.12: Comparison between the Stx1 prophages of EC3206 and EC14. Pair wise comparison between the Stx1 prophage of EC3206 and the Stx1 prophage of EC14. Regions of grey represent BLASTn matches

4.3. Discussion

In this chapter, the Stx prophages of three non-clinical STEC O157 strains were characterised and the contributions of the Stx on host expression of Stx was assessed through toxin production and phage quantification experiments. Whole genome sequencing of these three STEC O157 strains revealed that the Stx1 prophages characterised in this chapter were genetically similar to the Stx1 phages characterised in chapter 3 while the Stx2c prophages were genetically similar to the characterised Stx2c phage 2851. Genome analysis and phage quantification experiments also revealed new insights into the prophage-prophage interactions within a host which can affect Stx production in STEC O157 strains.

4.3.1. A dominant inducible Stx prophage within a bacterial strain harbouring two Stx prophages

It has been previously observed that in strains of STEC O157 (TW14359 and Sakai) harbouring two Stx prophages, one Stx prophage is the dominant phage being induced

at a higher rate over the other prophage (Asadulghani *et al.*, 2009; Kulasekara *et al.*, 2009). In both cases, it was the *podoviridae* Stx2a prophage that was observed to be dominant with the Stx2a prophage in Sakai being induced at a rate approximately 10 times higher than the Stx1 prophage (Asadulghani *et al.*, 2009) and the transcription of stx_2 encoded in the Stx2a prophage in STEC O157 strain TW14359 was approximately 300 times higher than that of stx_{2c} encoded in the Stx2c prophage in the same genome (Kulasekara *et al.*, 2009). Out of the three strains studied in this chapter, only two STEC O157 strains, EC571 and EC14, harboured two Stx prophages, however, the Stx1 prophage was found to be defective in EC571. In STEC O157 strain EC14, the Stx1 prophage was the main inducible prophage with phage encapsulated stx_1 detected at 147 times higher than phage encapsulated stx_{2c}. Despite all three studies, inclusive of this chapter, showing a similar observation of a dominant Stx prophage, different methods as mentioned above were utilised so a direct comparison cannot be made. However, the results presented in this chapter is the first time that Stx1 phages are shown to be dominant Stx phage

All Stx2c prophages, studied in this chapter carried an anti-repression operon located within the genomic region encoding tail morphogenesis genes (Figure 4.2). Using two STEC O157 strains, with one strain harbouring an Stx1 prophage and a Stx2c prophage with the anti-repression operon, and the other strain harbouring an almost identical Stx1 prophage, it was shown that the Stx1 prophage was induced at a higher rate in the strain also carrying the Stx2c prophage than the strain without. This indicated that the anti-repression operon in the Stx2c prophage may alleviate repression of the Stx1 prophage in-*trans*. However, there might be other genetic differences that might affect this result. To unambiguously prove that counter repression occurs, a panel of K-12 lysogens (Stx1 prophage only, Stx2c prophage with anti-repressor operon only and a dual Stx1 and Stx2c lysogen) should be formed and their induced phage titres and toxin production compared.

The same phage anti-repressor operon was also detected in the genome of the Stx2c prophage of STEC O157 strain TW14359, which was previously observed to transcribe stx_{2a} at a higher rate than stx_{2c} as described above. The authors of this study suggested that the Stx2a phage exerted regulatory control on the Stx2c phage (Kulasekara *et al.*, 2009). However, in light of the results presented in this chapter, a

plausible alternative explanation for the dominance of the Stx2a prophage could be that it is due to the Stx2c prophage alleviating the repression of the Stx2a prophage in-*trans*, forcing it into its lytic cycle and thus resulting in the higher *stx*$_2$ transcription. Moreover, strains that harbour both Stx2a and Stx2c prophages tend to fall within the hypervirulent clade of STEC O157 (Manning *et al.*, 2008). Based on the data generated in this Chapter, it is hypothesised that trans-alleviation of Stx2a prophage repression by the anti-repression operon of the Stx2c prophages, leading to the increase in Stx2a production is the reason for why STEC O157 strains that harbour both Stx2a and Stx2c prophages are more virulent than any other strains carrying either a single Stx prophage or harbouring both Stx1 and Stx2a.

4.3.2. Disruptions by IS elements contribute to variations in toxin production

Not all STEC O157 are equal in their ability to produce toxin and the amount of toxin produced was previously suggested to be strain specific (Mellor *et al.*, 2015). In this chapter, the three Australian STEC O157 non-clinical isolates showed different levels of toxin production with EC14 and EC2441 producing similar amounts of toxin while EC571 had lower toxin production, undetected by the limits of the toxin detection assay. In EC571, the Stx1 prophage was rendered defective in both phage replication and the ability to switch to lytic life cycle by a deletion, predicted to be mediated by IS*629*. In *E. coli* O157 strain Sakai, IS*629* was determined to be the most abundant IS element (Hayashi *et al.*, 2001) and was previously shown to induce a range of genomic rearrangements which included insertion, excision and deletions of flanking genomic regions (Ooka *et al.*, 2009). IS*629* was previously observed to disrupt the ability to produce Stx by insertion into the gene encoding the B subunit of Stx (Kusumoto *et al.*, 1999; Okitsu *et al.*, 2001) but this is the first mention of an IS*629* insertion into a genomic region encompassing phage regulatory genes, which led to the disruption of toxin production by rendering the Stx1 prophage defective. Moreover, IS*629* insertions were found localised to the accessory genome of STEC O157 strain Sakai (Ooka *et al.*, 2009), of which the Stx prophage are part of the accessory genome. It is therefore plausible that under the right circumstance, such as disruptions of phage regulatory genes by other mobile genetic element could contribute to variation of toxin production in STEC O157 strains.

4.3.3. DNA packaging and tail formation in the Stx2c prophages

From this in-*silico* analysis, it was predicted that these Stx2c phages are *cos* packaging phages due to similarity in gene organisation to other *cos* packaging phages, with the carriage of an endonuclease upstream of the terminase genes (Xu & Gupta, 2013; Quiles-Puchalt *et al.*, 2014).

Two different translational frameshifts were predicted to occur within the tail genes of the Stx2c prophage. Signatures for a -1 translational frameshift were identified within the tail assembly chaperone genes, a conserved mechanism previously demonstrated to be required for tail assembly in long-tailed phages (Xu *et al.*, 2004). The second translational frameshift predicted was the same +1 translational frameshift within the gene encoding the side tail fibres previously identified in the Stx1 phages in Chapter 3, and predicted to allow the production of a subpopulation of phages with alternate side tail fibres and extending the host range. In this chapter, an additional Cterm1 subtype (cterm1G) was identified in pp14Stx2c and pp571Stx2c, indicating that both pp14Stx2c and pp571Stx2c may have a different host range to other Stx phages. A premature stop codon was also identified in the alternative tail fibre tip gene of pp14Stx2c and pp571Stx2c, but read through of the termination signal was predicted to occur based on the presence of intact *selC* genes in its bacterial host chromosome. In addition, the difference in the tail tip protein observed between the podoviral Stx phages and the siphoviral Stx phages suggests that both these morphotypes of phages utilise different secondary receptors for irreversible binding. Stx2c is a clinically relevant Stx subtype and informed experiments on host range of these phages should be performed to determine the extent of their host range and the potential risk of emergence of other Stx2c producing pathogens.

4.4. Conclusion

In this Chapter, it was demonstrated that related Australian non-clinical STEC O157 strains belonging to the same O157 subtype and having the same complement of stx genes, produced different levels of toxin. Deletion of a 10kb region within a Stx1

prophage rendering it defective was identified as a contributing factor. Apart from this deletion, the Stx1 prophages were found to be highly similar to the genomes of the Stx1 phages in Chapter 3. The Stx2c prophages analysed in this Chapter are genetically similar to Stx2c phage 2851, the only Stx2c phage to be sequenced thus far. In addition, the in-*silico* analysis gave insights into the biology of these Stx2c prophages that showed that these Stx2c were siphoviruses and suggested that DNA packaging occurred by *cos*-packaging. In addition, translational frameshift in two different tail genes where the first region of translational frameshift was predicted to be required for tail assembly while the second region of translational frameshift is predicted to allow the phage to extend its host range, as also identified in the Stx1 and Stx2a phages (Chapter 3). Perhaps the most significant finding presented in this chapter was the enticing evidence to suggest that Stx2c prophages counteract repression of the lytic cycle of other Stx prophages in the same host genome, causing increased phage induction and toxin expression. In the current literature, it has been observed that STEC O157 strains harbouring both Stx2a and Stx2c prophages are more virulent (Kulasekara *et al.*, 2009) and this finding of counteracting repression in-*trans* is a potential explanation as to why STEC O157 strains harbouring both a Stx2a and Stx2c prophage are more highly virulent than strains with any other combinations of Stx prophages and are able to produce more toxin. While no insights into reduced virulence were gained in this Chapter, these results revealed new insights into how Stx phages can contribute to increased virulence of their host bacterium.

5. Mobilisation of the Locus of Enterocyte Effacement by an Stx2a prophage

5.1. Introduction

The Locus of Enterocyte Effacement (LEE) pathogenicity island has been detected in multiple genomic locations in attaching and effacing bacteria, which suggested that it could be mobilised (Deng *et al.*, 2001; Castillo *et al.*, 2005; Petty *et al.*, 2010). During a previous multiplex PCR screen of DNaseI treated phage lysates, in addition to the expected *stx*$_2$, the LEE-encoded *eae* was also detected within phage particles induced from a New Zealand clinical STEC O157 strain EC1810, and the DNA packaged in these phage virions (the phage metagenome) was subsequently sequenced (Petty *et al*., unpublished). The sequencing reads revealed that apart from induction of the Stx2a prophage of EC1810 (henceforth referred to as 1810Stx2a) which was integrated adjacent to tRNA-*argW*, there were also sequencing reads corresponding to the LEE, located over 1Mbp away in the genome (Petty *et al*., unpublished). This data suggested that either the Stx2a prophage or a cryptic prophage upstream of the LEE could mediate the mobilisation of the LEE (Petty *et al*., unpublished). The primary aim of the work presented in this chapter was to characterise the Stx2a phage, investigate the mechanism for prophage mediated LEE mobilisation and determine if the LEE could be transferred to a new susceptible bacterial host.

5.2. Results

5.2.1. Viability of phages induced from EC1810

Phages induced from EC1810 were investigated for suitable conditions for both short term storage (up to 28 days) and long term storage (up to 56 days). Short term storage conditions included storage in glass bottles and polypropylene tubes without chloroform, and under these conditions, phage lysates showed no reduced viability up to 3 days but a drop of 1 log pfu ml^{-1} per week thereafter was observed (Figure 5.1). Storage with chloroform in both glass bottles and polypropylene tubes, caused a loss in viability with a 1 log pfu ml^{-1} drop within 3 days and subsequently, a 1 log pfu ml^{-1} drop in titre per week (Figure 5.1). Data collection for short term conditions was stopped at day 14 as all conditions showed > 99% drop in phage titre. Long-term

storage in cryovials at both -20°C and -80°C in the presence of 10% glycerol showed that phage stocks could not be maintained under these conditions without a drop in titre. At 28 days, phages stored at -20°C had a titre drop of 2 logs pfu ml^{-1} while storage at -80°C resulted in a titre drop of 3 log pfu ml^{-1} (Figure 5.1). The time point at day 56 showed a further drop of 1 log pfu ml^{-1} for storage at both -20°C and -80°C (Figure 5.1). These experiments showed that phage lysates of EC1810 were unstable overtime and storage of more than three days was not recommended.

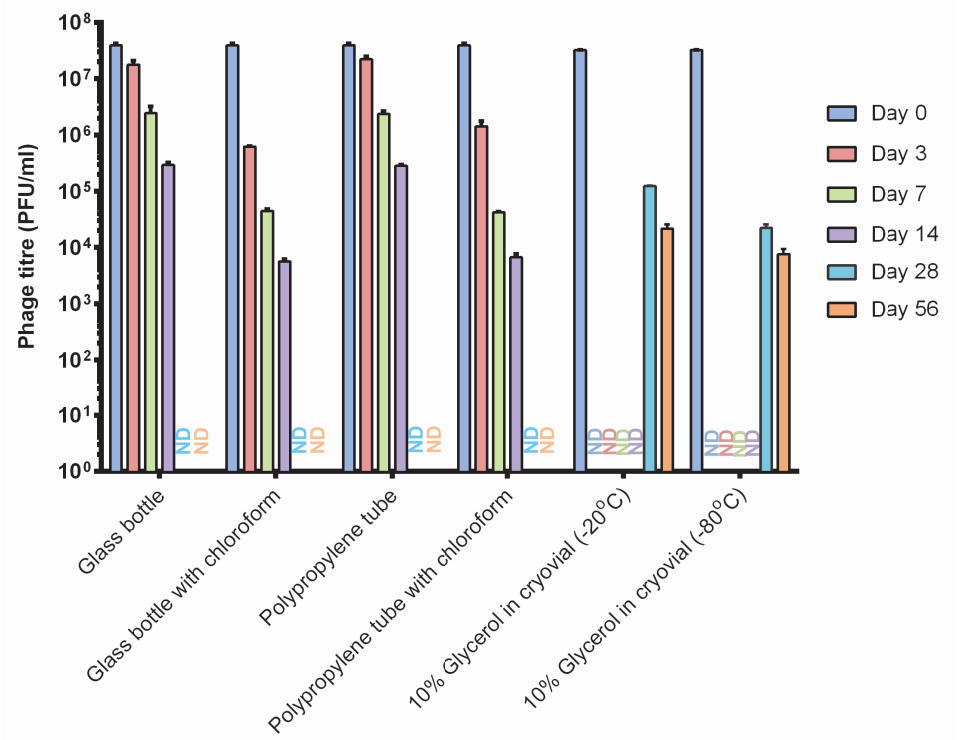


Figure 5.1: Investigation of 1810Stx2a stability over time, in various storage conditions. Phage titres for each storage condition are the mean of three biological replicates. Error bars represent standard deviation. Time points marked ND represents data points that were not determined due to either significant phage loss (short term storage) or unrequired data points (long term storage).

5.2.2. Morphology of the Stx2a prophage of EC1810

TEM of negatively stained induced phage lysates from EC1810 consisted of phages with a single morphology consisting of an icosahedral head of 68 ± 2 nm diameter (n = 48) with a short tail 20 ± 2 nm long (n = 48) (Figure 5.2) classifying 1810Stx2a as a podovirus (Ackermann, 2009).

The morphology and size of 1810Stx2a was cross-validated with phage lysates of STEC O157 EC3192 and laboratory strain K-12 strain Q358Sm^R along with their respective 1810Stx2a lysogenised derivatives FNS45 and EC3001. TEM of mitomycin C induced phage lysates of FNS45 and EC3001 revealed virions with the same morphology and size as in the EC1810 induced phage lysates (Figure 5.2, Figure 5.3) and they were able to form plaques on their indicator bacterial strains Q358Sm^R and MG1655 respectively. However, lysates from the respective K-12 parent strains, EC3192 and Q358Sm^R had no detectable phage virions in TEM studies, and were unable able to form plaques on lawns of the indicator strains. This similarity in phage morphology and size, coupled with no observed phage virions in the crude phage lysate preparations of 1810Stx2a negative strains, confirmed that the short tailed virions observed in FNS45 and EC3001 are representative of 1810Stx2a. The Stx2a phage of EC1810 also possessed the same morphology and size to AU5Stx1 and AU6Stx1 (Chapter 3) as well as to other characterised Stx2a phages, 933W (Plunkett *et al.*, 1999), 24_B (Allison *et al.*, 2003), VT2-Sakai (Asadulghani *et al.*, 2009) and P13374 (Beutin *et al.*, 2012).

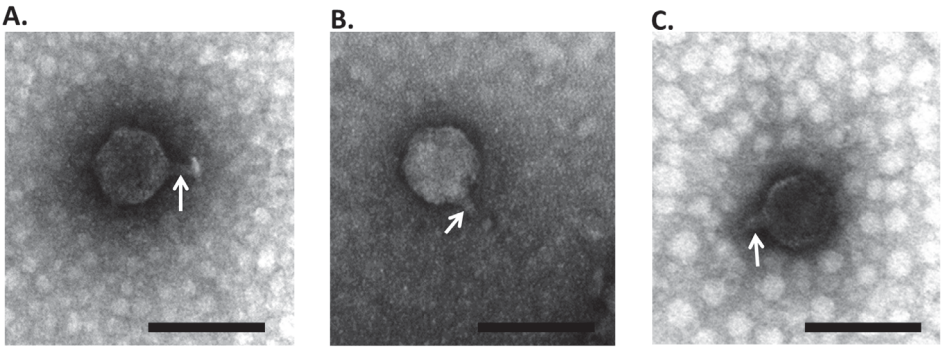


Figure 5.2: Morphology of 1810Stx2a. Representative electron micrographs of negatively stained (uranyl acetate) phage lysates induced from (A) EC1810, (B) FNS45 and (C) EC3001. Phages were viewed at 120,000x and phage tails are highlighted by a white arrow while the black bar represents a scale of 100 nm. Phages were view on the JEOL 1010.

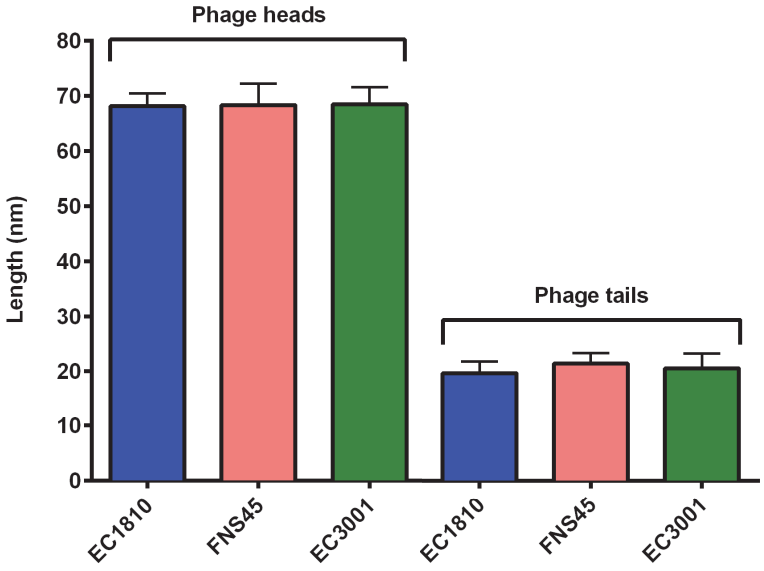


Figure 5.3: Histogram showing the measurements of phage virions viewed under the TEM. The mean length of phage heads and tails from phages induced from EC1810 (n= 48), FNS 45 (n=10) and EC3001 (n=10) is shown with error bars representing the standard deviation.

5.2.3. Host range of 1810Stx2a

A kanamycin marked (Δ*stx*$_2$::*kan*) derivative of 1810Stx2a (henceforth referred to as 1810Stx2K: Table 2.2), was used in a host range test against a panel of *E. coli* strains. Susceptibility to infection by 1810Stx2K was scored by both the ability to form plaques on a lawn of tested strains and the isolation of kanamycin resistant colonies harbouring Δ*stx*$_{2a}$::*kan* (when tested by PCR using primers in Table 2.4) after an infection in a liquid culture. The laboratory *E. coli* K-12 strains Q358sm^R, DH5α and MG1655 could be lysed and lysogenised by 1810Stx2K (Table 5.1). No plaques were detected on lawns from a clinical STEC O157 isolate (EC3192) but stable Δ*stx*$_{2a}$::*kan* lysogens could be formed, indicating that EC3192 was susceptible to infection by 1810Stx2K (Table 5.1). The ability to form Δ*stx*$_{2a}$::*kan* lysogens indicated that the non-plaquing observed on EC3192 was likely to be the result of preferential progression into the lysogenic life cycle, a phenomenon previously observed in the Stx2a phage from STEC O157 strain Sakai (Islam *et al.*, 2012).

Table 5.1: Susceptibility of bacterial strains to 1810Stx2K

	Infection with 1810Stx2K resulting in	
***E. coli* strains**	**Plaque formation**	**Δ*stx*$_{2a}$::*kan* lysogens**
Q358sm^R	+	+
DH5α	+	+
MG1655	+	+
EC3192	-	+

To determine if 1810Stx2a lysogens could resist infection to super infecting isogenic phages, phage lysates prepared from EC1810 were titrated onto a lawn of both Q358sm^R and a lawn of EC3001 (Q358sm^R lysogenised with 1810Stx2a). Phage titres at 10^7 pfu ml^{-1} were calculated from plaques observed on the lawn of Q358sm^R (does not harbour 1810Stx2a) but no plaques were seen on the lawn of EC3001 (1810Stx2a lysogen) (Figure 5.4). This suggested that the presence of an 1810Stx2a prophage in the recipient could confer immunity against superinfection of isogenic phages.

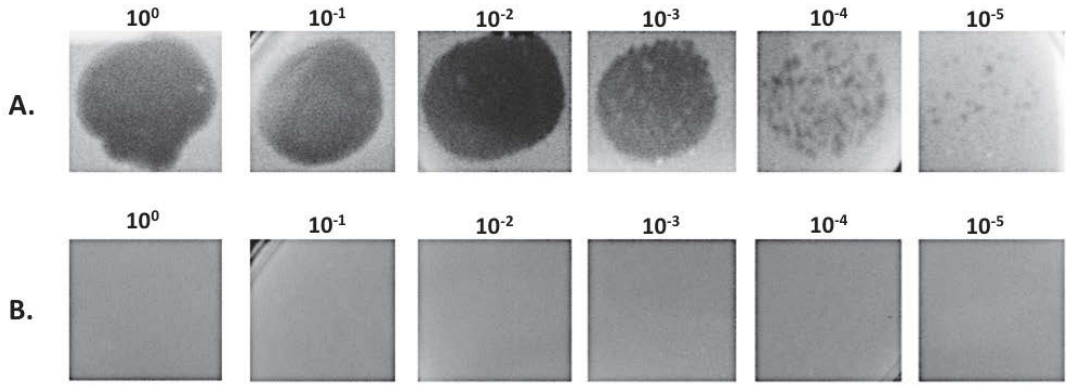


Figure 5.4: 1810Stx2a plaque formation on indicator strains. Spots of serially diluted phage lysate induced from EC1810 titrated onto a lawn of (A) Q358sm^R and (B) EC3001.

5.2.4. Transduction of Δ*eae*::*cat* to recipient *E. coli* strains

To test if the LEE could be transduced, phage lysates prepared from FNS8 (EC1810 with a *cat* cassette inserted into *eae*; formed by Michael Lietner from CSIRO), was used to infect a range of recipient strains including STEC O157 strains and laboratory *E. coli* K-12 strains (Figure 5.5). Included amongst the recipient strains are 1810Stx2a lysogenised derivatives. The formation of transducing particles (phage capsids containing bacterial DNA) occurs at low frequency as compared to phages particles

(Fineran *et al.*, 2009b) and it is possible that the infectious phages could infect and kill any transductants formed in the transduction experiment. The use of these 1810Stx2a lysogens as recipient strains could thus be exploited to increase the chance of stable transductant formation by preventing lysis from an infection by 1810Stx2a

Phage lysates were added into overnight liquid cultures of each recipient and together with controls for spontaneous mutation to chloramphenicol resistance and lysate contamination, were left to incubate over three days at 37^{o}C. Chloramphenicol resistant colonies were recovered from biological replicates of infection from FNS8 phage lysates to FNS9 and FNS45 (Figure 5.5). Only one chloramphenicol resistant colony was recovered from three biological replicates of infection from FNS8 phage lysates to EC3192 (Figure 5.5). No colonies were recovered from infections of *E. coli* K-12 strains (Figure 5.5). PCR analysis using primers eae_del_testF and eae_del_testR (Table 2.4) on the transductants derived from FNS9, EC3192 and FNS45 confirmed that chloramphenicol resistance is due to Δ*eae::cat* (Figure 5.6).

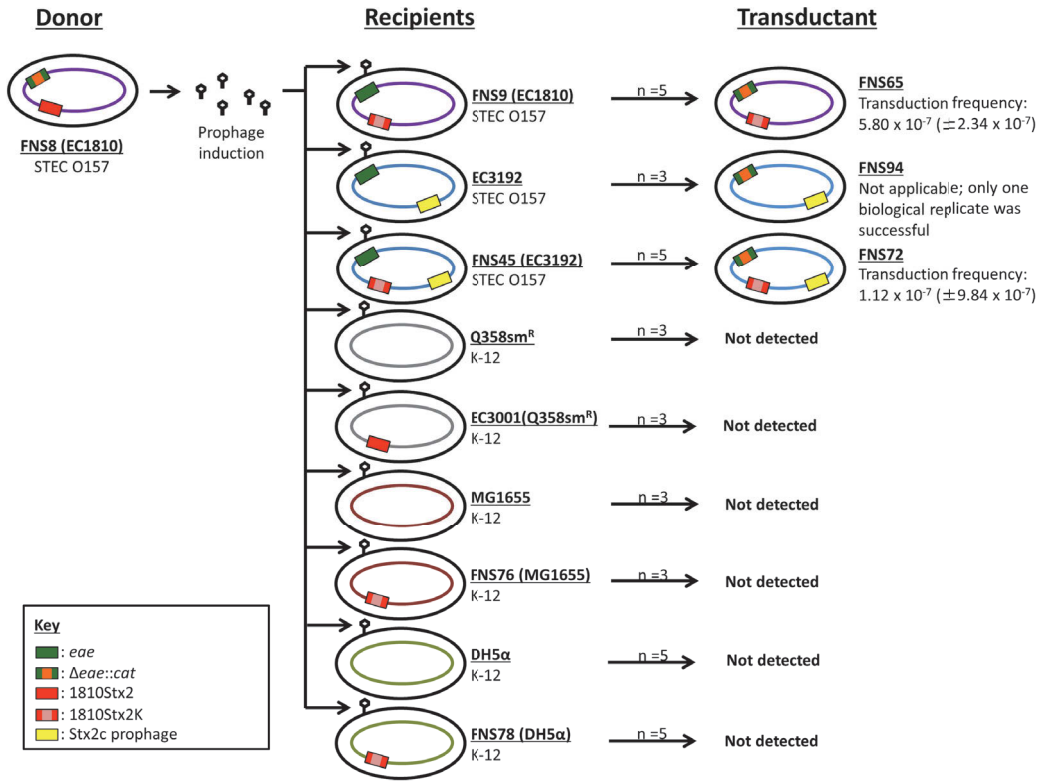


Figure 5.5: Overview of transduction experiment. Transduction of Δ*eae::cat* was performed by infecting various recipient strains with the phage lysate of FNS8. Strains with similar genetic background are colour coded with the same colour. Relevant genetic markers are coloured as per figure key. Number of biological replicates and the mean transduction frequency (number of transductants per pfu ml^{-1}) of successful transductions are reflected in the figure.

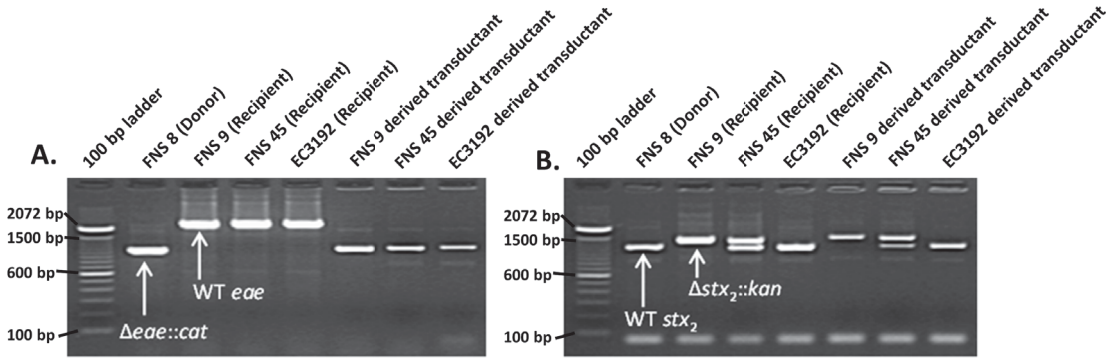


Figure 5.6: PCR analysis of donor, recipients and transductants. PCR was performed using the primers listed in table 2.4 and products were ran on a 2% agarose gel stained with ethidium bromide (A) PCR analysis detecting insertion of Δ*eae::cat*. WT *eae* will yield a product of 2805 bp while Δ*eae::cat* will yield of a product of 1094 bp (B) PCR analysis of carriage of *stx* gene. WT stx_{2a}/stx_{2c} will yield a product of 1241 bp while Δ stx_{2a}*::kan* will yield a product of 1576 bp.

There is a slight difference between the mean transduction frequency from FNS8 to FNS9 ($5.80 \times 10^{-7} \pm 2.34 \times 10^{-7}$) and FNS45 ($1.12 \times 10^{-7} \pm 9.84 \times 10^{-7}$). This difference in transduction frequency could be due the recipient genetic background. FNS8 and FNS9 are both derivatives of STEC O157 strain EC1810 while FNS45 is a different STEC O157 strain (derived from EC3192) (Table 2.3). The transduced DNA from FNS8 will not be recognised as foreign by FNS9 due to it being an isogenic strain while the transduced DNA will be recognised as foreign by FNS45 and may be subjected to degradation by any restriction-modification systems unique to the recipient (Bickle & Kruger, 1993) preventing transductant formation

The mean transduction frequency of FNS8 to FNS45 ($1.12 \times 10^{-7} \pm 9.84 \times 10^{-7}$) and EC3192 ($3.37 \times 10^{-9} \pm 5.83 \times 10^{-9}$) shows a difference of 2 logs. This difference in transduction frequency could be due to superinfection immunity as FNS45 is an 1810Stx2K lysogen, whereas EC3192 is not (Figure 5.5). This suggested that the 1810Stx2K prophage offered a protective effect, preventing the FNS45 derived transductants from lysis by any superinfecting 1810Stx2a phages.

5.2.5. Presence of 1810Stx2a is required for further dissemination of Δ*eae*::*cat*

Previous EC1810 phage metagenome sequencing data showed that apart from the Stx2 prophage, elevated sequencing reads corresponding to a cryptic prophage upstream of the LEE were also detected (Petty *et al*., unpublished data). To determine if the cryptic prophage was co-transduced with Δ*eae*::*cat*, primers 1810cpF and 1810cpR (Table 2.4) targeting a unique region of this cryptic prophage, were designed and tested on EC1810, FNS8, EC3192, FNS45 and FNS72 (Figure 5.7). The cryptic prophage was only detected in EC1810 and its derivative FNS8 but not in EC3192 or the EC3192 derived transductant FNS72 (Figure 5.5; Figure 5.7). This indicated that the cryptic prophage was not co-transduced with Δ*eae*::*cat*.

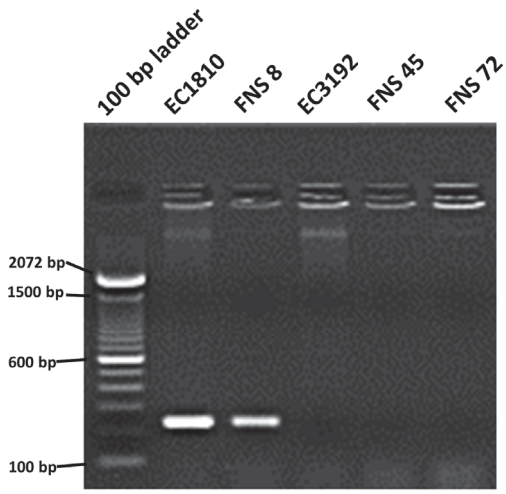


Figure 5.7: Detection of EC1810 cryptic prophage by PCR. Primers targeting a specific sequence within the cryptic prophage will yield an amplicon of 233 bp. High molecular weight faint bands near the well of agarose formed due to high DNA concentration in the template prior to PCR.

To determine if an 1810Stx2a prophage integrated in donor strains was necessary for the transduction of Δ*eae::cat*, phage lysates prepared from FNS72 (EC3192 derivative lysogenised with 1810Stx2K) and FNS94 (EC3192 derivative; Δ*eae::cat*; not lysogenised with 1810Stx2K) were used to infect (n=3) a spontaneous rifampicin resistant isogenic strain, FNS105 (Figure 5.8). Chloramphenicol, kanamycin and rifampicin triple resistant colonies were obtained following infection with the phage lysate from FNS72 but not FNS94 (Figure 5.8). All triple antibiotic resistant colonies all tested positive for Δ*eae*::*cat* by PCR. No detectable phages were produced by FNS94 ($< 10^2$ pfu ml^{-1}) compared to 9.09 pfu ml^{-1} produced by FNS72, and this, taken together with the lack of any colonies for any of the controls for spontaneous antibiotic resistance, confirmed that the method of Δ*eae*::*cat* transfer was phage-mediated. These results, coupled with the non-detection of the cryptic prophage in EC3192 strains, indicated that 1810Stx2a is key in mediating the mobilisation of Δ*eae*::*cat*.

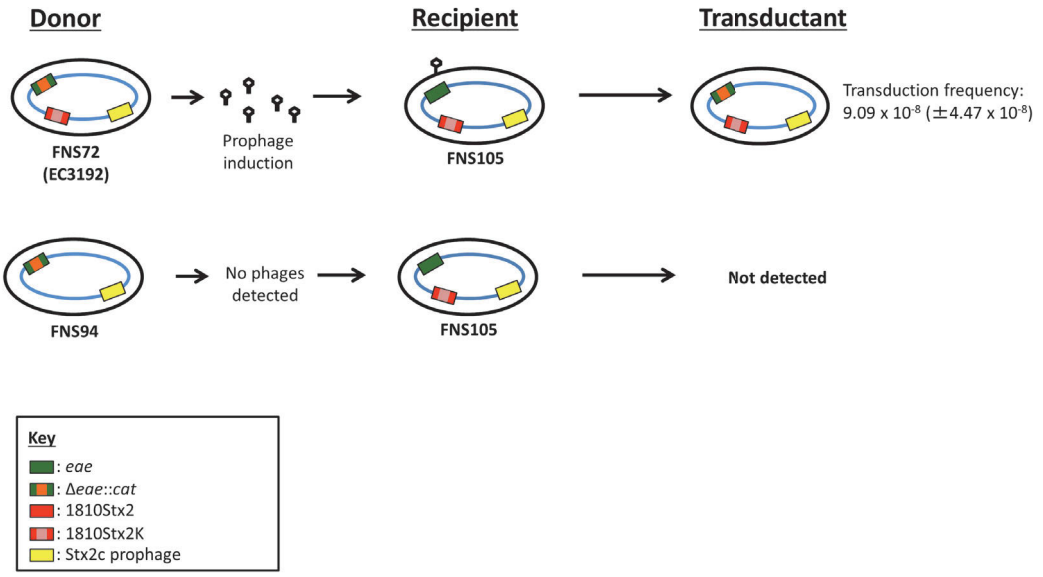


Figure 5.8: Schematic representation of transduction experiment to determine if 1810Stx2 is required for transduction. Strains with similar genetic background are colour coded with the same colour. Relevant genetic markers are marked up as per figure key. Three biological replicates were performed and the mean transduction frequency (number of transductants per pfu ml^{-1}) of successful transductions of Δ*eae::cat* to FNS105 is reflected in the figure.

5.2.6. Optimisation of transduction protocol to determine if transduction into non-STEC O157 strains could occur

Optimisation of the transduction protocol was performed in an attempt to improve the frequency of transduction to see if transduction into non-STEC O157 strains could be detected. Optimisation was performed using FNS8 as donor and FNS9 as recipient, which was the combination that yielded the highest transduction frequency (Figure 5.5). Three parameters were investigated sequentially: (1) duration of infection, (2) effects of divalent ions and (3) multiplicity of infection (MOI). The first parameter investigated was time of static incubation after the addition of phage lysates into recipient culture. The infection duration of three days was long which could lead to situation where a potential transductant could be killed by a phage in the experimental system or by other molecules (e.g. colicins) in the phage lysate. Time periods of 20 minutes, 1, 2, 4, 20, 44 and 68 hours post addition of phage lysates were investigated. Incubation duration of 20 minutes, 1, 2 and 4 hours yielded no detectable transductants while 20 (n=1; transduction frequency = 6.5 x 10^7 pfu ml^{-1}), 44 (n=1; transduction frequency = 8.1 x 10^7 pfu ml^{-1}) and 68 hours (n=1; transduction

frequency = 9.7 x 10^7 pfu ml^{-1}) each yielded transductants at a transduction frequency of 10^{-7} transductants per pfu. As the transduction frequency for a time period of 20, 44 and 68 hours are within the same log scale, the infection duration of 20 hours was determined to be the shortest time required to yield Δ*eae*::*cat* transductants.

Divalent ions have been previously shown to play a key role in phage infection (Bonhivers & Letellier, 1995; Islam *et al.*, 2012). When divalent ions were added to an overnight culture of FNS9 prior to phage infection, the frequency of transduction increased (Figure 5.9). The addition of either 10mM $CaCl_2$ or 10mM $MgCl_2$ significantly increased the transduction frequency (Figure 5.9). The highest transductant frequency obtained was with the addition of both 10mM $CaCl_2$ and 10mM $MgCl_2$ in combination and was determined to be an effective modification to the transduction protocol to increase Δ*eae*::*cat* transductant yield (Figure 5.9).

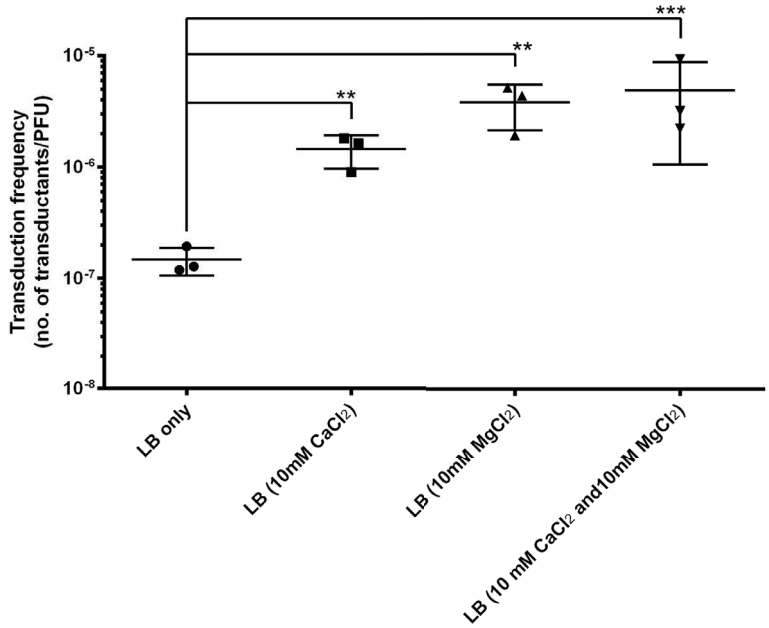


Figure 5.9: Effect of diavlaent ion addition on rate of transduction. Individual values from three separate experiments are plotted on the graph. The horizontal line represents the mean frequency of transduction from FNS8 to FNS9 while the error bars represent the standard deviation. Significance was calculated by unpaired *t*-test and reflected on the graph as stars. 2 (*) represents $P \leq 0.01$ while 3 (*) represents $P \leq 0.001$

The effect of MOI on transduction frequency was investigated over a range of cell numbers (Table 5.2). Some MOIs were not tested due to induced phage titre being a limiting factor capped at 10^8 pfu ml^{-1}. No transductants were detected from an infection of 1 x 10^7 recipient cells from all tested MOIs (Table 5.2). Recipient cultures with 1 x 10^8 cells were only observed to yield transductants at a MOI of 1, whereas recipient cultures with 1 x 10^9 and 1 x 10^{10} cells were able to yield Δ*eae*::*cat* transductants under all tested MOI conditions (Table 5.2). Of all the MOI tested, an infection of 1 x 10^{10} cells with 1 x 10^8 pfu ml^{-1} (MOI: 0.01) was observed to produce the highest transduction frequency of 1.78 x 10^{-5} transductants per pfu (Table 5.2).

Table 5.2: Frequency of transduction (transductants per pfu) of Δ*eae*::*cat* from FNS8 to FNS9 with varied MOI and cell numbers.

	Multiplicity of infection			
CFU	**0.001**	**0.01**	**0.1**	**1**
1 x 10^7	N.T	< 1 x 10^{-5}	< 1 x 10^{-6}	< 1 x 10^{-7}
1 x 10^8	N.T	< 1 x 10^{-6}	< 1 x 10^{-7}	2.40 x 10^{-7}
1 x 10^9	N.T	9.00 x 10^{-7}	6.40 x 10^{-7}	N.T
1 x 10^{10}	4.80 x 10^{-6}	1.78x 10^{-5}	N.T	N.T

Based on the results of optimisation, the optimised transduction protocol consisted of an infection of 1 x 10^{10} cells at an MOI of 0.01 for a period of 20 hours, in the presence of both 10 mM $CaCl_2$ and 10 mM $MgCl_2$ prior to plating onto selective media to select for transductants.

5.2.7. Repeat of transduction using optimised transduction protocol

The test for transduction was repeated using the optimised transduction protocol as aforementioned to determine if Δ*eae*::*cat* can be transduced from FNS8 into laboratory *E. coli* K-12 recipients (Figure 5.10). Similar to earlier results (Figure 5.5), transductants were only obtained from the recipient strains FNS9, EC3192 and FNS45 but transduction now occurred at higher frequencies (Figure 5.10). Even with the optimised conditions, no Δ*eae*::*cat* transductants could be obtained from an infection of laboratory K-12 strains. One explanation for this is that stable insertion of Δ*eae*::*cat* into the genome of transductants is likely to require homologous recombination with an existing *eae,* therefore presence of the LEE in the recipient

may be a requirement for successful transduction. Another explanation is that the DNA from the donor strain is recognised as foreign and degraded by the recipient defence systems.

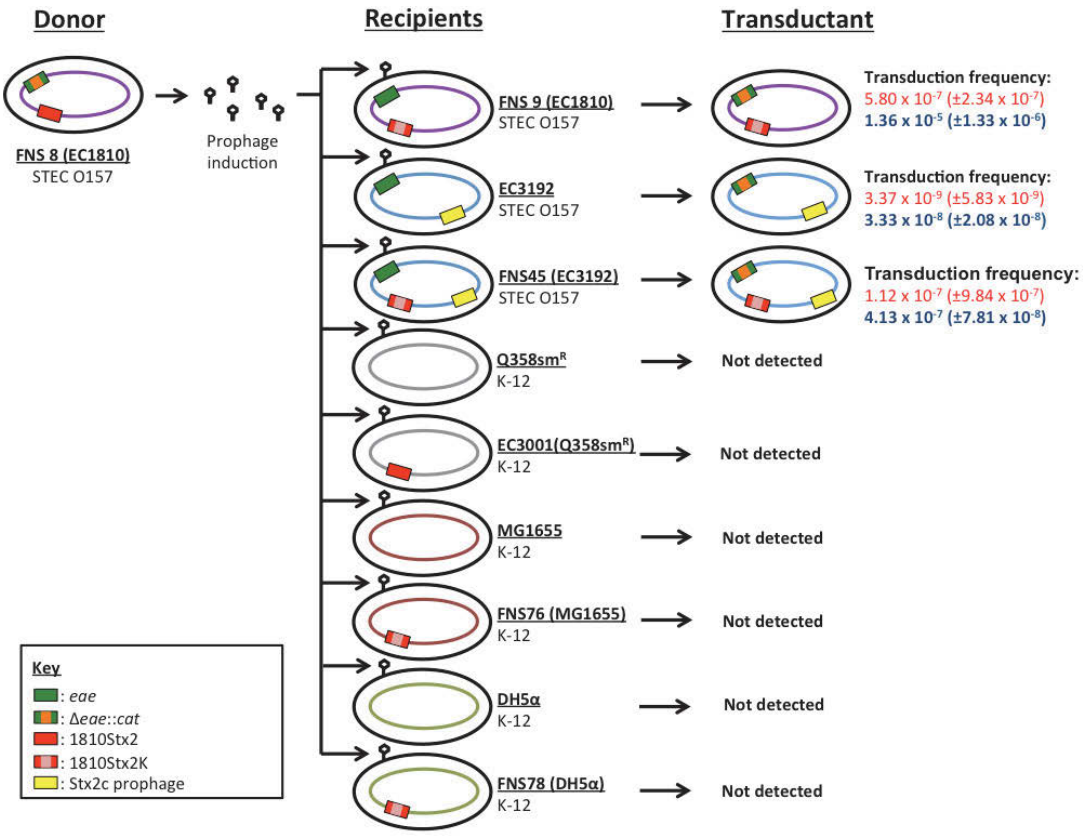


Figure 5.10: Repeat of transduction experiments using optimised conditions. Transduction of Δ*eae::cat* was performed by infecting various recipient strains with the phage lysate of FNS8 using optimised conditions. Strains with similar genetic background are colour coded with the same colour. Relevent genetic markers are marked up as per figure key. The mean (n=3) transduction frequency (number of transductants per pfu ml^{-1}) of successful transductions from previous transduction experiments (in red) and under optimised conditions (in blue) are reflected in the figure.

5.2.8. Genome assembly of STEC O157 strain FNS9, an attenuated derivative of EC1810

Bacterial DNA was extracted from the attenuated EC1810 derivative strain FNS9 (Δstx_{2a}::*kan*) sequenced using PacBio technology, *de-novo* assembled and annotated as described in Chapter 2. The assembly statistics of the genome sequence of FNS9 are listed in table 5.3.

Table 5.3: Genome assembly statistics of STEC O157 strain FNS9.

	FNS9 assembly
Total number of contigs	3
Total bases (bp)	5,705,932
Mean coverage	257.75
Maximum contig length (bp)	5,601,077
Minimum contig length (bp)	47,468

The assembly process generated three contigs with the bacterial chromosome assembling into a single contig of 5.6 Mb (contig 1; Table 5.3) Genome assembly also confirmed previous results of the LEE located adjacent to tRNA-*selC* with a cryptic prophage located upstream of the LEE and the Stx2 (Δstx_{2a}::*kan*) prophage located adjacent to tRNA-*argW* (Figure 5.11).

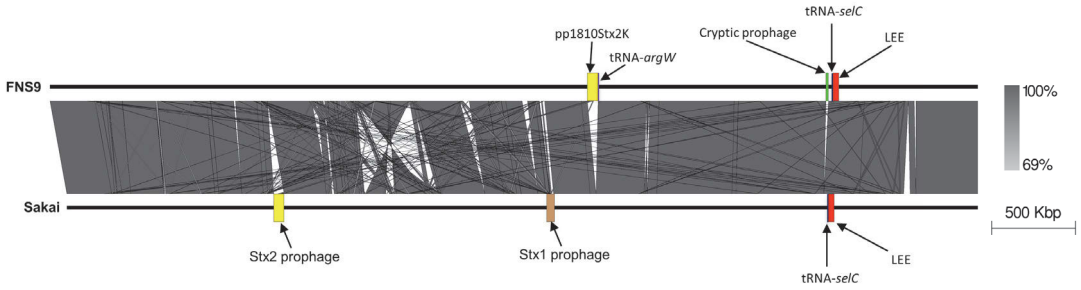


Figure 5.11: Comparison of the chromosomes of STEC O157 strains FNS9 and Sakai. BLASTn comparison between FNS9 and Sakai (Accession number: BA000007). Grey shading between the two genomes indicate regions of nucleotide similarity (BLASTn; minimum length of 1000 bp). Image was generated using Easyfig (Sullivan *et al.*, 2011).

5.2.9. Genomic analysis of pp1810Stx2K

The prophage pp1810Stx2K was 65,021 bp in length and has a G+C content of 50.6%, similar to that of FNS9 (50.5%). The annotated linear prophage genome is shown in Figure 5.12 and the predicted functional annotation list of pp1810Stx2K is provided in appendix VIII. When compared against published Stx phages (Figure 5.12), pp1810Stx2K was found to be highly similar to published Stx2a phages and two Stx1 phages, Stx1CPI and Hun2013, as well as the Australian Stx1 phages AU5Stx1 and AU6Stx1 (Chapter 3) and prophages pp14Stx1 and pp571Stx1 (Chapter 4), especially in the phage morphogenesis genes (Figure 5.12). The Stx1 phages

isolated from Australian STEC O157 strains, AU5Stx1 and AU6Stx1 (Chapter 3), together with the Stx1 phage Hun2013 (Sváb *et al.*, 2015) and Stx2a phages 933W (Plunkett *et al.*, 1999), phage 24_B (Allison *et al.*, 2003) and P13374 (Beutin *et al.*, 2012), as well as 1810Stx2a (Figure 5.2) were all previously shown to be podoviruses and this similarity in morphology corroborated with the sequence similarity in the phage morphogenesis genes. Previously, in Chapter 3, two host recognition domains (cterm1 and cterm2) were detected in the majority of known Stx phages within overlapping genes encoding a tail fibre and alternative tip. In the region of overlap, signatures for a +1 translation frameshift were detected which would allow for the production of a subpopulation of progeny phages with an alternative host recognition domain, extending the host range of the phage. Prophage pp1810Stx2K was found to encode cterm1 (subtype B) and cterm2 (subtype Y), the same subtypes found in all other Stx2a phages, which possessed a side tail fibre that could potentially undergo +1 translational frameshift (Table 3.3).

In the genome of FNS9, pp1810Stx2K was integrated adjacent to tRNA-*argW* and flanked by two identical 24 bp repeats (5'-GTCCTCTTAGTTAAATGGATATAA-3') identical to the *attB* sequence identified in other Australian Stx1 phages, AU5Stx1 and AU6Stx1 (Chapter 3) and Stx1 prophages pp14Stx1 and pp571Stx1 (Chapter 4) and the *attB* of 1810Stx2K was also confirmed to be the first 24 bp of the 5' end of tRNA-*argW*.

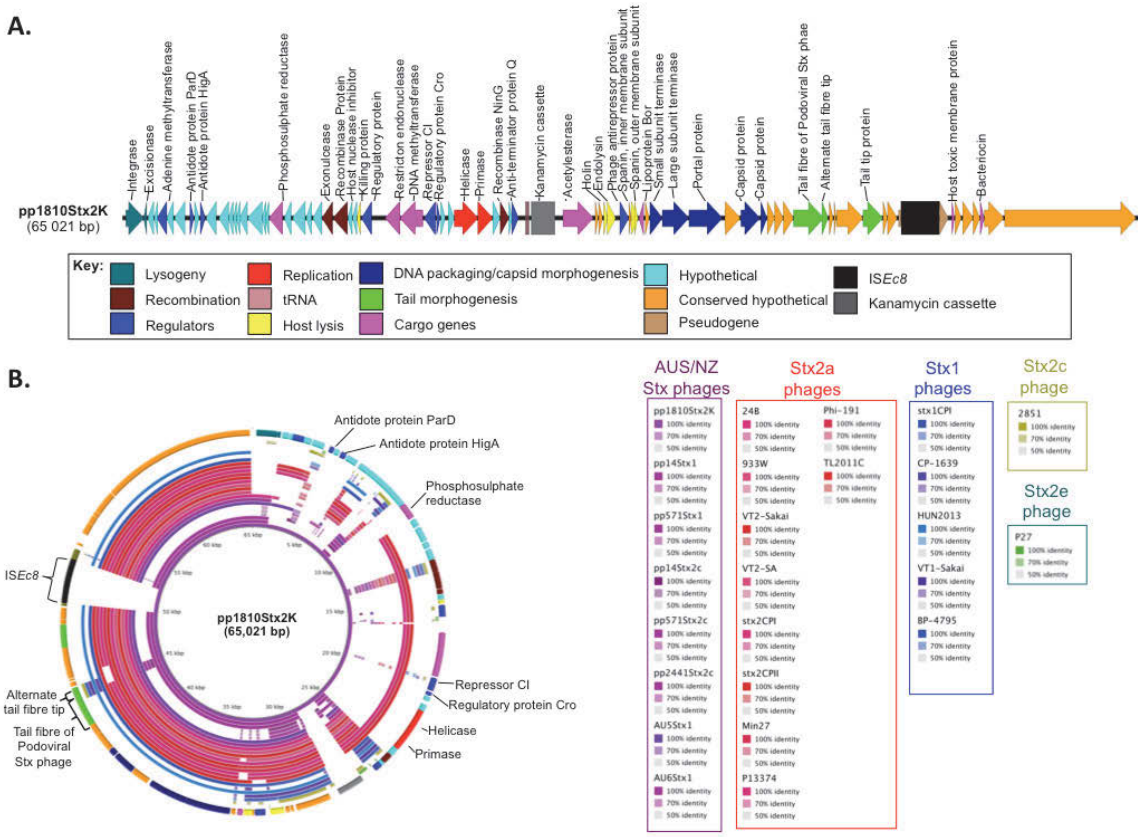


Figure 5.12: Genome of pp1810Stx2K. (A) The linear genome of pp1810Stx2K is presented here in its prophage state and genes were colour coded according to their function as per the key. Image was generated using Easyfig (Sullivan *et al.*, 2011). (B) pp1810Stx2K is presented as a reference genome in the centre with a circular representation of its genetic map. The outer most ring are the phage annotation, colour coded as per the key above. Each of the coloured rings represent a different query phage genome, and the coloured regions are BLASTn matches to the reference. Query genomes and their nucleotde identity are colour coded as listed in the figure key. Image was generated using BRIG (Alikhan *et al.*, 2011)

Genes predicted to encode the phage CI repressor and the regulatory protein Cro were identified in the genome of pp1810Stx2K based on similarity to CI (61% amino acid identity over entire length) and Cro (50% amino acid identity over entire length) of phage lambda. Downstream of *cro* in phage lambda is the gene *cII* that encodes a transcriptional activator, an important regulator for the progression to the lysogenic life cycle (Echols & Green, 1971). A common motif found in phage regulators is the presence of a helix-turn-helix motif, structural motifs that can bind DNA (Brennan & Matthews, 1989; Ho *et al.*, 1992; Neufing *et al.*, 1996; Datta *et al.*, 2005). No helix-turn-helix was found from an Interpro (Hunter *et al.*, 2012) scan of the CDS downstream of *cro* in pp1810Stx2K, which suggested that pp1810Stx2K might not possess the transcriptional activator CII. Inspection of the entire prophage genome did not reveal a putative *cII* gene elsewhere in the genome. Not all temperate phages encode *cII*, and this gene has only previously been detected in the genomes of phages lambda (Schwarz *et al.*, 1978), P22 (Ho *et al.*, 1992) and 434 (Grosschedl & Schwarz, 1979). This suggested that the 'decision' for the progression to the lysogenic life cycle after an infection with pp1810Stx2K might be regulated differently from phage lambda. Two genes downstream of *cro* were predicted to encode a helicase and a primase (Figure 5.12), both of which are involved in DNA replication (Weigel & Seitz, 2006). These genes were similar (>96% nucleotide identity over the entire length) to genes in Stx2a phages P13374, phi-191 and TL2011c but no homologues were found in the genomes of other known Stx phages (Figure 5.12)

The cargo gene *lom,* which confers its host the ability to adhere to epithelial cells (Vica Pacheco *et al.*, 1997), was present in pp1810Stx2K but was disrupted by an IS*Ec8* insertion. This IS element is not found in the genomes of other Stx phages (Figure 5.12). Three additional IS*Ec*8 copies were detected elsewhere in the genome of FNS9, fewer than that of STEC O157 strain Sakai which harboured eight copies (Hayashi *et al.*, 2001). Previous studies have shown that IS*Ec8* is an active IS element in STEC O157 strain Sakai, actively transposing and showing preferential insertion into mobile genetic elements (Ooka *et al.*, 2009). Using the study by Ooka *et al.*, (2009) as circumstantial evidence for IS*Ec8* mobiliy, this IS element in 1810Stx2K might have been transposed into its current location from its host and this could account for the detection of IS*Ec8* in pp1810Stx2K.

The prophage pp1810Stx2K encodes three unique coding sequences not previously found in other known Stx phages. Two of these coding sequences are predicted to encode proteins that are structurally similar to ParD of *Caulobacter crescentus* (Dalton & Crosson, 2010) and and HigA of *Proteus vulgaris* (Schureck *et al.*, 2014) (Figure 5.13). Both ParD and HigA are the anti-toxin components of the toxin-antitoxin systems *parDE* and *higAB* respectively. One mode of bacterial defence against phage infection is abortive infection mediated by the toxin-antitoxin system (Fineran *et al.*, 2009a). However, phages can overcome abortive infection by producing anti-toxin molecules that can counteract the toxin produced by its bacterial host (Blower *et al.*, 2012; Otsuka & Yonesaki, 2012) and the genes that encode anti-toxin proteins in pp1810Stx2K might be a countermeasure used by the phage to circumvent host defence and establish its infection. As both coding sequences have a function that benefits the phage, they are likely to be phage borne regulator genes. The third coding sequence (locus tag: 1810STX2_200) was predicted to encode a protein that was structurally similar (Figure 5.13) to a phosphosulfate reductase in *E. coli* (Chartron *et al*., 2007). *E. coli* strains harbour a chromosomally encoded gene (*cysH*) that expresses a sulfonucleotide reductase, which reduces activated sulphate into sulphite (Krone *et al.*, 1991). A copy of *cysH* was also located in the genome of FNS9 but no significant nucleotide similarity was detected between the chromosomally encoded *cysH* of FNS9 and 1810STX2_200. In addition and both protein products shared only 38% amino acid identity over 23% of their amino acid sequence. As the phage borne phosphosulfate reductase is dissimilar to the chromosomally encoded phosphosulfate reductase gene, it is likely that both genes may function differently from each other. As this reductase is unlikely to influence the phage life cycle, it is likely to be a cargo gene.

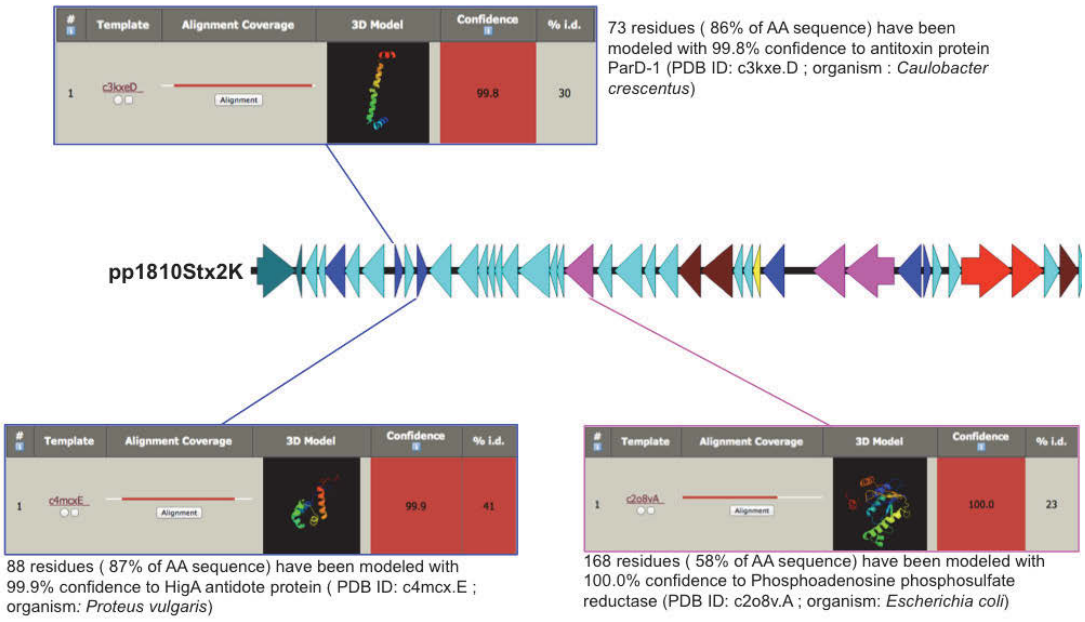


Figure 5.13: Protein folding prediction of the three unique genes carried by pp1810Stx2K. Protein folding prediction was performed using Phyre2 (Kelley *et al.*, 2015). Genome subregion image was generated using Easyfig (Sullivan *et al.*, 2011).

5.3. Discussion

In this chapter, biological evidence for the transduction of a LEE encoded gene by a Stx2a prophage was presented. In addition, in-*silico* analysis of the Stx2a prophage showed that this prophage had a different replication and repressor system and also harboured additional genes never previously observed in Stx phages.

5.3.1. A new DNA replication system for Stx phages

It has been established that Stx phages are genetic mosaics of each other and show variation in their genomes, around the common factor of the *stx* operon (Allison, 2007). A multi-locus PCR screen was previously developed to characterise Stx phages, based on the genomic sequences publicly available at that time (Smith *et al.*, 2007b; Smith *et al.*, 2012) and it was observed that there were five genetically different replication modules (O_1/P_1 through O_5/P_5) associated with Stx phages (Smith *et al.*, 2012). The replication module of pp1810Stx2K, which is genetically highly similar to those of other Stx2a phages P13374, phi-191 and TL2011c, are genetically distinct from the five-replication systems (Figure 5.12) which indicates that there are now six different replication systems that are associated with Stx phages.

5.3.2. pp1810Stx2k carries genes that enhances its chances of infection

Prophage pp1810Stx2K harboured two genes predicted to encode two different yet structurally related anti-toxins, with similarity to the antitoxins from two distinct toxin-antitoxin systems, ParDE and HigAB, which were first described in a broad host plasmid RK2 (Johnson *et al.*, 1996) and in the chromosome of *E. coli* (Christensen-Dalsgaard *et al.*, 2010) respectively. Toxin-antitoxin systems (also known as addiction modules) consist of two genes that encode a toxin and its neutralising antitoxin and was first demonstrated to have a functional role in plasmid inheritance (Thisted *et al.*, 1994). Since then, a growing body of evidence has shown that toxin-antitoxin systems have a wider role in the bacterial cellular biology and that chromosomally encoded toxin-antitoxin systems are a form of anti-phage defence utilised by bacteria (Pecota & Wood, 1996; Hazan & Engelberg-Kulka, 2004; Fineran *et al.*, 2009a). A phage infection will result in the imbalance of the ratio of anti-toxin and toxin which will allow the toxin to induce cell death, leading to the abortion of phage infection (Samson *et al.*, 2013). However, some phages can evade this host defence by producing an anti-toxin molecule, which allows the phage to neutralise the toxic effect of the host-encoded toxin thus preventing an abortive infection (Blower *et al.*, 2012; Otsuka & Yonesaki, 2012).

The carriage of anti-toxin genes, coupled with the potential to form alternative tail fibres with different host recognition domains (as described in Chapters 3 and 4) confers upon pp1810Stx2K two measures to extend its host range by countering the bacterial defence as well as recognising more than one surface receptor to establish infection. In addition, the ability to form alternative tail fibres with different host recognition domains also means that there could be a mixed population of transducing particles (packaged with the LEE) with different tail fibres, which would potentially allow for the dissemination of the LEE to a broad host range, not limited to *E. coli*.

5.3.3. 1810Stx2a is likely to conform to the superinfection immunity paradigm

The ability of lysogenic bacteria to resist infection by superinfecting isogenic phages is due to expression of a phage repressor through a process called superinfection immunity (Nasi *et al.*, 1974). In the archetypical phage lambda, the phage repressor harboured by a phage is not only required for the maintenance of the lysogenic state but also to block the expression of the integrase from superinfecting isogenic phages (Nasi *et al.*, 1974). However, in repressor mutants of phage lambda, the process of superinfection immunity could be circumvented (Freifelder & Kirschner, 1971). The results presented in this chapter showed that 1810Stx2a conformed to the superinfection model whereby laboratory K-12 strains lysogenised with 1810Stx2a were able to resist superinfection.

Of all phages with an intact repressor studied thus far, only the Stx2a phage $24_{B,}$ has been previously shown to circumvent superinfection immunity, leading to the integration of up to three isogenic Stx phages in three different locations in the bacterial genome (Fogg *et al.*, 2007). The genome of pp1810Stx2K is highly similar to the genome of $24_{B,}$ especially in the phage morphogenesis modules, but it possessed a different phage repressor genetically distinct from 24_B (Figure 5.12). The only other Stx phages that possessed an identical repressor to pp1810Stx2K were P13374, Phi-191 and TL2011c (Figure 5.12), indicating they would also conform to the lambda superinfection immunity model. However, there is also a possibility that superinfection might have occurred undetected by this assay, as only lysis on a lawn of recipient bacterial cells, and not multiple lysogeny was investigated. Future experiments on multiple lysogeny would have to be conducted to unequivocally prove that 1810Stx2a conforms to the superinfection immunity paradigm.

5.3.4. Evidence for transduction of the LEE

The potential for phage-mediated transfer of the LEE was investigated by transduction of a LEE marker (Δ*eae*::*cat*) conferring a phenotypic change to chloramphenicol resistance in the recipient strain and subsequent PCR confirmation of carriage of Δ*eae*::*cat* by this resistant strain. Successful transduction of Δ*eae*::*cat* into an isogenic STEC O157 strain, a heterogenic STEC O157 strain and an

1810Stx2a immune derivative of the heterogenic STEC O157 strain was detected and no transduction was detected for *E. coli* K-12 recipients (Figure 5.5). It was observed that the frequency of transduction was higher in bacterial strains harbouring an 1810Stx2a prophage as opposed to recipient strains without a 1810Stx2a prophage. This was likely due to superinfection immunity, whereby the 1810Stx2a prophage prevents blocks infection of other phages with similar phage repressors and protecting its bacterial host from succumbing to the lytic life. Transduction frequencies were also higher in isogenic recipients compared to heterogenic LEE-encoding strains. This difference in transduction frequency is likely to be due to bacterial defence against incoming foreign DNA in the recipient strains. In a heterogenic strain, the incoming foreign DNA injected from the transducing particle could be recognised by any recipient DNA restriction-modification (R-M) systems as non-self and would be subsequently cleaved by the restriction endonuclease of the R-M system (Bickle & Kruger, 1993). When the bacteria chromosome of FNS9 was compared to *E. coli* K-12 MG1655, it was observed that FNS9 did not possess R-M genes of *E. coli* K-12 (*hsdSMR*, *mrr*, *mcrA*, and *mcrBC*) indicative that both strains have a different set of R-M systems, further providing evidence for a barrier against transduction.

Another plausible reason for the difference in transduction rate between an isogenic and heterogenic recipient could be degradation by RecBCD, an enzyme that degrades unwanted linear dsDNA (Dillingham & Kowalczykowski, 2008). However, RecBCD also has a contradictory function where the enzyme also participates in homologous recombination (Dillingham & Kowalczykowski, 2008), which in the case of DNA injected from a transducing particle might be a double edge sword, as it might conversely aid in the recombination of the transferred donor DNA with the genome of recipient strains to form transductants.

5.3.5. Mechanism of LEE mobilisation

Phages induced from Δ*eae*::*cat* transductants that harboured 1810Stx2a were also capable of disseminating Δ*eae*::*cat* to other heterogenic STEC O157 strains harbouring the LEE (Figure 5.8). However, phage-mediated transfer of Δ*eae*::*cat* from transductants without 1810Stx2a was not detected, confirming that the mechanism for transfer was by 1810Stx2a-mediated transduction.

The cryptic prophage previously identified upstream of the LEE was only present in EC1810 derivative bacterial strains which showed that the cryptic prophage was not co-transduced with Δ*eae*::*cat*. Moreover, the demonstration that transductants harbouring both Δ*eae*::*cat* and 1810Stx2a are able to further disseminate Δ*eae*::*cat,* eliminated the cryptic prophage as a phage responsible for the mobilisation of the LEE. These data confirm that 1810Stx2a is the prophage responsible for the mobilisation of the LEE. In the literature, the best-studied systems of pathogenicity island (PAI) transfer by phages are the *Staphylococcus aureus* Pathogenicity Islands (SaPIs). The transfer of SaPIs was found to be due to the P4-like parasitic reliance of this PAI on helper prophages within the same genome (Novick *et al.*, 2010). SaPIs encode integrases and phage DNA packaging genes and when the helper phage gets induced by the SOS response or if the bacterial host is infected by a new temperate helper phage, the pathogenicity island excises and subsequently hijacks the capsid produced by the helper phage (Novick *et al.*, 2010). Another documented PAI transfer event is the transfer of a *Vibrio* pathogenicity island mediated by a generalised transducing phage (O'shea & Boyd, 2002). However, this generalised transducing phage was not an endogenous prophage in the donor strain carrying the pathogenicity island, but instead was a temperate phage introduced into the donor strain through laboratory infection (O'shea & Boyd, 2002).

No phage genes have been identified in the LEE PAI (Elliott *et al.*, 1998), and none were found in the LEE of EC1810. As such, it is unlikely that the mechanism of LEE mobilisation would be similar to that of the SaPIs which are phage-related chromosomal islands. Moreover no phages of different sizes or morphology were detected from phages lysates of EC1810, which further indicated that the mobilisation of the LEE was dissimilar to that of SaPIs where phages with smaller phage heads, which encapsulates the DNA of the SaPI, are formed.

Based on comparative genomic analysis of pp1810Stx2K to other Stx phages, it was observed that it is genomically highly similar to other published Stx2 phages, especially in the phage morphogenesis genes. This includes similarity to the large terminase of the podovirus Stx2a phage 933W. The amino acid sequence of the large terminase, which can be used to predict packaging mechanism (Casjens *et al.*, 2005), of 1810Stx2k has 99% amino acid sequence identity over its entire length to the large

terminase of 933W [which has been showed to cluster with other headful packaging phages (Casjens *et al.*, 2005)], suggesting that 1810Stx2K is a headful packaging phage. Phages that can perform generalised transduction are headful packaging phages, and generalised transduction is likely to account for the mobilisation mediated by 1810Stx2a. However, to confirm generalised transduction, the transduction assay should be repeated using a donor strain containing a range of different genomic markers, as performed for the classification of other generalised transducing phages (Budzik *et al.*, 2004; Petty *et al.*, 2006; Petty *et al.*, 2007; Battaglioli *et al.*, 2011).

In addition, the study conducted in this chapter showed that Δ*eae*::*cat* could only be transduced from *E. coli* strains containing the prophage 1810Stx2a. This suggested that the mobilisation of Δ*eae*::*cat,* and potentially the entire LEE pathogenicity island, could be due to a novel transduction event between the LEE and the Stx2a prophage: the specific mobilisation of non-adjacent bacteria genomic DNA in *trans* at high frequency.

In transduction experiments using laboratory *E. coli* K-12 strains and their 1810Stx2a immune derivatives, no transductants were detected. Unlike the STEC O157 strains, which were successfully transduced, *E. coli* K-12 strains do not harbour the LEE. As no transductants could be obtained from LEE-negative recipients, it may be that the incorporation of a transduced LEE is dependent on homologous recombination with an existing LEE in the recipient strain. Alternatively, it is also possible that transduction events could occur but incorporation of a large genomic region like the LEE occurred less frequently and thus was below the limit of detection of the assay. It will also be interesting to see if other Stx phages also have this ability to mobilise the LEE, through further experiments. For that purpose, the strain FNS94 formed in this chapter will prove valuable as a host to be lysogenised with other characterised Stx phages due to it containing a LEE marked gene. Once a panel of FNS94 Stx lysogens were formed, transduction assays could be performed using this panel of donors and the optimised transduction protocol described in this Chapter.

5.4 Conclusions

The Stx2a prophage 1810Stx2a is key to the mobilisation of the LEE, potentially via a novel transduction mechanism. The incorporation of Δ*eae*::*cat* into LEE positive recipient strains occurred most likely by homologous recombination. The horizontal transfer of the LEE has never been previously demonstrated and the work presented in this chapter shows that phages have a role in the dissemination of the LEE. The ability for a phage to package the LEE could be a mechanism for the dissemination of the LEE to other bacterial hosts, which can be enhanced by the predicted ability for the Stx phage to extend its host range by +1 translational frameshifting during tail fibre morphogenesis as well as evading the bacterial host anti-phage defence by encoding anti-toxins to prevent host death by the TA system. Taken together, this demonstrates the contribution of Stx phages to the formation of new pathogens.

6. Final discussion

This chapter summarises the findings in this thesis and provides an overview of the Stx phages of Australasian STEC O157. Genomic characterisation on all the Stx phages in this thesis showed that they are capable of infecting a broad host range due to the potential of producing a subpopulation of progeny phages possessing an alternate tail fibre tip. It was also observed that Stx production can be influenced by either the bacterial genetic background or by phage encoded anti-repressors. Two other significant results presented in this thesis is the suggestion that the *stx* operon might have first entered the phage genome as part of a MITE and secondly, the first mention of the ability for the mobilisation of an antibiotic resistant cassette marked LEE encoded gene by an endogenous Stx2a prophage. These combined findings are used to draw conclusions on the role of Stx phages in influencing the pathogenicity of their host.

6.1. Insights into the reduced virulence of Australian STEC O157

The production of Stx has been linked to the induction of the Stx prophage from its bacterial host (Neely & Friedman, 1998a; Neely & Friedman, 1998b; Plunkett et al., 1999). However, it has been documented that Stx production amongst different STEC O157 strains varies, with STEC strains isolated from patients showing severe clinical manifestations producing more toxin (Neupane et al., 2011; Mellor et al., 2015). In this thesis, circumstantial evidence suggests that a possible contributing factor to reduced toxin production in STEC O157 was disruption of phage genes by IS elements (Chapter 4). In STEC O157 strain Sakai, it was determined that IS*629* was the most abundant IS element (Hayashi et al., 2001). In addition to being an active IS element, IS*629* can also cause deletions of genomic regions adjacent to its insertion site (Ooka *et al*., 2009) and is able to excise imperfectly, leaving behind a truncated 3' region of the IS element at the insertion site (Kusumoto et al., 2011). One of the Stx1 prophages characterised in this thesis, pp571Stx1, has a truncated 3' region of IS*629* element located within its genome in a region which encodes genes crucial for the phage to switch to its lytic cycle, resulting in the deletion of regulation genes and the 5' of a replication gene (Chapter 4). Insertions by IS*629* have been previously shown

to directly disrupt the production of Stx by inserting into the *stx2B* gene (Kusumoto *et al.*, 1999). However, presented in Chapter 4 was the observation of an indirect disruption to Stx production where IS*629* deleted the genes required for the switch to the lytic life cycle in the Stx1 prophage, leading to the phage inability to switch to its lytic life cycle. While it is unable to switch to its lytic cycle, the genes for excision from the bacterial chromosome are intact and it from the results presented in Chapter 4, pp571Stx1 is likely to be capable of excision and packaged into empty phage heads formed by other prophages. In addition, this disruption resulted in the non-detection of toxin production from the bacterium. Taken together, these results show that the detection of *stx* genes in STEC O157 strains is not always indicative of their pathogenic potential, as disruptions to phage genome can affect toxin output, regardless of the presence of the *stx* genes.

From the analysis conducted in this thesis it was observed that out of the four Stx1 phages studied, only one Stx1 phage carried extra anti-repressor genes while the other three Stx1 phage did not (Chapter 3 and 4). It was also shown in a laboratory K-12 genetic background that the Stx1 prophage harbouring an extra anti-repressor gene could be induced at a higher titre than the other Stx1 prophage, which subsequently led to higher Stx production (Chapter 3). In addition, the Stx2c prophages characterised in this study were all found to carry an anti-repression operon, and results showing strains that carried both Stx1 and Stx2c prophages having higher induction of the Stx1 prophages as well as producing more Stx than a strain that only carries the Stx1 prophage. This indicates that this anti-repressor encoded by the Stx2c phage may alleviate the repression of the lytic cycle of the Stx1 prophages in-*trans* (Chapter 4). As such, the carriage of either an extra-antirepressor gene or an anti-repression operon by the Stx phages is likely to contribute to the observed variation in toxin production by different STEC strains.

Multiple studies on the Australian STEC O157 population showed that the dominant *stx* genotype is stx_{2c} either alone or in combination with stx_1. (Fegan & Desmarchelier, 2002; Mellor *et al.*, 2012; Mellor *et al.*, 2013). Quantitative PCR performed in Chapter 4 showed that in Australian STEC O157 strains that harboured two Stx prophages, one Stx prophage was induced at higher titres as compared to the other, a phenomenon that was also previously observed in two other STEC O157

strains harbouring two Stx prophages, strain Sakai from the 1996 outbreak in Japan and strain TW14359 from the 2006 outbreak in North America (Asadulghani *et al.*, 2009; Kulasekara *et al.*, 2009). In the case of the strains tested in chapter 4, the Stx1 prophage was observed to be the dominant inducible prophage from Australian STEC O157 strains carrying both intact Stx1 and Stx2c prophages which in turn suggested that Stx1 was the dominant toxin produced. However, the results in Chapter 4 also indicated that Stx1 prophage being the dominant prophage might not be true for all Australian STEC O157 as one strain (EC571) showed that the Stx2c prophage was dominant due to the Stx1 prophage in that genome being a defective. The carriage of Stx subtypes has previously been suggested to influence the pathogenicity of STEC O157 with subtypes Stx2a, Stx2c or Stx2d being more likely to cause the progression to HUS than Stx1(Friedrich *et al.*, 2002; Orth *et al.*, 2007). The lack of stx_{2a} detected in Australian STEC O157, coupled with the dominance of the Stx1 prophage in Australia STEC O157 could be a contributing factor to the low rates of STEC O157 clinical disease in Australia as compared to the rest of the world.

In addition to the Stx phages, there are other factors that may account for the reduced incidence of STEC O157 associated disease in Australia. One possible factor is the lack of motility in majority of the STEC O157 strains isolated from Australia, resulting in majority of the STEC O157 strains being typed as H- strains (Fegan & Desmarchelier, 2002; Mellor *et al.*, 2012). Motility is conferred by the production of the flagella. In terms of host immunological response, flagella is recognised by the Toll-like receptor 5 which will upregulate the expression of proinflammatory chemokines and neutrophils which will lead to the onset of Shiga toxin production due to the hydrogen peroxide production by neutrophils acting as an inducing agent for Stx prophages which can lead to HUS (Wagner *et al.*, 2001; Miyamoto *et al.*, 2006; Jandhyala *et al.*, 2010). Without the flagella, this immune response cascade leading to Stx production does not occur and this can account for the lower disease incidence and virulence of Australian STEC O157 strains

Another possible reason could be that the Australian STEC O157 strains, both clinical and non-clinical strains, possessing a SNP typing profile different to strains that cause significant disease (Mellor *et al.*, 2012; Mellor *et al.*, 2015). This SNP typing assay developed by Manning *et al*., (2008) subtyped STEC O157 strains into nine distinct

clades of which clades 1, 2, 3, 8 consist of outbreak strains of STEC O157. In Australia, STEC O157 strains were predominantly typed as clade 6 and clade 7 which are not typically associated with outbreak strains and a minority (2% of all strains tested) typed as clade 8 (Mellor *et al.*, 2012). This suggested that Australian STEC O157 strains may be genetically predisposed to be less virulent that their other international counterparts.

6.2. Broad host range of Stx phages leads to the potential of new Shiga toxin producing pathogens

The ability to recognise and bind to bacterial surface receptors lies with the tail fibres of phages. Tailed phages possess two tail fibres, a side tail fibre responsible for initial reversible binding onto its bacterial host and a central tail fibre responsible for the irreversible binding of the phage prior to DNA injection (Fokine & Rossmann, 2014). The genomes of the eight Stx phages analysed in this thesis all possessed a gene that encodes a tail fibre harbouring a host recognition domain in the tail fibre tip and an overlapping CDS that encodes a different tail fibre tip with a distinct host recognition domain. Genomic signatures for a +1 translational frameshift were detected in the region of overlap, which would allow the formation of a subpopulation of progeny viruses with alternative tail fibre tips. These alternative tail fibre tips were also found to be encoded in the genomes of most Stx phages sequenced to date, with the exception of three Stx phages which possess a different tail fibre with no alternate tail fibre tip, and this could be an explanation for the broad host range of the Stx phages. Previously it was thought that podoviral Stx phages only encode a tail fibre for host recognition (Plunkett *et al.*, 1999; Smith *et al.*, 2012). However, genomic analysis of the podoviral Stx phages (Chapters 3 to 5) revealed that podoviral Stx phages also possess a tail tip protein, which indicates that the podoviral Stx phages attach to their bacterial host reversibly with their tail fibres before attaching to their bacterial host irreversibly with their central tail fibre.

The genes encoding the tail fibre tips were also detected in the genomes of prophages in two other attaching and effacing bacteria, *Citrobacter rodenitum* ICC168 and *Escherichia albertii* KF1, suggesting that Stx phages expressing these tail fibres could also infect these strains (Chapter 3). Of note was the potential to infect *Escherichia*

albertii, an emerging human enteric pathogen (Nimri, 2013). *Escherichia albertii* causes gastrointestinal illness in humans and has been shown to harbour the LEE pathogenicity island (Ooka *et al.*, 2012; Nimri, 2013; Fiedoruk *et al.*, 2014). In addition, some *Escherichia albertii* strains have been observed to carry the genes encoding Stx2f (Ooka *et al.*, 2012) a toxin subtype previously identified in STEC strains isolated from pigeons (Schmidt *et al.*, 2000) as well as being associated with human STEC infections with mild clinical manifestations (Friesema *et al.*, 2014). There is, therefore, a potential that when lysogenised with a phage carrying a Stx subtype that is commonly associated with severe clinical manifestations, such as Stx2a, the resulting Stx-producing *Escherichia albertii* could be capable of causing HUS in humans.

Other than *E. coli*, *stx* genes have also been detected in both environmental and clinical isolates of *Enterobactericeae*, namely *Shigella sonnei* (Strauch *et al.*, 2001), *Citrobacter freundii* (Schmidt *et al.*, 1993; Tschape *et al.*, 1995), *Enterobacter cloacae* (Paton & Paton, 1996), *Acinetobacter haemolyticus* (Grotiuz *et al.*, 2006) and *Enterococcus* spp (Casas *et al.*, 2011). In-*vitro* studies on host range have been previously conducted on some Stx phages. A Stx2 phage isolated from *Shigella sonnei*, was shown to be able to infect *Shigella dysenteriae*, *Shigella boydii*, *Shigella. flexneri* and *E. coli* strains lacking the O antigen (Strauch *et al.*, 2001). This phage was not fully sequenced and it was unknown if it possessed tail fibres similar to that of the Stx phages investigated in this thesis. Host range studies using the Stx2a phage 933W was also attempted on a range of bacterial strains, inclusive of *E. coli* and other non *E. coli* faecal isolates (Gamage *et al.*, 2004). Lysogens could be obtained from the majority of the *E. coli* strains tested but not from the non *E. coli* faecal isolates, which included *Citrobacter freundii*, *Klebsiella pneumonia*, *Klebsiella oxytoca*, *Klebsiella ozaenae*, *Enterobacter amnigenes*, *Enterobacter aerogenes*, *Serratia marcesscens*, *Acinetobacter Iwoffii* and *Edwardsiella* spp, which suggested that 933W was not able to infect strains other than *E. coli* (Gamage *et al.*, 2004). Host range studies on the Stx2a phage, phage $24_{B,}$ were also conducted. Lysogens of phage 24_{B} could be formed from infections of *Shigella flexneri*, *Shigella sonnei ,* all tested laboratory K-12 strains and some *E. coli strains* isolated from animal and human sources (James *et al.*, 2001). Further adsorption studies have also shown that phage

24_B can attach to *Salmonella enterica* serovar Choleraesius strain SC-B67 and *C. rodentium* strain ICC168 (Smith *et al.*, 2007a).

Both 24_B and 933W also possess tail fibre genes and a tail tip protein gene similar to those of the eight Stx phages in this thesis (Chapters 3- 5). The ability of 24_B to infect a range of *Enterobactericeae* strains provided an indication that all eight Stx phages in this thesis are also likely to have a broad host range. Based on their cterm1/2 profile, it is likely that the Stx1 phages investigated in this thesis all share the same host range due to all four phages having the same cterm1/2 profile. The podoviral Stx2a phage 1810Stx2a will have an identical host range to other characterised Stx2a podoviruses while the siphoviral Stx phages will have a different host range based on the difference to the cterm1/2 profile to the podoviral Stx phages.

The bacterial surface receptors of the Stx phages have been studied to some extent with three candidates previously documented: BamA (Smith *et al.*, 2007a), FadL (Watarai *et al.*, 1998) and LamB (Watarai *et al.*, 1998). However, the results are still currently ambiguous with conflicting results despite phages possessing similar tail fibres (Islam *et al.*, 2012) and further investigations using purified receptors and tail fibre gene knockouts are required to determine the bacterial surface receptor(s) of the Stx phages. In addition, future investigation on surface receptors of Stx phages will also have to take into account the ability of some Stx phages to utilise different primary surface receptors based on the ability to encode different tail fibre tips for reversible attachment as well as a different secondary receptor recognised by the tail tip protein for irreversible attachment.

6.3. Horizontal gene transfer: the possibility of more than just *stx* and phage encoded cargo genes

Transduction mediated by Stx phages was previously mentioned twice in the literature. The first documentation of the possibility of specialised transduction mediated by Stx phages was made when prophages were induced from STEC O157 strain Sakai and amplification of bacterial DNA adjacent to both the Stx1 and Stx2a prophages was detected via DNA microarrays (Asadulghani *et al.*, 2009). The second documentation of transduction by a Stx phage was generalised transduction of random

genomic markers from one laboratory *E. coli* K-12 strain to another laboratory K-12 strain by the Stx2a phage 933W (Marinus & Poteete, 2013). Phages that mediate generalised transduction are generally headful packaging phages due to the non-specificity in DNA cleavage by the phage encoded terminase, prior to packaging in phage heads (Fineran *et al.*, 2009b). The short tailed Stx phage 933W was previously postulated to be a headful packaging phage based on sequencing data (Plunkett *et al.*, 1999) and the clustering of the amino acid sequence of the large terminase of 933W with other headful packaging phages (Casjens *et al.*, 2005). The amino acid sequence of the large terminase can be utilised to predict the DNA packaging mechanism (Casjens *et al.*, 2005) and comparative genomics performed in this thesis showed that all previously sequenced short tailed Stx phages possessed a large terminase that is similar to the large terminase of 933W and are likely to all be headful packaging phages (Chapters 3- 5).

It was determined in Chapter 5 that the Stx2 prophage, pp1810Stx2a, was key in the mediation of transduction of an antibiotic resistance marker within the LEE (Δ*eae*::*cat*), from STEC O157 strains carrying the marker to other STEC O157 strains. The physical separation of the LEE, encoded over 1 Mb away from pp1810Stx2a in the EC1810 genome, and the ability of pp1810Stx2 to mobilise Δ*eae*::*cat*, suggested that transduction mediated by pp1810Stx2 was possibly via a novel transduction mechanism. This proposed transduction event, whereby specific, non-adjacent bacterial genomic DNA is mobilised in-*trans,* is the first description of an endogenous prophage within an attaching and effacing bacterial strain that can mobilise the LEE. The Stx2a prophage pp1810Stx2a also possessed a replication and regulatory system that was highly similar to that of the more recently characterised Stx2a phages, P13374, phi-191 and TL2011C isolated from STEC isolates O104:H4 strain CB13374 (Beutin *et al.*, 2012), O111:H2 strain ED 191(Grande *et al.*, 2014) and O103:H25 strain NVH-734 (L'abee-Lund *et al.*, 2012), respectively, of which only strain NVH-734 possessed the LEE pathogenicity island. Based on the similarity of the replication and regulatory system of TL2011C to pp1810Stx2, it might be possible that TL2011C could also participate in LEE mobilisation and it would be interesting to analyse the phage metagenome of strain NVH-734 to determine if the LEE could also be mobilised in this strain.

6.4. MITES as a possible cargo gene acquisition mechanism

While the carriage of a *stx* operon is the defining characteristic of Stx phages, it is important to appreciate that *stx* genes are cargo genes that are not involved in the life cycle of the phage. It has been previously speculated that cargo genes could have inserted site specifically into phage genomes via a yet unknown mechanism (Nilsson *et al.*, 2004) and in this thesis, a candidate mechanism has been postulated upon.

The transfer of the *stx* operon and its flanking genomic region from the Stx1 prophage into the Stx2a prophage was previously documented in STEC O157 strain Sakai, and it was speculated to have occurred by homologous recombination due to the high levels of nucleotide similarity between the flanking regions of *stx* operons in both Stx1 and Stx2a phages (Asadulghani *et al.*, 2009; Sváb *et al.*, 2015). However, comparative genomic analysis performed in this thesis showed that there are two non Stx prophages in bacterial genomes namely P04 of *E. coli* O111:H- strain 11128 and a prophage integrated adjacent to *ynfG* in the genomes of *E. coli* O104:H4 strains 2011C-3493, 2009EL-2050 and 2009EL-2071 that are genomically highly similar to the podoviral Stx phages inclusive of the region flanking where the *stx* operon is located (Chapter 3). In addition, the comparisons of Stx phage genomes revealed a high level of sequence similarity in the genes flanking the *stx* operon but a clear homology break point observed at the *stx* operon of Stx phages carrying different *stx* genes (Chapter 3). Near the edge of this breakpoint are the detection of imperfect inverted repeats, highly similar to the terminal inverted repeats of IS*102*, flanking the *stx* operon, suggesting that the *stx* operon could have previously entered the phage as part of a MITE and that this could be a potential mechanism by which cargo genes, and not limited to *stx* genes, initially entered a phage genome.

Apart from the *stx* operon, other cargo genes that have been associated with Stx phages are homologs of genes that encode an acetylesterase which mediates colonisation of its host bacterium to intestinal epithelial cells, non LEE encoded Type III secretion system effectors (Creuzburg *et al.*, 2005b; Strauch *et al.*, 2008) as well as *stk*, *lom* and *bor*, which encode a serine/threonine kinase, an outer membrane protein associated with binding to epithelial cells and a protein that confers resistance

to host serum, respectively (Plunkett *et al.*, 1999; Tyler & Friedman, 2004; Strauch *et al.*, 2008; Smith *et al.*, 2012). Not all sequenced Stx phages carry all of the aforementioned cargo genes within their genomes with the cargo genes *bor* and *stk* typically found in the genomes of the podoviral Stx phages (Plunkett *et al.*, 1999; Smith *et al.*, 2012) while the non-LEE effectors are typically found in the genomes of with the siphoviral Stx phages (Strauch *et al.*, 2008). With the exception of the myoviral Stx2e phage and Stx1 phage CP1639, all sequenced Stx phages carry the gene that encodes the acetylesterase (Chapter 3-5). In addition to these cargo genes, genomic characterisations performed in the thesis (Chapter 4-5) identified four additional cargo genes, harboured by Stx phages that can increase the fitness of its bacterium. Two of these are cargo genes that are involved in substrate metabolism while the other two are genes that provide the phage with a defence against host toxin-antitoxin induced abortive infection. Prophages can increase the fitness of the bacterial host due to lysogenic conversion (Brussow & Hendrix, 2002) and the possibility of cargo genes being part of a MITE, though experimentally untested, provides a new candidate mechanism to account for the acquisition of cargo genes in phage genomes.

6.5. Conclusions

The work presented in this thesis has provided insights into Stx phage genome-encoded factors that could enhance the pathogenicity of STEC O157 strains. Toxin production is likely to be influenced by means of phage borne anti-repressors alleviating repression of Stx prophages either in-*cis* or in-*trans* leading to toxin production. Importantly, this thesis also identified the potential for production of alternative tail fibre tips, with different host recognition domains, and found it is a conserved mechanism in all Stx phages and is a likely a reason for their broad host range. This is significant as there is currently no mechanism to account for the ability for the Stx phage to infect across a broad host range of *Enterobactericeae* strains, which could lead to the emergence of new Stx producing pathogens.

Prior to this study, there was no mention of a Stx phage transduction, only a single mention of a Stx phage mediating the transduction of genomic markers in a flawed study published recently. The results presented in this thesis, building from previous

phage metagenome analysis, demonstrated for the first time, the transduction of a gene encoded within the LEE to a different STEC O157 strain by an a endogenous Stx phage, and thus the potential for the entire pathogenicity island to be transduced. This work has significant implications in understanding the acquisition of the LEE pathogenicity island in attaching and effacing bacterial pathogens.

Overall, the work presented in this thesis added to the current understanding of Stx phages. Results suggested that Stx phages contributed more to bacterial pathogenesis than just the conference of the *stx* operon, with phage encoded genes being implicated to regulate levels of toxin production and the emergence of new Stx producing pathogens, as well as the possibility of Stx phages to mediate transduction, provided new insights into how Stx phages can influence the evolution and pathogenicity of STEC O157, and indeed other bacteria.

References

Abbani, M. A., C. V. Papagiannis, M. D. Sam, D. Cascio, R. C. Johnson & R. T. Clubb, (2007) Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc Natl Acad Sci USA* **104**: 2109-2114

Abuladze, N. K., M. Gingery, J. Tsai & F. A. Eiserling, (1994) Tail length determination in bacteriophage T4. *Virology* **199**: 301-310

Ackermann, H. W., (1998) Tailed bacteriophages: the order *caudovirales*. *Adv Virus Res* **51**: 135-201

Ackermann, H. W., (2009) Phage classification and characterization. *Methods Mol Biol* **501**: 127-140

Ackermann, H. W., A. Audurier, L. Berthiaume, L. A. Jones, J. A. Mayo & A. K. Vidaver, (1978) Guidelines for bacteriophage characterization. *Adv Virus Res* **23**: 1-24

Ackermann, H. W. & D. Prangishvili, (2012) Prokaryote viruses studied by electron microscopy. *Arch Virol* **157**: 1843-1849

Aertsen, A., D. Faster & C. W. Michiels, (2005) Induction of Shiga toxin-converting prophage in *Escherichia coli* by high hydrostatic pressure. *Appl Environ Microbiol* **71**: 1155-1162

Ahmed, S. A., J. Awosika, C. Baldwin, K. A. Bishop-Lilly, B. Biswas, S. Broomall, P. S. Chain, O. Chertkov, O. Chokoshvili, S. Coyne, K. Davenport, J. C. Detter, W. Dorman, T. H. Erkkila, J. P. Folster, K. G. Frey, M. George, C. Gleasner, M. Henry, K. K. Hill, K. Hubbard, J. Insalaco, S. Johnson, A. Kitzmiller, M. Krepps, C. C. Lo, T. Luu, L. A. McNew, T. Minogue, C. A. Munk, B. Osborne, M. Patel, K. G. Reitenga, C. N. Rosenzweig, A. Shea, X. Shen, N. Strockbine, C. Tarr, H. Teshima, E. van Gieson, K. Verratti, M. Wolcott, G. Xie, S. Sozhamannan, H. S. Gibbons & T. C. Consortium, (2012) Genomic comparison of *Escherichia coli* O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including Shiga toxin encoding phage stx2. *PLoS One* **7**: e48228.doi:10.1371/journal.pone.0048228

Aksyuk, A. A., P. G. Leiman, L. P. Kurochkina, M. M. Shneider, V. A. Kostyuchenko, V. V. Mesyanzhinov & M. G. Rossmann, (2009) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J* **28**: 821-829

Aksyuk, A. A. & M. G. Rossmann, (2011) Bacteriophage assembly. *Viruses* **3**: 172-203

Alikhan, N.-F., N. K. Petty, N. L. Ben Zakour & S. A. Beatson, (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**.doi:10.1186/1471-2164-12-402

Allison, H. E., (2007) Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiol* **2**: 165-174

Allison, H. E., M. J. Sergeant, C. E. James, J. R. Saunders, D. L. Smith, R. J. Sharp, T. S. Marks & A. J. McCarthy, (2003) Immunity profiles of wild-type and recombinant Shiga-like toxin-encoding bacteriophages and characterisation of novel double lysogens. *Infect Immun* **71**: 3409-3418

Asadulghani, M., Y. Ogura, T. Ooka, T. Itoh, A. Sawaguchi, A. Iguchi, K. Nakayama & T. Hayashi, (2009) The defective prophage pool of *Escherichia coli* O157:

prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog* **5**: e1000408.doi: 10.1371/journal.ppat.1000408

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke & O. Zagnitko, (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.doi:10.1186/1471-2164-9-75

Backhaus, H. & J. B. Petri, (1984) Sequence analysis of a region from the early right operon in phage P22 including the replication genes 18 and 12. *Gene* **32**: 289-303

Barondess, J. J. & J. Beckwith, (1995) *bor* gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein. *J Bacteriol* **177**: 1247-1253

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero & P. Horvath, (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712

Battaglioli, E. J., G. A. Baisa, A. E. Weeks, R. A. Schroll, A. J. Hryckowian & R. A. Welch, (2011) Isolation of generalized transducing bacteriophages for uropathogenic strains of *Escherichia coli*. *Appl Environ Microbiol* **77**: 6630-6635

Bentancor, A. B., L. A. Ameal, M. F. Calvino, M. C. Martinez, L. Miccio & O. J. Degregorio, (2012) Risk factors for Shiga toxin-producing *Escherichia coli* infections in preadolescent schoolchildren in Buenos Aires, Argentina. *J Infect Dev Ctries* **6**: 378-386

Bergan, J., A. B. D. Lingelem, R. Simm, T. Skotland & K. Sandvig, (2012) Shiga toxins. *Toxicon* **60**: 1085-1107

Bergh, O., K. Y. Borsheim, G. Bratbak & M. Heldal, (1989) High abundance of viruses found in aquatic environments. *Nature* **340**: 467-468

Besser, R. E., S. M. Lett, J. T. Weber, M. P. Doyle, T. J. Barrett, J. G. Wells & P. M. Griffin, (1993) An outbreak of diarrhea and hemolytic uremic syndrome from *Escherichia coli* O157:H7 in fresh-pressed apple cider. *JAMA* **269**: 2217-2220

Beutin, L., J. A. Hammerl, E. Strauch, J. Reetz, R. Dieckmann, Y. Kelner-Burgos, A. Martin, A. Miko, N. A. Strockbine, B. A. Lindstedt, D. Horn, H. Monse, B. Huettel, I. Muller, K. Stuber & R. Reinhardt, (2012) Spread of a distinct Stx2-encoding phage prototype among *Escherichia coli* O104:H4 strains from outbreaks in Germany, Norway, and Georgia. *J Virol* **86**: 10444-10455

Bickle, T. A. & D. H. Kruger, (1993) Biology of DNA restriction. *Microbiol Rev* **57**: 434-450

Billard-Pomares, T., S. Fouteau, M. E. Jacquet, D. Roche, V. Barbe, M. Castellanos, J. Y. Bouet, S. Cruveiller, C. Medigue, J. Blanco, O. Clermont, E. Denamur & C. Branger, (2014) Characterization of a P1-like bacteriophage carrying an SHV-2 extended-spectrum beta-lactamase from an *Escherichia coli* strain. *Antimicrob Agents Chemother* **58**: 6550-6557

Biswas, T., H. Aihara, M. Radman-Livaja, D. Filman, A. Landy & T. Ellenberger, (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* **435**: 1059-1066.10.1038/nature03657

Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau & Y. Shao,

(1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1462

Blower, T. R., T. J. Evans, R. Przybilski, P. C. Fineran & G. P. Salmond, (2012) Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. *PLoS Genet* **8**: e1003023.doi:10.1371/journal.pgen.1003023

Bock, A., K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek & F. Zinoni, (1991) Selenocysteine: the 21st amino acid. *Mol Microbiol* **5**: 515-520

Bohm, H. & H. Karch, (1992) DNA fingerprinting of *Escherichia coli* O157:H7 strains by pulsed-field gel electrophoresis. *J Clin Microbiol* **30**: 2169-2172

Bondy-Denomy, J., A. Pawluk, K. L. Maxwell & A. R. Davidson, (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429-432

Bonhivers, M. & L. Letellier, (1995) Calcium controls phage T5 infection at the level of the *Escherichia coli* cytoplasmic membrane. *FEBS Lett* **374**: 169-173

Bono, J. L., T. P. Smith, J. E. Keen, G. P. Harhay, T. G. McDaneld, R. E. Mandrell, W. K. Jung, T. E. Besser, P. Gerner-Smidt, M. Bielaszewska, H. Karch & M. L. Clawson, (2012) Phylogeny of Shiga toxin-producing *Escherichia coli* O157 isolated from cattle and clinically ill humans. *Mol Biol Evol* **29**: 2047-2062

Botstein, D., (1980) A theory of modular evolution for bacteriophages. *Ann N Y Acad Sci* **354**: 484-490

Bougdour, A., C. Cunning, P. J. Baptiste, T. Elliott & S. Gottesman, (2008) Multiple pathways for regulation of sigmaS (RpoS) stability in *Escherichia coli* via the action of multiple anti-adaptors. *Mol Microbiol* **68**: 298-313

Bradley, M. D., M. B. Beach, A. P. de Koning, T. S. Pratt & R. Osuna, (2007) Effects of Fis on *Escherichia coli* gene expression during different growth stages. *Microbiology* **153**: 2922-2940

Brennan, R. G. & B. W. Matthews, (1989) The helix-turn-helix DNA binding motif. *J Biol Chem* **264**: 1903-1906

Brunder, W., H. Schmidt & H. Karch, (1996) KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* **142** 3305-3315

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

Brussow, H., C. Canchaya & W. D. Hardt, (2004) Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**: 560-602

Brussow, H. & R. W. Hendrix, (2002) Phage genomics: small is beautiful. *Cell* **108**: 13-16

Budzik, J. M., W. A. Rosche, A. Rietsch & G. A. O'Toole, (2004) Isolation and characterization of a generalized transducing phage for *Pseudomonas aeruginosa* strains PAO1 and PA14. *J Bacteriol* **186**: 3270-3273

Burland, V., Y. Shao, N. T. Perna, G. Plunkett, H. J. Sofia & F. R. Blattner, (1998) The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Research* **26**: 4196-4204

Byrne, L., C. Jenkins, N. Launders, R. Elson & G. K. Adak, (2015) The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. *Epidemiol Infect* **143**: 3475-3487

Cabezon, E., J. Ripoll-Rozada, A. Pena, F. de la Cruz & I. Arechaga, (2015) Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev* **39**: 81-95

Calendar, R., E. Ljungquist, G. Deho, D. C. Usher, R. Goldstein, P. Youderian, G. Sironi & E. W. Six, (1981) Lysogenization by satellite phage P4. *Virology* **113**: 20-38

Campbell, A., (1994) Comparative molecular biology of lambdoid phages. *Annu Rev Microbiol.* **48**: 193-222

Campbell, A., (2003a) The future of bacteriophage biology. *Nat Rev Genet* **4**: 471-477

Campbell, A., (2003b) Prophage insertion sites. *Res Microbiol.* **154**: 277-282

Campbell, A. M., (1992) Chromosomal insertion sites for phages and plasmids. *J Bacteriol* **174**: 7495-7499

Canchaya, C., G. Fournous & H. Brussow, (2004) The impact of prophages on bacterial chromosomes. *Mol Microbiol* **53**: 9-18

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

Carver, T., S. R. Harris, T. D. Otto, M. Berriman, J. Parkhill & J. A. McQuillan, (2013) BamView: visualizing and interpretation of next-generation sequencing read alignments. *Brief Bioinform* **14**: 203-212

Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell & J. Parkhill, (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422-3423

Casas, V., G. Sobrepena, B. Rodriguez-Mueller, J. AhTye & S. R. Maloy, (2011) Bacteriophage-encoded Shiga toxin gene in atypical bacterial host. *Gut Pathog* **3**.doi: 10.1186/1757-4749-3-10

Casjens, S., (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* **49**: 277-300

Casjens, S. & M. Hayden, (1988) Analysis in vivo of the bacteriophage P22 headful nuclease. *J Mol Biol* **199**: 467-474

Casjens, S. & I. Molineux, (2012) Short Noncontractile Tail Machines: Adsorption and DNA Delivery by Podoviruses. In: Viral Molecular Machines. M. G. Rossmann & V. B. Rao (eds). Springer US, pp. 143-179.

Casjens, S. R., (2005) Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* **8**: 451-458

Casjens, S. R. & E. B. Gilcrease, (2008) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. In: Bacteriophages methods and protocols voume 2: Molecular and applied aspects. A. M. Kropinski & M. R. J. Clokie (eds). Humana Press, pp. 91-111.

Casjens, S. R., E. B. Gilcrease, D. A. Winn-Stapley, P. Schicklmaier, H. Schmieger, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull & R. W. Hendrix, (2005) The generalized transducing *Salmonella* bacteriophage ES18: Complete genome sequence and DNA packaging strategy. *J Bacteriol* **187**: 1091-1104

Castillo, A., L. E. Eguiarte & V. Souza, (2005) A genomic population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli***:** the search for the unit of selection. *Proc Natl Acad Sci U S A* **102**: 1542-1547

Caugant, D. A., B. R. Levin, I. Ørskov, F. Ørskov, C. Svanborg Eden & R. K. Selander, (1985) Genetic diversity in relation to serotype in *Escherichia coli*. *Infect Immun* **49**: 407-413

Chan, P. P. & T. M. Lowe, (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37**: D93-D97 doi:10.1093/nar/gkn1787.10.1093/nar/gkn787

Chandry, P. S., S. Gladman, S. C. Moore, T. Seemann, K. A. Crandall & N. Fegan, (2012) A Genomic Island in *Salmonella enterica* ssp. *salamae* provides new insights on the genealogy of the Locus of Enterocyte Effacement. *PLoS ONE* **7**: e41615.doi:10.1371/journal.pone.0041615

Chartron, J., C. Shiau, C. D. Stout & K. S. Carroll, (2007) 3'-Phosphoadenosine-5'-phosphosulfate reductase in complex with thioredoxin: a structural snapshot in the catalytic cycle. *Biochemistry* **46**: 3942-3951

Chin, C.-S., D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E. E. Eichler, S. W. Turner & J. Korlach, (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Meth* **10**: 563-569

Christensen-Dalsgaard, M., M. G. Jorgensen & K. Gerdes, (2010) Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol Microbiol* **75**: 333-348

Christie, G. E. & R. Calendar, (1990) Interactions between satellite bacteriophage P4 and its helpers. *Annu. Rev. Genet.* **24**: 465-490

Christie, G. E. & T. Dokland, (2012) Pirates of the *Caudovirales*. *Virology* **434**: 210-221

Christie, G. E., L. M. Temple, B. A. Bartlett & T. S. Goodwin, (2002) Programmed translational frameshift in the bacteriophage P2 *FETUD* tail gene operon. *J Bacteriol* **184**: 6522-6531

Conradi, H., (1903) Uber Iosliche, durch asptische Autolyse erhaltene Giftstoffe vonRuhr- und Typhus-Bazillen. *Dtsch Med Wochenschr* **29**: 26-28

Cooper, K. K., R. E. Mandrell, J. W. Louie, J. Korlach, T. A. Clark, C. T. Parker, S. Huynh, P. S. Chain, S. Ahmed & M. Q. Carter, (2014a) Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7. *BMC Genomics* **15**: 17.doi:10.1186/1471-2164-15-17

Cooper, K. K., R. E. Mandrell, J. W. Louie, J. Korlach, T. A. Clark, C. T. Parker, S. Huynh, P. S. G. Chain, S. Ahmed & M. Q. Carter, (2014b) Complete genome sequences of two *Escherichia coli* O145:H28 outbreak strains of food origin. *Genome Announce* **2**: e00482-00414.doi:10.1128/genomeA.00482-14

Cornelissen, A., P. J. Ceyssens, V. N. Krylov, J. P. Noben, G. Volckaert & R. Lavigne, (2012) Identification of EPS-degrading activity within the tail spikes of the novel *Pseudomonas putida* phage AF. *Virology* **434**: 251-256

Cote, R., R. Katani, M. R. Moreau, I. T. Kudva, T. M. Arthur, C. DebRoy, M. M. Mwangi, I. Albert, J. A. Raygoza Garay, L. Li, M. T. Brandl, M. Q. Carter & V. Kapur, (2015) Comparative analysis of super-shedder strains of *Escherichia coli* O157:H7 reveals distinctive genomic features and a strongly aggregative adherent phenotype on bovine rectoanal junction squamous epithelial cells. *PLoS One* **10**: e0116743.doi:10.1371/journal.pone.0116743

Cowden, J. M., S. Ahmed, M. Donaghy & A. Riley, (2001) Epidemiological investigation of the central Scotland outbreak of *Escherichia coli* O157 infection, November to December 1996. *Epidemiol Infect* **126**: 335-341

Creuzburg, K., B. Kohler, H. Hempel, P. Schreier, E. Jacobs & H. Schmidt, (2005a) Genetic structure and chromosomal integration site of the cryptic prophage CP-1639 encoding Shiga toxin 1. *Microbiology* **151**: 941-950

Creuzburg, K., E. Recktenwald, V. Kuhle, S. Herold, M. Hensel & H. Schmidt, (2005b) The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-Like type III effector protein. *J Bacteriol* **187**: 8494-8498

Crump, J. A., C. R. Braden, M. E. Dey, R. M. Hoekstra, J. M. Rickelman-Apisa, D. A. Baldwin, S. J. De Fijter, S. F. Nowicki, E. M. Koch, T. L. Bannerman, F. W. Smith, J. P. Sarisky, N. Hochberg & P. S. Mead, (2003) Outbreaks of *Escherichia coli* O157 infections at multiple county agricultural fairs: a hazard of mixing cattle, concession stands and children. *Epidemiol Infect* **131**: 1055-1062

Crump, J. A., A. C. Sulka, A. J. Langer, C. Schaben, A. S. Crielly, R. Gage, M. Baysinger, M. Moll, G. Withers, D. M. Toney, S. B. Hunter, R. M. Hoekstra, S. K. Wong, P. M. Griffin & T. J. Van Gilder, (2002) An outbreak of *Escherichia coli* O157:H7 infections among visitors to a dairy farm. *N Engl J Med* **347**: 555-560

d'Herelle, F., (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad Sci Paris*: 373-375

Dai, W., A. Hodes, W. H. Hui, M. Gingery, J. F. Miller & Z. H. Zhou, (2010) Three-dimensional structure of tropism-switching Bordetella bacteriophage. *Proc Natl Acad Sci USA* **107**: 4347-4352

Dalton, K. M. & S. Crosson, (2010) A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. *Biochemistry* **49**: 2205-2215

Darling, A. C., B. Mau, F. R. Blattner & N. T. Perna, (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394-1403

Datta, A. B., S. Panjikar, M. S. Weiss, P. Chakrabarti & P. Parrack, (2005) Structure of lambda CII: implications for recognition of direct-repeat DNA by an unusual tetrameric organization. *Proc Natl Acad Sci U S A* **102**: 11242-11247

Davidson, A. R., L. Cardarelli, L. G. Pell, D. R. Radford & K. L. Maxwell, (2012) Long noncontractile tail machines of bacteriophages. *Adv Exp Med Biol* **726**: 115-142

De Paepe, M., G. Hutinet, O. Son, J. Amarir-Bouhram, S. Schbath & M. A. Petit, (2014) Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet* **10**: e1004181.doi:10.1371/journal.pgen.1004181

De Palmenaer, D., C. Vermeiren & J. Mahillon, (2004) IS*231*-MIC*231* elements from *Bacillus cereus* sensu lato are modular. *Mol. Microbiol* **53**: 457-467

Deho, G., D. Ghisotti, P. Alano, S. Zangrossi, M. G. Borrello & G. Sironi, (1984) Plasmid mode of propagation of the genetic element P4. *J Mol Biol* **178**: 191-207

Delihas, N., (2011) Impact of small repeat sequences on bacterial genome evolution. *Genome Biol Evol* **3**: 959-973

Deng, W., Y. Li, B. A. Vallance & B. B. Finlay, (2001) Locus of Enterocyte Effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. *Infect Immun* **69**: 6323-6335

Deng, W., J. L. Puente, S. Gruenheid, Y. Li, B. A. Vallance, A. Vázquez, J. Barba, J. A. Ibarra, P. O'Donnell, P. Metalnikov, K. Ashman, S. Lee, D. Goode, T. Pawson & B. B. Finlay, (2004) Dissecting virulence: Systematic and functional analyses of a pathogenicity island. *Proc Natl Acad Sci USA* **101**: 3597-3602

Desmarchelier, P. M., S. S. Bilge, N. Fegan, L. Mills, J. C. Vary, Jr. & P. I. Tarr, (1998) A PCR specific for *Escherichia coli* O157 based on the *rfb* locus encoding O157 lipopolysaccharide. *J Clin Microbiol* **36**: 1801-1804

Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath & S. Moineau, (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**: 1390-1400

Dillingham, M. S. & S. C. Kowalczykowski, (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* **72**: 642-671

Dodd, I. B. & J. B. Egan, (1996) DNA binding by the coliphage 186 repressor protein CI. *J Biol Chem* **271**: 11532-11540

Dokland, T., B. H. Lindqvist & S. D. Fuller, (1992) Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2-P4 bacteriophage system. *EMBO J* **11**: 839-846

Drozdz, M., A. Piekarowicz, J. M. Bujnicki & M. Radlinska, (2012) Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Res* **40**: 2119-2130

Ebel-Tsipis, J., M. S. Fox & D. Botstein, (1972) Generalized transduction by bacteriophage P22 in *Salmonella typhimurium*. II. Mechanism of integration of transducing DNA. *J Mol Biol* **71**: 449-469

Echols, H. & L. Green, (1971) Establishment and maintenance of repression by bacteriophage lambda: the role of the CI, CII, and CIII proteins. *Proc Natl Acad Sci U S A* **68**: 2190-2194

Elliott, S. J., L. A. Wainwright, T. K. McDaniel, K. G. Jarvis, Y. K. Deng, L. C. Lai, B. P. McNamara, M. S. Donnenberg & J. B. Kaper, (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic Escherichia coli E2348/69. *Mol Microbiol* **28**: 1-4

Enquist, L. W. & A. Skalka, (1973) Replication of bacteriophage λ DNA dependent on the function of host and viral genes: I. Interaction of *red, gam* and *rec*. *J Mol Biol* **75**: 185-212

Eppinger, M., M. K. Mammel, J. E. Leclerc, J. Ravel & T. A. Cebula, (2011) Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc Natl Acad Sci USA* **108**: 20142-20147

Escherich, T., (1885) Die Darmbakterien des neugeborenen und säuglings. *Fortschr Med* **3**: 547-554

Farabaugh, P. J., (1996) Programmed translational frameshifting. *Microbiol Rev* **30**: 103-134

Fegan, N. & P. Desmarchelier, (2002) Comparison between human and animal isolates of Shiga toxin-producing *Escherichia coli* O157 from Australia. *Epidemiol Infect* **128**: 357-362

Feiss, M., R. A. Fisher, M. A. Crayton & C. Egner, (1977) Packaging of the bacteriophage lambda chromosome: effect of chromosome length. *Virology* **77**: 281-293

Feng, P., (1995) *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants. *Emerg Infect Dis* **1**: 47-52

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

Ferens, W. A. & C. J. Hovde, (2011) *Escherichia coli* O157:H7: Animal Reservoir and Sources of Human Infection. *Foodborne Pathog Dis* **8**: 465-487

Fiedoruk, K., T. Daniluk, I. Swiecicka, E. Murawska, M. Sciepuk & K. Leszczynska, (2014) First complete genome sequence of *Escherichia albertii* strain KF1, a

new potential human enteric pathogen. *Genome Announc* **2**.doi:10.1128/genomeA.00004-14

Fineran, P. C., T. R. Blower, I. J. Foulds, D. P. Humphreys, K. S. Lilley & G. P. Salmond, (2009a) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* **106**: 894-899

Fineran, P. C., N. K. Petty & G. P. C. Salmond, (2009b) Transduction: Host DNA Transfer by Bacteriophages. In: Encyclopedia of microbiology. M. Schaechter (ed). Oxford: Elsevier, pp. 666-679.

Fitzmaurice, W. P., A. S. Waldman, R. C. Benjamin, P. C. Huang & J. J. Scocca, (1984) Nucleotide sequence and properties of the cohesive DNA termini from bacteriophage HP1c1 of *Haemophilus influenzae* Rd. *Gene* **31**: 197-203

Fogg, P. C., S. M. Gossage, D. L. Smith, J. R. Saunders, A. J. McCarthy & H. E. Allison, (2007) Identification of multiple integration sites for Stx-phage Phi24B in the *Escherichia coli* genome, description of a novel integrase and evidence for a functional anti-repressor. *Microbiology* **153**: 4098-4110

Fokine, A. & M. G. Rossmann, (2014) Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**: e28281. doi:28210.24161/bact.28281.10.4161/bact.28281

Frankel, G. & A. D. Phillips, (2008) Attaching effacing *Escherichia coli* and paradigms of Tir-triggered actin polymerization: getting off the pedestal. *Cell Microbiol* **10**: 549-556

Franz, E., P. Delaquis, S. Morabito, L. Beutin, K. Gobius, D. A. Rasko, J. Bono, N. French, J. Osek, B. A. Lindstedt, M. Muniesa, S. Manning, J. LeJeune, T. Callaway, S. Beatson, M. Eppinger, T. Dallman, K. J. Forbes, H. Aarts, D. L. Pearl, V. P. Gannon, C. R. Laing & N. J. Strachan, (2014) Exploiting the explosion of information associated with whole genome sequencing to tackle Shiga toxin-producing *Escherichia coli* (STEC) in global food production systems. *Int J Food Microbiol* **187**: 57-72

Franz, E., A. H. van Hoek, F. J. van der Wal, A. de Boer, A. Zwartkruis-Nahuis, K. van der Zwaluw, H. J. Aarts & A. E. Heuvelink, (2012) Genetic features differentiating bovine, food, and human isolates of shiga toxin-producing *Escherichia coli* O157 in The Netherlands. *J Clin Microbiol* **50**: 772-780

Fraser, J. S., Z. Yu, K. L. Maxwell & A. R. Davidson, (2006) Ig-like domains on bacteriophages: A tale of promiscuity and deceit. *J Mol Biol* **359**: 496-507

Freeman, N. L., D. V. Zurawski, P. Chowrashi, J. C. Ayoob, L. Huang, B. Mittal, J. M. Sanger & J. W. Sanger, (2000) Interaction of the enteropathogenic *Escherichia coli* protein, translocated intimin receptor (Tir), with focal adhesion proteins. *Cell Motil Cytoskeleton* **47**: 307-318

Freifelder, D. & I. Kirschner, (1971) The formation of homoimmune double lysogens of phage lambda and the segregation of single lysogens from them. *Virology* **44**: 633-637

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

Friesema, I., K. van der Zwaluw, T. Schuurman, M. Kooistra-Smid, E. Franz, Y. van Duynhoven & W. van Pelt, (2014) Emergence of *Escherichia coli* encoding Shiga toxin 2f in human Shiga toxin-producing *E. coli* (STEC) infections in the Netherlands, January 2008 to December 2011. *Euro Surveill* **19**: 26-32

Fuchs, S., I. Muhldorfer, A. Donohue-Rolfe, M. Kerenyi, L. Emody, R. Alexiev, P. Nenkov & J. Hacker, (1999) Influence of RecA on in vivo virulence and Shiga toxin 2 production in *Escherichia coli* pathogens. *Microb Pathog* **27**: 13-23

Fuller, C. A., C. A. Pellino, M. J. Flagler, J. E. Strasser & A. A. Weiss, (2011) Shiga toxin subtypes display dramatic differences in potency. *Infect Immun* **79**: 1329-1337

Furth, M. E. & S. H. Wickner, (1983) Lambda DNA replication. *Cold Spring Harbor Monograph Archive* **13**: 145-173

Gamage, S. D., A. K. Patton, J. F. Hanson & A. A. Weiss, (2004) Diversity and host range of Shiga toxin-encoding phage. *Infect Immun* **72**: 7131-7139

Garcia-Aljaro, C., M. Muniesa, J. Jofre & A. R. Blanch, (2009) Genotypic and phenotypic diversity among induced, stx(2)-carrying bacteriophages from environmental *Escherichia coli* strains. *Appl Environ Microbiol* **75**: 329-336

Garred, O., B. van Deurs & K. Sandvig, (1995) Furin-induced cleavage and activation of Shiga toxin. *J Biol Chem* **270**: 10817-10821

Ghosh, N., T. J. McKillop, T. A. Jowitt, M. Howard, H. Davies, D. F. Holmes, I. S. Roberts & J. Bella, (2012) Collagen-Like Proteins in Pathogenic *E. coli* Strains. *PLoS ONE* **7**: e37872.doi:10.1371/journal.pone.0037872

Gobert, A. P., M. Vareille, A. L. Glasser, T. Hindre, T. de Sablet & C. Martin, (2007) Shiga toxin produced by enterohemorrhagic *Escherichia coli* inhibits PI3K/NF-kappaB signaling pathway in globotriaosylceramide-3-negative human intestinal epithelial cells. *J Immunol* **178**: 8168-8174

Gomez-Alvarez, V., T. K. Teal & T. M. Schmidt, (2009) Systematic artifacts in metagenomes from complex microbial communities. *ISME J* **3**: 1314-1317

Gould, L. H., L. Demma, T. F. Jones, S. Hurd, D. J. Vugia, K. Smith, B. Shiferaw, S. Segler, A. Palmer, S. Zansky, P. M. Griffin & t. E. I. P. F. W. Group, (2009) Hemolytic uremic syndrome and death in persons with *Escherichia coli* O157:H7 infection, foodborne diseases active surveillance network sites, 2000–2006. *Clin Infect Dis* **49**: 1480-1485

Grande, L., V. Michelacci, R. Tozzoli, P. Ranieri, A. Maugliani, A. Caprioli & S. Morabito, (2014) Whole genome sequence comparison of vtx2-converting phages from Enteroaggregative Haemorrhagic *Escherichia coli* strains. *BMC Genomics* **15**: 574.doi: 10.1186/1471-2164-15-574

Grant, S. G., J. Jessee, F. R. Bloom & D. Hanahan, (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci USA* **87**: 4645-4649

Grosschedl, R. & E. Schwarz, (1979) Nucleotide sequence of the *cro-cII*-oop region of bacteriophage 434 DNA. *Nucleic Acids Res* **6**: 867-881

Groth, A. C. & M. P. Calos, (2004) Phage integrases: biology and applications. *J Mol Biol* **335**: 667-678

Grotiuz, G., A. Sirok, P. Gadea, G. Varela & F. Schelotto, (2006) Shiga toxin 2-producing *Acinetobacter haemolyticus* associated with a case of bloody diarrhea. *J Clin Microbiol* **44**: 3838-3841

Guh, A., Q. Phan, R. Nelson, K. Purviance, E. Milardo, S. Kinney, P. Mshar, W. Kasacek & M. Cartter, (2010) Outbreak of *Escherichia coli* O157 associated with raw milk, Connecticut, 2008. *Clin Infect Dis* **51**: 1411-1417

Hamano, S.-i., Y. Nakanishi, T. Nara, T. Seki, T. Ohtani, T. Oishi, K. Joh, T. Oikawa, Y. Muramatsu, Y. Ogawa & S. Akashi, (1993) Neurological manifestations of hemorrhagic colitis in the outbreak of *Escherichia coli* O157:H7 infection in Japan. *Acta Pædiatrica* **82**: 454-458

Hammad, A. M. M., (1999) Evaluation of alginate-encapsulated *Azotobacter chroococcum* as a phage-resistant and an effective inoculum. *J Basic Microbiol* **38**: 9-16

Hatfull, G. F., (2008) Bacteriophage genomics. *Curr Opin Microbiol* **11**: 447-453

Hatfull, G. F., M. L. Pedulla, D. Jacobs-Sera, P. M. Cichon, A. Foley, M. E. Ford, R. M. Gonda, J. M. Houtz, A. J. Hryckowian, V. A. Kelchner, S. Namburi, K. V. Pajcini, M. G. Popovich, D. T. Schleicher, B. Z. Simanek, A. L. Smith, G. M. Zdanowicz, V. Kumar, C. L. Peebles, W. R. Jacobs, Jr., J. G. Lawrence & R. W. Hendrix, (2006) Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet* **2**: e92.doi:10.1371/journal.pgen.0020092

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

Hazan, R. & H. Engelberg-Kulka, (2004) *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol Genet Genomics* **272**: 227-234

Hendrix, R. W. & R. L. Duda, (1992) Bacteriophage lambda PaPa: not the mother of all lambda phages. *Science* **258**: 1145-1148

Hendrix, R. W., G. F. Hatfull & M. C. Smith, (2003) Bacteriophages with tails: chasing their origins and evolution. *Res Microbiol* **154**: 253-257

Hendrix, R. W., J. G. Lawrence, G. F. Hatfull & S. Casjens, (2000) The origins and ongoing evolution of viruses. *Trends Microbiol* **8**: 504-508

Hershey, A. D. & E. Burgi, (1965) Complementary structure of interacting sites at the ends of lambda DNA molecules. *Proc Natl Acad Sci USA* **53**: 325-328

Hershey, A. D. & W. Dove, (1983) Introduction to Lambda. In: Lambda II. R. W. Hendrix, J. W. Roberts, F. W. Stahl & R. A. Weisberg (eds). Cold Spring Harbor Laboratory, pp. 3-12.

Herzer, P. J., S. Inouye, M. Inouye & T. S. Whittam, (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* **172**: 6175-6181

Higgins, R. R., H. J. Lucko & A. Becker, (1988) Mechanism of cos DNA cleavage by bacteriophage lambda terminase: multiple roles of ATP. *Cell* **54**: 765-775

Highton, P. J., Y. Chang & R. J. Myers, (1990) Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. *Mol Microbiol* **4**: 1329-1340

Hilborn, E. D., J. H. Mermin, P. A. Mshar, J. L. Hadler, A. Voetsch, C. Wojtkunski, M. Swartz, R. Mshar, M. A. Lambert-Fair, J. A. Farrar, M. K. Glynn & L. Slutsker, (1999) A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce. *Arch Intern Med* **159**: 1758-1764

Hillyar, C. R. T., (2012) Genetic recombination in bacteriophage lambda. *Bioscience Horizons* **5**.doi:10.1093/biohorizons/hzs001

Ho, Y. S., D. Pfarr, J. Strickler & M. Rosenberg, (1992) Characterization of the transcription activator protein C1 of bacteriophage P22. *J Biol Chem* **267**: 14388-14397

Holst Sorensen, M. C., L. B. van Alphen, C. Fodor, S. M. Crowley, B. B. Christensen, C. M. Szymanski & L. Brondsted, (2012) Phase variable expression of capsular polysaccharide modifications allows *Campylobacter jejuni* to avoid bacteriophage infection in chickens. *Front Cell Infect Microbiol* **2**: 11.doi:10.3389/fcimb.2012.00011. eCollection 2012.

Honish, L., G. Predy, N. Hislop, L. Chui, K. Kowalewska-Grochowska, L. Trottier, C. Kreplin & I. Zazulak, (2005) An outbreak of *E. coli* O157:H7 hemorrhagic colitis associated with unpasteurized gouda cheese. *Can J Public Health* **96**: 182-184

Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Y. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. A. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P. D. Thomas, C. H. Wu, C. Yeats & S. Y. Yong, (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research* **40**: D306-D312.doi:10.1093/nar/gkr948

Hyytia-Trees, E., P. Lafon, P. Vauterin & E. M. Ribot, (2010) Multilaboratory validation study of standardized multiple-locus variable-number tandem repeat analysis protocol for shiga toxin-producing *Escherichia coli* O157: a novel approach to normalise fragment size data between capillary electrophoresis platforms. *Foodborne Pathog Dis* **7**: 129-136

Iida, S., M. B. Streiff, T. A. Bickle & W. Arber, (1987) Two DNA antirestriction systems of bacteriophage P1, *darA*, and *darB*: characterization of *darA*-phages. *Virology* **157**: 156-166

Ikeda, H. & J. Tomizawa, (1968) Prophage P1, and extrachromosomal replication unit. *Cold Spring Harb Symp Quant Biol* **33**: 791-798

Imamovic, L., J. Jofre, H. Schmidt, R. Serra-Moreno & M. Muniesa, (2009) Phage-mediated Shiga toxin 2 gene transfer in food and water. *Appl Environ Microbiol* **75**: 1764-1768

Imamovic, L. & M. Muniesa, (2012) Characterizing RecA-independent induction of Shiga toxin2-encoding phages by EDTA treatment. *PLoS One* **7**: e32393.doi:10.1371/journal.pone.0032393

Islam, M. R., Y. Ogura, M. Asadulghani, T. Ooka, K. Murase, Y. Gotoh & T. Hayashi, (2012) A sensitive and simple plaque formation method for the Stx2 phage of *Escherichia coli* O157:H7, which does not form plaques in the standard plating procedure. *Plasmid* **67**: 227-235

Jakhetia, R., A. Marri, J. Stahle, G. Widmalm & N. K. Verma, (2014) Serotype-conversion in *Shigella flexneri*: identification of a novel bacteriophage, Sf101, from a serotype 7a strain. *BMC Genomics* **15**: 742.doi:10.1186/1471-2164-15-742

James, C. E., K. N. Stanley, H. E. Allison, H. J. Flint, C. S. Stewart, R. J. Sharp, J. R. Saunders & A. J. McCarthy, (2001) Lytic and lysogenic infection of diverse *Escherichia coli* and *Shigella* strains with a verocytotoxigenic bacteriophage. *Appl Environ Microbiol* **67**: 4335-4337

Jandhyala, D. M., T. J. Rogers, A. Kane, A. W. Paton, J. C. Paton & C. M. Thorpe, (2010) Shiga toxin 2 and flagellin from shiga-toxigenic *Escherichia coli*

superinduce interleukin-8 through synergistic effects on host stress-activated protein kinase activation. *Infect Immun* **78**: 2984-2994

Jerse, A. E., J. Yu, B. D. Tall & J. B. Kaper, (1990) A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proc Natl Acad Sci U S A* **87**: 7839-7843

Johnson, E. P., A. R. Strom & D. R. Helinski, (1996) Plasmid RK2 toxin protein ParE: purification and interaction with the ParD antitoxin protein. *J Bacteriol* **178**: 1420-1429

Johnston, C., B. Martin, G. Fichant, P. Polard & J. P. Claverys, (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* **12**: 181-196

Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull & R. W. Hendrix, (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* **299**: 27-51

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

Kajiura, T., M. Tanaka, H. Wada, K. Ito, Y. Koyama & F. Kato, (2001) Effects of disinfectants on Shiga-like toxin converting phage from enterohemorrhagic *Escherichia coli* O157:H7. *J Health Sci* **47**: 203-207

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

Karch, H., H. Bohm, H. Schmidt, F. Gunzer, S. Aleksic & J. Heesemann, (1993) Clonal structure and pathogenicity of Shiga-like toxin-producing, sorbitol-fermenting *Escherichia coli* O157:H. *J Clin Microbiol* **31**: 1200-1205

Karmali, M. A., B. T. Steele, M. Petric & C. Lim, (1983) Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* **1**: 619-620

Karn, J., S. Brenner, L. Barnett & G. Cesareni, (1980) Novel bacteriophage lambda cloning vector. *Proc Natl Acad Sci U S A* **77**: 5172-5176

Katsura, I. & R. W. Hendrix, (1984) Length determination in bacteriophage lambda tails. *Cell* **39**: 691-698

Keene, W. E., E. Sazie, J. Kok, D. H. Rice, D. D. Hancock, V. K. Balan, T. Zhao & M. P. Doyle, (1997) An outbreak of *Escherichia coli* O157:H7 infections traced to jerky made from deer meat. *JAMA* **277**: 1229-1231

Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass & M. J. Sternberg, (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845-858

Khan, F., F. Proulx & C. A. Lingwood, (2009) Detergent-resistant globotriaosyl ceramide may define verotoxin/glomeruli-restricted hemolytic uremic syndrome pathology. *Kidney Int* **75**: 1209-1216

Kim, J., J. Nietfeldt & A. K. Benson, (1999) Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. *Proc Natl Acad Sci USA* **96**: 13288-13293

Kim, J., J. Nietfeldt, J. Ju, J. Wise, N. Fegan, P. Desmarchelier & A. K. Benson, (2001) Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. *J Bacteriol* **183**: 6885-6897

Kim, M. & S. Ryu, (2012) Spontaneous and transient defence against bacteriophage by phase-variable glucosylation of O-antigen in *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* **86**: 411-425

Kobiler, O., A. Rokney, N. Friedman, D. L. Court, J. Stavans & A. B. Oppenheim, (2005) Quantitative kinetic analysis of the bacteriophage lambda genetic network. *Proc Natl Acad Sci USA* **102**: 4470-4475

Kourilsky, P., (1973) Lysogenization by bacteriophage lambda. I. Multiple infection and the lysogenic response. *Mol Gen Genet* **122**: 183-195

Kourilsky, P. & A. Knapp, (1974) Lysogenization by bacteriophage lambda. III. Multiplicity dependent phenomena occuring upon infection by lambda. *Biochimie* **56**: 1517-1523

Krone, F. A., G. Westphal & J. D. Schwenn, (1991) Characterisation of the gene cysH and of its product phospho-adenylylsulphate reductase from *Escherichia coli*. *Mol Gen Genet* **225**: 314-319

Kropinski, A. M., A. Mazzocco, T. E. Waddell, E. Lingohr & R. P. Johnson, (2008) Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. In: Bacteriophages : Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions. A. M. Kropinski & M. R. J. Clokie (eds). pp. 69-76.

Kruger, D. H., G. J. Barcak, M. Reuter & H. O. Smith, (1988) EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res* **16**: 3997-4008

Kulasekara, B. R., M. Jacobs, Y. Zhou, Z. N. Wu, E. Sims, C. Saenphimmachak, L. Rohmer, J. M. Ritchie, M. Radey, M. McKevitt, T. L. Freeman, H. Hayden, E. Haugen, W. Gillett, C. Fong, J. Chang, V. Beskhlebnaya, M. K. Waldor, M. Samadpour, T. S. Whittam, R. Kaul, M. Brittnacher & S. I. Miller, (2009) Analysis of the genome of the *Escherichia coli* O157:H7 2006 spinach-associated outbreak isolate indicates candidate genes that may enhance virulence. *Infect Immun* **77**: 3713-3721

Kusumoto, M., Y. Nishiya, Y. Kawamura & K. Shinagawa, (1999) Identification of an insertion sequence, IS*1203* variant, in a Shiga toxin 2 gene of *Escherichia coli* O157:H7. *J Biosci Bioeng* **87**: 93-96

Kusumoto, M., T. Ooka, Y. Nishiya, Y. Ogura, T. Saito, Y. Sekine, T. Iwata, M. Akiba & T. Hayashi, (2011) Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. *Nat Commun* **2**: 152.doi:10.1038/ncomms1152

Kuzminov, A., (1999) Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**: 751-813

Kyle, J. L., C. A. Cummings, C. T. Parker, B. Quinones, P. Vatta, E. Newton, S. Huynh, M. Swimley, L. Degoricija, M. Barker, S. Fontanoz, K. Nguyen, R. Patel, R. Fang, R. Tebbs, O. Petrauskene, M. Furtado & R. E. Mandrell, (2012) *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. *J Bacteriol* **194**: 1885-1896

L'Abee-Lund, T. M., H. J. Jorgensen, K. O'Sullivan, J. Bohlin, G. Ligard, P. E. Granum & T. Lindback, (2012) The highly virulent 2006 Norwegian EHEC O103:H25 outbreak strain is related to the 2011 German O104:H4 outbreak strain. *PLoS One* **7**: e31413.doi:10.1371/journal.pone.0031413

Lathem, W. W., T. E. Grys, S. E. Witowski, A. G. Torres, J. B. Kaper, P. I. Tarr & R. A. Welch, (2002) StcE, a metalloprotease secreted by *Escherichia coli*

O157:H7, specifically cleaves C1 esterase inhibitor. *Mol Microbiol* **45**: 277-288

Latif, H., H. J. Li, P. Charusanti, B. O. Palsson & R. K. Aziz, (2014) A gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Genome Announc* **2**: e00821-00814.doi:10.1128/genomeA.00821-14.

Launders, N., L. Byrne, N. Adams, K. Glen, C. Jenkins, D. Tubin-Delic, M. Locking, C. Williams & D. Morgan, (2013) Outbreak of Shiga toxin-producing *E. coli* O157 associated with consumption of watercress, United Kingdom, August to September 2013. *Euro Surveill* **18**: 20624.doi:10.2807/1560-7917.ES2013.18.44.20624

Lee, K., N. P. French, Y. Hara-Kudo, S. Iyoda, H. Kobayashi, Y. Sugita-Konishi, H. Tsubone & S. Kumagai, (2011) Multivariate analyses revealed distinctive features differentiating human and cattle isolates of Shiga toxin-producing Escherichia coli O157 in Japan. *J Clin Microbiol* **49**: 1495-1500.10.1128/jcm.02640-10

Leiman, P. G. & M. M. Shneider, (2012) Contractile tail machines of bacteriophages. *Adv Exp Med Biol* **726**: 93-114

Leinfelder, W., K. Forchhammer, F. Zinoni, G. Sawers, M. A. Mandrand-Berthelot & A. Bock, (1988) *Escherichia coli* genes whose products are involved in selenium metabolism. *J Bacteriol* **170**: 540-546

Lemire, S., N. Figueroa-Bossi & L. Bossi, (2011) Bacteriophage crosstalk: coordination of prophage induction by trans-acting antirepressors. *PLoS Genet* **7**: e1002149.doi:10.1371/journal.pgen.1002149

Leotta, G. A., E. S. Miliwebsky, I. Chinen, E. M. Espinosa, K. Azzopardi, S. M. Tennant, R. M. Robins-Browne & M. Rivas, (2008) Characterisation of Shiga toxin-producing *Escherichia coli* O157 strains isolated from humans in Argentina, Australia and New Zealand. *BMC Microbiol* **8**: 46.doi:10.1186/1471-2180-8-46

Levin, M. E., R. W. Hendrix & S. R. Casjens, (1993) A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *J Mol Biol* **234**: 124-139

Levine, M. M., J. G. Xu, J. B. Kaper, H. Lior, V. Prado, B. Tall, J. Nataro, H. Karch & K. Wachsmuth, (1987) A DNA probe to identify enterohemorrhagic *Escherichia coli* of O157:H7 and other serotypes that cause hemorrhagic colitis and hemolytic uremic syndrome. *J Infect Dis* **156**: 175-182

Lewis, J. A. & G. F. Hatfull, (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* **29**: 2205-2216

Lhuillier, S., M. Gallopin, B. Gilquin, S. Brasiles, N. Lancelot, G. Letellier, M. Gilles, G. Dethan, E. V. Orlova, J. Couprie, P. Tavares & S. Zinn-Justin, (2009) Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proc Natl Acad Sci USA* **106**: 8507-8512

Li, H., (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **1303.3997v1 [q-bio.GN]**

Lieberthal, W. & J. S. Levine, (1996) Mechanisms of apoptosis and its potential role in renal tubular epithelial cell injury. *Am J Physiol* **271**: 477-488

Lindberg, A. A., M. Sarvas & P. H. Makela, (1970) Bacteriophage attachment to the somatic antigen of *Salmonella*: effect of O-specific structures in leaky R mutants and s, t1 hybrids. *Infect Immun* **1**: 88-97

Lingwood, C. A., (1994) Verotoxin-binding in human renal sections. *Nephron* **66**: 21-28

Lingwood, C. A., H. Law, S. Richardson, M. Petric, J. L. Brunton, S. De Grandis & M. Karmali, (1987) Glycolipid binding of purified and recombinant *Escherichia coli* produced verotoxin in *vitro*. *J Biol Chem* **262**: 8834-8839

Liu, B., X. Yin, Y. Feng, J. R. Chambers, A. Guo, J. Gong, J. Zhu & C. L. Gyles, (2010) Verotoxin 2 enhances adherence of enterohemorrhagic *Escherichia coli* O157:H7 to intestinal epithelial cells and expression of beta1-integrin by IPEC-J2 cells. *Appl Environ Microbiol* **76**: 4461-4468

Livny, J. & D. I. Friedman, (2004) Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. *Mol Microbiol* **51**: 1691-1704

Loenen, W. A. & N. E. Murray, (1986) Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J Mol Biol* **190**: 11-22

Łoś, J. M., Łoś, M. , A. Węgrzyn & G. Węgrzyn, (2012) Altruism of Shiga toxin-producing *Escherichia coli*: recent hypothesis versus experimental results. *Front Cell Infect Microbiol* **2**: 166.10.3389/fcimb.2012.00166

Łoś, J. M., M. Łoś, A. Węgrzyn & G. Węgrzyn, (2010) Hydrogen peroxide-mediated induction of the Shiga toxin-converting lambdoid prophage ST2-8624 in *Escherichia coli* O157:H7. *FEMS Immunol Med Microbiol* **58**: 322-329

Łoś, M. & G. Węgrzyn, (2012) Pseudolysogeny. In: Advances in Virus Research. Ł. Małgorzata & T. S. Wacław (eds). Academic Press, pp. 339-349.

Lovett, P. S., (1972) PBPI: a flagella specific bacteriophage mediating transduction in *Bacillus pumilus*. *Virology* **47**: 743-752

Low, J. C., I. J. McKendrick, C. McKechnie, D. Fenlon, S. W. Naylor, C. Currie, D. G. Smith, L. Allison & D. L. Gally, (2005) Rectal carriage of enterohemorrhagic *Escherichia coli* O157 in slaughtered cattle. *Appl Environ Microbiol* **71**: 93-97

Lowe, T. M. & S. R. Eddy, (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955-964

Lu, M. J. & U. Henning, (1994) Superinfection exclusion by T-even-type coliphages. *Trends Microbiol* **2**: 137-139

Lu, M. J., Y. D. Stierhof & U. Henning, (1993) Location and unusual membrane topology of the immunity protein of the *Escherichia coli* phage T4. *J Virol* **67**: 4905-4913

Lwoff, A., (1953) Lysogeny. *Bacteriol Rev* **17**: 269-337

MacDonald, D. M., M. Fyfe, A. Paccagnella, A. Trinidad, K. Louie & D. Patrick, (2004) *Escherichia coli* O157:H7 outbreak linked to salami, British Columbia, Canada, 1999. *Epidemiol Infect* **132**: 283-289

Makino, K., K. Ishii, T. Yasunaga, M. Hattori, K. Yokoyama, C. H. Yutsudo, Y. Kubota, Y. Yamaichi, T. Iida, K. Yamamoto, T. Honda, C. G. Han, E. Ohtsubo, M. Kasamatsu, T. Hayashi, S. Kuhara & H. Shinagawa, (1998) Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res* **5**: 1-9

Makino, K., K. Yokoyama, Y. Kubota, C. H. Yutsudo, S. Kimura, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, T. Iida, K. Yamamoto, M. Onishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa & H. Shinagawa, (1999) Complete nucleotide sequence of the prophage VT2-Sakai carrying the

verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157 : H7 derived from the Sakai outbreak. *Genes Genet Syst* **74**: 227-239

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

Mannix, M., N. O'Connell, E. McNamara, A. Fitzgerald, T. Prendiville, T. Norris, T. Greally, R. Fitzgerald, D. Whyte, D. Barron, R. Monaghan, E. Whelan, A. Carroll, A. Curtin, C. Collins, J. Quinn, F. O'Dea, M. O'Riordan, J. Buckley, J. McCarthy, P. Mc Keown & T. Multi-agency Outbreak Control, (2005) Large *E. coli* O157 outbreak in Ireland, October-November 2005. *Euro Surveill* **10**: E051222

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

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley & J. M. Rothberg, (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380

Marinus, M. G. & A. R. Poteete, (2013) High efficiency generalized transduction in *Escherichia coli* O157:H7. *F1000Research* **2**: 7.doi:10.12688/f1000research.2-7.v1

Martinsohn, J. T., M. Radman & M. A. Petit, (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet* **4**: e1000065.doi:10.1371/journal.pgen.1000065

Mazzocco, A., T. E. Waddell, E. Lingohr & R. P. Johnson, (2008) Enumeration of bacteriophages using the small drop plaque assaysystem. In: Bacteriophages : Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions. A. M. Kropinski & M. R. J. Clokie (eds). Humana Press, pp. 81-85.

McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg & J. B. Kaper, (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* **92**: 1664-1668

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

Mellor, G. E., N. Fegan, K. S. Gobius, H. V. Smith, A. V. Jennison, B. A. D'Astek, M. Rivas, S. Shringi, K. N. Baker & T. E. Besser, (2015) Geographically

distinct *Escherichia coli* O157 isolates differ by lineage, Shiga toxin genotype, and total Shiga toxin production. *J Clin Microbiol* **53**: 579-586

Mellor, G. E., E. M. Sim, R. S. Barlow, B. A. D'Astek, L. Galli, I. Chinen, M. Rivas & K. S. Gobius, (2012) Phylogenetically related Argentinean and Australian *Escherichia coli* O157 isolates are distinguished by virulence clades and alternative Shiga toxin 1 and 2 prophages. *Appl Environ Microbiol* **78**: 4724-4731

Meltz Steinberg, K. & B. R. Levin, (2007) Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. *Proc Biol Sci* **274**: 1921-1929

Meyer, J. R., D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick & R. E. Lenski, (2012) Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**: 428-432

Michino, H., K. Araki, S. Minami, S. Takaya, N. Sakai, M. Miyazaki, A. Ono & H. Yanagawa, (1999) Massive outbreak of *Escherichia coli* O157:H7 infection in school children in Sakai City, Japan, associated with consumption of white radish sprouts. *Am J Epidemiol* **150**: 787-796

Milne, I., G. Stephen, M. Bayer, P. J. Cock, L. Pritchard, L. Cardle, P. D. Shaw & D. Marshall, (2013) Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* **14**: 193-202

Mitraki, A., S. Miller & M. J. van Raaij, (2002) Review: conformation and folding of novel beta-structural elements in viral fiber proteins: the triple beta-spiral and triple beta-helix. *J Struct Biol* **137**: 236-247

Miyamoto, H., W. Nakai, N. Yajima, A. Fujibayashi, T. Higuchi, K. Sato & A. Matsushiro, (1999) Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions. *DNA Res* **6**: 235-240

Miyamoto, Y., M. Iimura, J. B. Kaper, A. G. Torres & M. F. Kagnoff, (2006) Role of Shiga toxin versus H7 flagellin in enterohaemorrhagic *Escherichia coli* signalling of human colon epithelium in vivo. *Cell Microbiol* **8**: 869-879

Moon, H. W., S. C. Whipp, R. A. Argenzio, M. M. Levine & R. A. Giannella, (1983) Attaching and effacing activities of rabbit and human enteropathogenic *Escherichia coli* in pig and rabbit intestines. *Infect Immun* **41**: 1340-1351

Morse, M. L., E. M. Lederberg & J. Lederberg, (1956) Transduction in *Escherichia coli* K-12. *Genetics* **41**: 142-156

Mühldorfer, I., J. Hacker, G. T. Keusch, D. W. Acheson, H. Tschäpe, A. V. Kane, A. Ritter, T. Olschläger & A. Donohue-Rolfe, (1996) Regulation of the Shiga-like toxin II operon in *Escherichia coli*. *Infect Immun* **64**: 495-502

Muniesa, M., J. E. Blanco, M. De Simon, R. Serra-Moreno, A. R. Blanch & J. Jofre, (2004a) Diversity of Stx2 converting bacteriophages induced from Shiga-toxin-producing *Escherichia coli* strains isolated from cattle. *Microbiology* **150**: 2959-2971

Muniesa, M., J. Recktenwald, M. Bielaszewska, H. Karch & H. Schmidt, (2000) Characterization of a Shiga toxin 2e-converting bacteriophage from an *Escherichia coli* strain of human origin. *Infect Immun* **68**: 4850-4855

Muniesa, M., R. Serra-Moreno & J. Jofre, (2004b) Free Shiga toxin bacteriophages isolated from sewage showed diversity although the *stx* genes appeared conserved. *Environ Microbiol* **6**: 716-725

Nasi, S., L. Paolozzi & E. Calef, (1974) Mutations in the CI gene of lambda bacteriophage which retain immunity and confer a rex. *Mol Gen Genet*

130: 297-305

Neely, M. N. & D. I. Friedman, (1998a) Arrangement and functional identification of genes in the regulatory region of lambdoid phage H-19B, a carrier of a Shiga-like toxin. *Gene* **223**: 105-113

Neely, M. N. & D. I. Friedman, (1998b) Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. *Mol Microbiol* **28**: 1255-1267

Neil, K. P., G. Biggerstaff, J. K. MacDonald, E. Trees, C. Medus, K. A. Musser, S. G. Stroika, D. Zink & M. J. Sotir, (2012) A novel vehicle for transmission of *Escherichia coli* O157:H7 to humans: multistate outbreak of *E. coli* O157:H7 infections associated with consumption of ready-to-bake commercial prepackaged cookie dough--United States, 2009. *Clin Infect Dis* **54**: 511-518

Neisser, M. & K. Shiga, (1903) Ueber freie Receptoren von Typhusund Dysenterie-Bazillen und ueber das Dysenterie Toxin. *Dtsch Med Wochenschr* **29**: 61-62

Neufing, P. J., K. E. Shearwin, J. Camerotto & J. B. Egan, (1996) The CII protein of bacteriophage 186 establishes lysogeny by activating a promoter upstream of the lysogenic promoter. *Mol Microbiol* **21**: 751-761

Neupane, M., G. S. Abu-Ali, A. Mitra, D. W. Lacher, S. D. Manning & J. T. Riordan, (2011) Shiga toxin 2 overexpression in *Escherichia coli* O157:H7 strains associated with severe human disease. *Microb Pathog* **51**: 466-470

Nilsson, A. S., J. L. Karlsson & E. Haggård-Ljungquist, (2004) Site-specific recombination links the evolution of p2-like coliphages and pathogenic Enterobacteria. *Mol Biol Evol* **21**: 1-13

Nilsson, L. & V. Emilsson, (1994) Factor for inversion stimulation-dependent growth rate regulation of individual tRNA species in *Escherichia coli*. *J Biol Chem* **269**: 9460-9465

Nimri, L. F., (2013) *Escherichia albertii*, a newly emerging enteric pathogen with poorly defined properties. *Diagn Microbiol Infect Dis* **77**: 91-95

Noller, A. C., M. C. McEllistrem, O. C. Stine, J. G. Morris, Jr., D. J. Boxrud, B. Dixon & L. H. Harrison, (2003) Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. *J Clin Microbiol* **41**: 675-679

Novick, R. P., G. E. Christie & J. R. Penades, (2010) The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol* **8**: 541-551

Nubling, S., T. Eisele, H. Stober, J. Funk, S. Polzin, L. Fischer & H. Schmidt, (2014) Bacteriophage 933W encodes a functional esterase downstream of the Shiga toxin 2a operon. *Int J Med Microbiol* **304**: 269-274

O'Brien, A. D. & R. K. Holmes, (1987) Shiga and Shiga-like toxins. *Microbiol Rev* **51**: 206-220

O'Brien, A. D., J. W. Newland, S. F. Miller, R. K. Holmes, H. W. Smith & S. B. Formal, (1984) Shiga like toxin converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226**: 694-696

O'Shea, Y. A. & E. F. Boyd, (2002) Mobilization of the Vibrio pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction. *FEMS Microbiol Lett* **214**: 153-157

Obata, F., K. Tohyama, A. D. Bonev, G. L. Kolling, T. R. Keepers, L. K. Gross, M. T. Nelson, S. Sato & T. G. Obrig, (2008) Shiga toxin 2 affects the central

nervous system through receptor globotriaosylceramide localized to neurons. *J Infect Dis* **198**: 1398-1406

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

Ochman, H. & R. K. Selander, (1984) Evidence for clonal population structure in *Escherichia coli*. *Proc Natl Acad Sci USA* **81**: 198-201

Ogura, Y., K. Kurokawa, T. Ooka, K. Tashiro, T. Tobe, M. Ohnishi, K. Nakayama, T. Morimoto, J. Terajima, H. Watanabe, S. Kuhara & T. Hayashi, (2006) Complexity of the genomic diversity in enterohemorrhagic *Escherichia coli* O157 revealed by the combinational use of the O157 Sakai OligoDNA microarray and the Whole Genome PCR scanning. *DNA Res* **13**: 3-14

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

Ohtsubo, H., M. Zenilman & E. Ohtsubo, (1980) Insertion element IS*102* resides in plasmid pSC101. *J Bacteriol* **144**: 131-140

Okitsu, T., M. Kusumoto, R. Suzuki, S. Sata, Y. Nishiya, Y. Kawamura & S. Yamai, (2001) Identification of shiga toxin-producing *Escherichia coli* possessing insertionally inactivated Shiga toxin gene. *Microbiol Immunol* **45**: 319-322

Olorunniji, F. J., D. E. Buck, S. D. Colloms, A. R. McEwan, M. C. Smith, W. M. Stark & S. J. Rosser, (2012) Gated rotation mechanism of site-specific recombination by varphiC31 integrase. *Proc Natl Acad Sci USA* **109**: 19661-19666

Ooka, T., Y. Ogura, M. Asadulghani, M. Ohnishi, K. Nakayama, J. Terajima, H. Watanabe & T. Hayashi, (2009) Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Res* **19**: 1809-1816

Ooka, T., K. Seto, K. Kawano, H. Kobayashi, Y. Etoh, S. Ichihara, A. Kaneko, J. Isobe, K. Yamaguchi, K. Horikawa, T. A. Gomes, A. Linden, M. Bardiau, J. G. Mainil, L. Beutin, Y. Ogura & T. Hayashi, (2012) Clinical significance of *Escherichia albertii*. *Emerg Infect Dis* **18**: 488-492

Orlova, E. V., B. Gowen, A. Droge, A. Stiege, F. Weise, R. Lurz, M. van Heel & P. Tavares, (2003) Structure of a viral DNA gatekeeper at 10 A resolution by cryo-electron microscopy. *EMBO J* **22**: 1255-1262

Ørskov, I., F. Ørskov, B. Jann & K. Jann, (1977) Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol Rev* **41**: 667-710

Orth, D., K. Grif, A. B. Khan, A. Naim, M. P. Dierich & R. Wurzner, (2007) The Shiga toxin genotype rather than the amount of Shiga toxin or the cytotoxicity of Shiga toxin in vitro correlates with the appearance of the hemolytic uremic syndrome. *Diagn Microbiol Infect Dis* **59**: 235-242

Ostroff, S. M., P. I. Tarr, M. A. Neill, J. H. Lewis, N. Hargrett-Bean & J. M. Kobayashi, (1989) Toxin genotypes and plasmid profiles as determinants of systemic sequelae in *Escherichia coli* O157:H7 infections. *J Infect Dis* **160**: 994-998

Otsuka, Y. & T. Yonesaki, (2012) Dmd of bacteriophage T4 functions as an antitoxin against *Escherichia coli* LsoA and RnlA toxins. *Mol Microbiol* **83**: 669-681

Padmanabhan, R., R. Wu & R. Calendar, (1974) Complete nucleotide sequence of the cohesive ends of bacteriophage P2 deoxyribonucleic acid. *J Biol Chem* **249**: 6197-6207

Parma, D. H., M. Snyder, S. Sobolevski, M. Nawroz, E. Brody & L. Gold, (1992) The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev* **6**: 497-510

Paton, A. W. & J. C. Paton, (1996) *Enterobacter cloacae* producing a Shiga-like toxin II-related cytotoxin associated with a case of hemolytic-uremic syndrome. *Journal of Clinical Microbiology* **34**: 463-465

Paton, A. W. & J. C. Paton, (1998) Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx*1, *stx*2, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfbO111*, and *rfbO157*. *J Clin Microbiol* **36**: 598-602

Peacock, E., V. W. Jacob & S. M. Fallone, (2001) *Escherichia coli* O157:H7: etiology, clinical features, complications, and treatment. *Nephrol Nurs J* **28**: 547-550

Pecota, D. C. & T. K. Wood, (1996) Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *J Bacteriol* **178**: 2044-2050

Perna, N. T., G. F. Mayhew, G. Posfai, S. Elliott, M. S. Donnenberg, J. B. Kaper & F. R. Blattner, (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **66**: 3810-3817

Perna, N. T., G. Plunkett, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Limk, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Y. Lin, G. Yen, D. C. Schwartz, R. A. Welch & F. R. Blattner, (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**: 529-533

Peters, M., A. Tomikas & A. Nurk, (2004) Organization of the horizontally transferred *pheBA* operon and its adjacent genes in the genomes of eight indigenous *Pseudomonas* strains. *Plasmid* **52**: 230-236

Petty, N. K., R. Bulgin, V. F. Crepin, A. M. Cerdeno-Tarraga, G. N. Schroeder, M. A. Quail, N. Lennard, C. Corton, A. Barron, L. Clark, A. L. Toribio, J. Parkhill, G. Dougan, G. Frankel & N. R. Thomson, (2010) The *Citrobacter rodentium* genome sequence reveals convergent evolution with human pathogenic *Escherichia coli*. *J Bacteriol* **192**: 525-538

Petty, N. K., T. Feltwell, D. Pickard, S. Clare, A. L. Toribio, M. Fookes, K. Roberts, R. Monson, S. Nair, R. A. Kingsley, R. Bulgin, S. Wiles, D. Goulding, T. Keane, C. Corton, N. Lennard, D. Harris, D. Willey, R. Rance, L. Yu, J. S. Choudhary, C. Churcher, M. A. Quail, J. Parkhill, G. Frankel, G. Dougan, G. P. Salmond & N. R. Thomson, (2011) *Citrobacter rodentium* is an unstable pathogen showing evidence of significant genomic flux. *PLoS Pathog* **7**: e1002018.doi:10.1371/journal.ppat.1002018

Petty, N. K., I. J. Foulds, E. Pradel, J. J. Ewbank & G. P. Salmond, (2006) A generalized transducing phage (phiIF3) for the genomically sequenced *Serratia marcescens* strain Db11: a tool for functional genomics of an opportunistic human pathogen. *Microbiology* **152**: 1701-1708

Petty, N. K., A. L. Toribio, D. Goulding, I. Foulds, N. Thomson, G. Dougan & G. P. Salmond, (2007) A generalised transducing phage for the murine pathogen *Citrobacter rodentium*. *Microbiology* **153**: 2984-2988

Pingoud, A., M. Fuxreiter, V. Pingoud & W. Wende, (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* **62**: 685-707

Plunkett, G., 3rd, D. J. Rose, T. J. Durfee & F. R. Blattner, (1999) Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J Bacteriol* **181**: 1767-1778

Poteete, A. R., (2001) What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS microbiol lett* **201**: 9-14

Ptashne, M., (2004) *A Genetic Switch: Phage Lambda Revisited*. Cold Spring Harbor Laboratory Press.

Quiles-Puchalt, N., N. Carpena, J. C. Alonso, R. P. Novick, A. Marina & J. R. Penadés, (2014) Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 6016-6021.10.1073/pnas.1320538111

Randall-Hazelbauer, L. & M. Schwartz, (1973) Isolation of the bacteriophage lambda receptor from *Escherichia coli*. *J Bacteriol* **116**: 1436-1446

Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull & R. W. Hendrix, (2000) Genomic sequence and analysis of the atypical temperate bacteriophage N15. *J Mol Biol* **299**: 53-73

Raya, R. R. & E. M. Hébert, (2008) Isolation of phage via induction of lysogens. In: Bacteriophages : Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions. A. M. Kropinski & M. R. J. Clokie (eds). pp. 23-32.

Recktenwald, J. & H. Schmidt, (2002) The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage φp27 is not related to other Sx phage genomes, but the modular genetic structure is conserved. *Infect Immun* **70**: 1896-1908

Refardt, D., (2012) Real-time quantitative PCR to discriminate and quantify lambdoid bacteriophages of *Escherichia coli* K-12. *Bacteriophage* **2**: 98-104

Reida, P., M. Wolff, H. W. Pohls, W. Kuhlmann, A. Lehmacher, S. Aleksic, H. Karch & J. Bockemuhl, (1994) An outbreak due to enterohaemorrhagic *Escherichia coli* O157:H7 in a children day care centre characterized by person-to-person transmission and environmental contamination. *Zentralbl Bakteriol* **281**: 534-543

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

Rivas, M., I. Chinen, E. Miliwebsky & M. Masana, (2014) Risk factors for Shiga toxin-producing *Escherichia coli*-associated human diseases. *Microbiol Spectr* **2**.doi:10.1128/microbiolspec.EHEC-0002-2013

Roberts, J. W. & J. E. Steitz, (1967) The reconstitution of infective bacteriophage R17. *Proc Natl Acad Sci USA* **58**: 1416-1421

Robinson, C. M., J. F. Sinclair, M. J. Smith & A. D. O'Brien, (2006) Shiga toxin of enterohemorrhagic *Escherichia coli* type O157:H7 promotes intestinal colonization. *Proc Natl Acad Sci U S A* **103**: 9667-9672

Rode, T. M., L. Axelsson, P. E. Granum, E. Heir, A. Holck & T. M. L'Abee-Lund, (2011) High stability of Stx2 phage in food and under food-processing conditions. *Appl Environ Microbiol* **77**: 5336-5341

Roessner, C. A. & G. M. Ihler, (1984) Proteinase sensitivity of bacteriophage lambda tail proteins GpJ and pH in complexes with the lambda receptor. *J Bacteriol* **157**: 165-170

Romig, W. R. & A. M. Brodetsky, (1961) Isolation and preliminary characterization of bacteriophages for *Bacillus subtilis*. *J Bacteriol* **82**: 135-141

Rumer, L., J. Jores, P. Kirsch, Y. Cavignac, K. Zehmke & L. H. Wieler, (2003) Dissemination of *pheU* and *pheV* located genomic islands among enteropathogenic (EPEC) and enterohemorrhagic (EHEC) *E. coli* and their possible role in the horizontal transfer of the locus of enterocyte effacement (LEE). *Int J Med Microbiol* **292**: 463-475

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream & B. Barrell, (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945

Sam, M. D., D. Cascio, R. C. Johnson & R. T. Clubb, (2004) Crystal structure of the excisionase-DNA complex from bacteriophage lambda. *J Mol Biol* **338**: 229-240

Sambrook, J. & D. Russell, (2001) *Molecular cloning: a laboratory manual (third edition)*. Cold Spring Harbor Laboratory Press.

Samson, J. E., A. H. Magadan, M. Sabri & S. Moineau, (2013) Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**: 675-687

Sandvig, K., O. Garred, K. Prydz, J. V. Kozlov, S. H. Hansen & B. van Deurs, (1992) Retrograde transport of endocytosed Shiga toxin to the endoplasmic reticulum. *Nature* **358**: 510-512

Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill & G. B. Petersen, (1982) Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* **162**: 729-773

Sato, T., T. Shimizu, M. Watarai, M. Kobayashi, S. Kano, T. Hamabata, Y. Takeda & S. Yamasaki, (2003a) Distinctiveness of the genomic sequence of Shiga toxin 2-converting phage isolated from *Escherichia coli* O157:H7 Okayama strain as compared to other Shiga toxin 2-converting phages. *Gene* **309**: 35-48

Sato, T., T. Shimizu, M. Watarai, M. Kobayashi, S. Kano, T. Hamabata, Y. Takeda & S. Yamasaki, (2003b) Genome analysis of a novel Shiga toxin 1 (Stx1)-converting phage which is closely related to Stx2-converting phages but not to other Stx1-converting phages. *J Bacteriol* **185**: 3966-3971

Sauer, R. T., W. Krovatin, J. DeAnda, P. Youderian & M. M. Susskind, (1983) Primary structure of the *immI* immunity region of bacteriophage P22. *J Mol Biol* **168**: 699-713

Savill, J., A. Mooney & J. Hughes, (1996) Apoptosis and renal scarring. *Kidney Int Suppl* **54**: 14-17

Scheiring, J., S. P. Andreoli & L. B. Zimmerhackl, (2008) Treatment and outcome of Shiga-toxin-associated hemolytic uremic syndrome (HUS). *Pediatr Nephrol* **23**: 1749-1760

Scherer, G., (1978) Nucleotide sequence of the O gene and of the origin of replication in bacteriophage lambda DNA. *Nucleic Acids Res* **5**: 3141-3156

Scheutz, F., L. D. Teel, L. Beutin, D. Pierard, G. Buvens, H. Karch, A. Mellmann, A. Caprioli, R. Tozzoli, S. Morabito, N. A. Strockbine, A. R. Melton-Celsa, M. Sanchez, S. Persson & A. D. O'Brien, (2012) Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J Clin Microbiol* **50**: 2951-2963

Schmidt, H., B. Henkel & H. Karch, (1997) A gene cluster closely related to type II secretion pathway operons of gram-negative bacteria is located on the large

plasmid of enterohemorrhagic *Escherichia coli* O157 strains. *FEMS Microbiol Lett* **148**: 265-272

Schmidt, H., M. Montag, J. Bockemuhl, J. Heesemann & H. Karch, (1993) Shiga-like toxin II-related cytotoxins in *Citrobacter freundii* strains from humans and beef samples. *Infect Immun* **61**: 534-543

Schmidt, H., J. Scheef, S. Morabito, A. Caprioli, L. H. Wieler & H. Karch, (2000) A new Shiga toxin 2 variant (Stx2f) from Escherichia coli isolated from pigeons. *Appl Environ Microbiol* **66**: 1205-1208

Schmidt, M. A., (2010) LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* **12**: 1544-1552

Schmitt, C. K., P. Kemp & I. J. Molineux, (1991) Genes 1.2 and 10 of bacteriophages T3 and T7 determine the permeability lesions observed in infected cells of *Escherichia coli* expressing the F plasmid gene *pifA*. *J Bacteriol* **173**: 6507-6514

Schuller, S., R. Heuschkel, F. Torrente, J. B. Kaper & A. D. Phillips, (2007) Shiga toxin binding in normal and inflamed human intestinal mucosa. *Microbes Infect* **9**: 35-39

Schureck, M. A., T. Maehigashi, S. J. Miles, J. Marquez, S. E. Cho, R. Erdman & C. M. Dunham, (2014) Structure of the Proteus vulgaris HigB-(HigA)2-HigB toxin-antitoxin complex. *J Biol Chem* **289**: 1060-1070

Schwarz, E., G. Scherer, G. Hobom & H. Kossel, (1978) Nucleotide sequence of *cro*, *cII* and part of the *O* gene in phage lambda DNA. *Nature* **272**: 410-414

Scotland, S. M., H. R. Smith, G. A. Willshaw & B. Rowe, (1983) Vero cyto-toxin production in strain of *Escherichia coli* is determined by genes carried on bacteriophage. *Lancet* **2**: 216-216

Seed, K. D., D. W. Lazinski, S. B. Calderwood & A. Camilli, (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**: 489-491

Serra-Moreno, R., J. Jofre & M. Muniesa, (2007) Insertion site occupancy by Stx2 bacteriophages depends on the locus availability of the host strain chromosome. *J Bacteriol* **189**: 6645-6654

Shin, H., J. H. Lee, H. Kim, Y. Choi, S. Heu & S. Ryu, (2012) Receptor diversity and host interaction of bacteriophages infecting *Salmonella enterica* serovar Typhimurium. *PLoS One* **7**: e43392.doi:10.1371/journal.pone.0043392

Shringi, S., A. García, K. K. Lahmers, K. A. Potter, S. Muthupalani, A. G. Swennes, C. J. Hovde, D. R. Call, J. G. Fox & T. E. Besser, (2012) Differential virulence of clinical and bovine-biased enterohemorrhagic *Escherichia coli* O157:H7 genotypes in piglet and dutch belted rabbit models. *Infect Immun* **80**: 369-380

Siegler, R. L., T. G. Obrig, T. J. Pysher, V. L. Tesh, N. D. Denkers & F. B. Taylor, (2003) Response to Shiga toxin 1 and 2 in a baboon model of hemolytic uremic syndrome. *Pediatr Nephrol* **18**: 92-96

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson & D. G. Higgins, (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539. doi:510.1038/msb.2011.1075.10.1038/msb.2011.75

Siguier, P., J. Filee & M. Chandler, (2006a) Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* **9**: 526-531

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon & M. Chandler, (2006b) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32-36.doi: 10.1093/nar/gkj014

Sim, E. M., (2010) Comparative genomic analysis of Shiga toxin bacteriophages from Australian human and food source *Escherichia coli* O157. In: The University of Queensland (Honours Thesis).

Simon, M. N., R. W. Davis & N. Davidson, (1971) Heteroduplexes of DNA Molecules of Lambdoid Phages: Physical Mapping of Their Base Sequence Relationships by Electron Microscopy. In: The Bacteriophage Lambda. A. D. Hershey (ed). Cold Spring Harbour Monographs, pp. 313-328.

Simonsen, J., M. Frisch & S. Ethelberg, (2008) Socioeconomic risk factors for bacterial gastrointestinal infections. *Epidemiology* **19**: 282-290

Six, E. W., (1975) The helper dependence of satellite bacteriophage P4: which gene functions of bacteriophage P2 are needed by P4? *Virology* **67**: 249-263

Smith, D. L., C. E. James, M. J. Sergeant, Y. Yaxian, J. R. Saunders, A. J. McCarthy & H. E. Allison, (2007a) Short-tailed Stx phages exploit the conserved YaeT protein to disseminate Shiga Toxin genes among enterobacteria. *Journal of Bacteriology* **189**: 7223-7233

Smith, D. L., D. J. Rooks, P. C. M. Fogg, A. C. Darby, N. R. Thomson, A. J. McCarthy & H. E. Allison, (2012) Comparative genomics of Shiga toxin encoding bacteriophages. *Bmc Genomics* **13**.doi:10.1186/1471-2164-13-311

Smith, D. L., B. M. Wareing, P. C. Fogg, L. M. Riley, M. Spencer, M. J. Cox, J. R. Saunders, A. J. McCarthy & H. E. Allison, (2007b) Multilocus characterization scheme for Shiga toxin-encoding bacteriophages. *Appl Environ Microbiol* **73**: 8032-8040

Smith, H. W., P. Green & Z. Parsell, (1983) Vero cell toxins in *Escherichia coli* and related bacteria: transfer by phage and conjugation and toxic action in laboratory animals, chickens and pigs. *J Gen Microbiol* **129**: 3121-3137

Smith, H. W. & M. A. Linggood, (1971) Transmissible nature of enterotoxin production in a human enteropathogenic strain of *Escherichia coli*. *J Med Microbiol* **4**: 301-305

Smith, M. C., N. Burns, J. R. Sayers, J. A. Sorrell, S. R. Casjens & R. W. Hendrix, (1998) Bacteriophage collagen. *Science* **279**: 1834

Smith, M. C., R. Till, K. Brady, P. Soultanas, H. Thorpe & M. C. Smith, (2004) Synapsis and DNA cleavage in phiC31 integrase-mediated site-specific recombination. *Nucleic Acids Res* **32**: 2607-2617

Snedeker, K. G., D. J. Shaw, M. E. Locking & R. J. Prescott, (2009) Primary and secondary cases in *Escherichia coli* O157 outbreaks: a statistical analysis. *BMC Infect Dis* **9**: 144.doi:10.1186/1471-2334-9-144

Snyder, L., (1995) Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol Microbiol* **15**: 415-420

Solheim, H. T., C. Sekse, A. M. Urdahl, Y. Wasteson & L. L. Nesse, (2013) Biofilm as an environment for dissemination of stx genes by transduction. *Appl Environ Microbiol* **79**: 896-900

Sorek, R., V. Kunin & P. Hugenholtz, (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181-186

Sperschneider, J., A. Datta & M. J. Wise, (2011) Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. *Rna* **17**: 27-38

Spina, N., S. Zansky, N. Dumas & S. Kondracki, (2005) Four laboratory-associated cases of infection with *Escherichia coli* O157:H7. *J Clin Microbiol* **43**: 2938-2939

Stearns-Kurosawa, D. J., V. Collins, S. Freeman, V. L. Tesh & S. Kurosawa, (2010) Distinct physiologic and inflammatory responses elicited in baboons after challenge with Shiga toxin type 1 or 2 from enterohemorrhagic *Escherichia coli*. *Infect Immun* **78**: 2497-2504

Steven, A. C. & B. L. Trus, (1986) The structure of bacteriophage T7. In: Electron microscopy of proteins, viral strutures. Volume 5. J. R. Harris & R. W. Horne (eds). London academic press, pp. 1-33.

Stockdale, S. R., J. Mahony, P. Courtin, M. P. Chapot-Chartier, J. P. van Pijkeren, R. A. Britton, H. Neve, K. J. Heller, B. Aideh, F. K. Vogensen & D. van Sinderen, (2013) The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. *J Biol Chem* **288**: 5581-5590

Stocker, B. A., (1956) Abortive transduction of motility in *Salmonella*; a nonreplicated gene transmitted through many generations to a single descendant. *J Gen Microbiol* **15**: 575-598

Stolfa, G. & G. B. Koudelka, (2012) Entry and killing of *Tetrahymena thermophila* by bacterially produced Shiga toxin. *MBio* **4**: e00416-00412.doi:10.1128/mBio.00416-12

Strauch, E., J. A. Hammerl, A. Konietzny, S. Schneiker-Bekel, W. Arnold, A. Goesmann, A. Puhler & L. Beutin, (2008) Bacteriophage 2851 Is a Prototype Phage for Dissemination of the Shiga Toxin Variant Gene 2c in *Escherichia coli* O157:H7. *Infect Immun* **76**: 5466-5477

Strauch, E., R. Lurz & L. Beutin, (2001) Characterization of a Shiga toxin-encoding temperate bacteriophage of *Shigella sonnei*. *Infect Immun* **69**: 7588-7595

Strauch, E., C. Schaudinn & L. Beutin, (2004) First-time isolation and characterization of a bacteriophage encoding the Shiga toxin 2c variant, which is globally spread in strains of *Escherichia coli* O157. *Infection and Immunity* **72**: 7030-7039

Studier, F. W. & N. R. Movva, (1976) SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J Virol* **19**: 136-145

Su, L. K., C. P. Lu, Y. Wang, D. M. Cao, J. H. Sun & Y. X. Yan, (2010) Lysogenic infection of a Shiga toxin 2-converting bacteriophage changes host gene expression, enhances host acid resistance and motility. *Mol Biol* **44**: 60-73

Su, L. K., C. P. Lu, L. M. Xia, Y. Wang, J. Wang, J. H. Sun & Y. Yan, (2009) Characterization of a Shiga toxin 2-encoding bacteriophage phimin27 isolated from *Escherichia* coli O157:H7 strain of china. *Afr J Microbiol Res* **3**: 799-808

Sullivan, M. J., N. K. Petty & S. A. Beatson, (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009-1010

Susskind, M. M. & D. Botstein, (1975) Mechanism of action of *Salmonella* phage P22 antirepressor. *J Mol Biol* **98**: 413-424

Susskind, M. M. & D. Botstein, (1978) Molecular genetics of bacteriophage P22. *Microbiol Rev* **42**: 385-413

Suttle, C. A., (2007) Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801-812

Sváb, D., B. Bálint, G. Maróti & I. Tóth, (2015) A novel transducible chimeric phage from *Escherichia coli* O157:H7 Sakai strain encoding Stx1 production. *Infect Genet Evol* **29**: 42-47

Swerdlow, D. L., B. A. Woodruff, R. C. Brady, P. M. Griffin, S. Tippen, H. D. Donnell, Jr., E. Geldreich, B. J. Payne, A. Meyer, Jr., J. G. Wells & et al., (1992) A waterborne outbreak in Missouri of *Escherichia coli* O157:H7 associated with bloody diarrhea and death. *Ann Intern Med* **117**: 812-819

Tarkowski, T. A., D. Mooney, L. C. Thomason & F. W. Stahl, (2002) Gene products encoded in the *ninR* region of phage lambda participate in Red-mediated recombination. *Genes to Cells* **7**: 351-363

Tarr, P. I., C. A. Gordon & W. L. Chandler, (2005) Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* **365**: 1073-1086

Tatsuno, I., M. Horie, H. Abe, T. Miki, K. Makino, H. Shinagawa, H. Taguchi, S. Kamiya, T. Hayashi & C. Sasakawa, (2001) *toxB* gene on pO157 of enterohemorrhagic *Escherichia coli* O157:H7 is required for full epithelial cell adherence phenotype. *Infect Immun* **69**: 6660-6669

Tauschek, M., R. A. Strugnell & R. M. Robins-Browne, (2002) Characterization and evidence of mobilization of the LEE pathogenicity island of rabbit-specific strains of enteropathogenic *Escherichia coli*. *Mol Microbiol* **44**: 1533-1550

Taylor, K., (1965) Enzymatic deacetylation of Vi-polysaccharide by Vi-phage. II. *Biochem Biophys Res Commun* **20**: 752-756

Tesh, V. L., (2010) Induction of apoptosis by Shiga toxins. *Future Microbiol* **5**: 431-453

Tesh, V. L., J. A. Burris, J. W. Owens, V. M. Gordon, E. A. Wadolkowski, A. D. O'Brien & J. E. Samuel, (1993) Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. *Infect Immun* **61**: 3392-3402

Thisted, T., N. S. Sorensen, E. G. Wagner & K. Gerdes, (1994) Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5'-end single-stranded leader and competes with the 3'-end of Hok mRNA for binding to the mok translational initiation region. *EMBO J* **13**: 1960-1968

Tobe, T., S. A. Beatson, H. Taniguchi, H. Abe, C. M. Bailey, A. Fivian, R. Younis, S. Matthews, O. Marches, G. Frankel, T. Hayashi & M. J. Pallen, (2006) An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc Natl Acad Sci U S A* **103**: 14941-14946

Tock, M. R. & D. T. Dryden, (2005) The biology of restriction and anti-restriction. *Curr Opin Microbiol* **8**: 466-472.10.1016/j.mib.2005.06.003

Toshima, H., A. Yoshimura, K. Arikawa, A. Hidaka, J. Ogasawara, A. Hase, H. Masaki & Y. Nishikawa, (2007) Enhancement of Shiga toxin production in enterohemorrhagic *Escherichia coli* serotype O157:H7 by DNase colicins. *Appl Environ Microbiol* **73**: 7582-7588

Tozzoli, R., L. Grande, V. Michelacci, P. Ranieri, A. Maugliani, A. Caprioli & S. Morabito, (2014) Shiga toxin-converting phages and the emergence of new pathogenic *Escherichia coli*: a world in motion. *Front Cell Infect Microbiol* **4**.doi:10.3389/fcimb.2014.00080

Trotz-Williams, L. A., N. J. Mercer, J. M. Walters, A. M. Maki & R. P. Johnson, (2012) Pork implicated in a Shiga toxin-producing *Escherichia coli* O157:H7 outbreak in Ontario, Canada. *Can J Public Health* **103**: 322-326

Tschape, H., R. Prager, W. Streckel, A. Fruth, E. Tietze & G. Bohme, (1995) Verotoxinogenic *Citrobacter freundii* associated with severe gastroenteritis

and cases of hemolytic uremic syndrome in a nursery school green butter as the infection source. *Epidemiol Infect* **114**: 441-450

Tsurimoto, T. & K. Matsubara, (1981) Purified bacteriophage lambda O protein binds to four repeating sequences at the lambda replication origin. *Nucleic Acids Res* **9**: 1789-1799

Tuttle, J., T. Gomez, M. P. Doyle, J. G. Wells, T. Zhao, R. V. Tauxe & P. M. Griffin, (1999) Lessons from a large outbreak of *Escherichia coli* O157:H7 infections: insights into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiol Infect* **122**: 185-192

Twort, F. W., (1915) An investigation on the nature of ultra-microscopic viruses. *Lancet* **2**: 1241-1243

Tye, B. K., J. A. Huberman & D. Botstein, (1974) Non-random circular permutation of phage P22 DNA. *J Mol Biol* **85**: 501-528

Tyler, J. S. & D. I. Friedman, (2004) Characterization of a eukaryotic-like tyrosine protein kinase expressed by the Shiga toxin-encoding bacteriophage 933W. *J Bacteriol* **186**: 3472-3479

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm & S. G. Rozen, (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.doi:10.1093/nar/gks596

van Setten, P. A., L. A. Monnens, R. G. Verstraten, L. P. van den Heuvel & V. W. van Hinsbergh, (1996) Effects of verocytotoxin-1 on nonadherent human monocytes: binding characteristics, protein synthesis, and induction of cytokine release. *Blood* **88**: 174-183

Verma, A., F J. Bolton, D. Fiefield, P. Lamb, E. Woloschin, N. Smith & R. McCann, (2007) An outbreak of *E. coli* O157 associated with a swimming pool: an unusual vehicle of transmission. *Epidemiol Infect* **135**: 989-992

Vica Pacheco, S., O. Garcia Gonzalez & G. L. Paniagua Contreras, (1997) The *lom* gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. *FEMS Microbiol Lett* **156**: 129-132

Wagner, L. A., R. B. Weiss, R. Driscoll, D. S. Dunn & R. F. Gesteland, (1990) Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucleic Acids Res* **18**: 3529-3535

Wagner, P. L., D. W. Acheson & M. K. Waldor, (2001) Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic *Escherichia coli*. *Infect Immun* **69**: 1934-1937

Wang, X., Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino & T. K. Wood, (2010) Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**.doi:10.1038/ncomms1146

Warren, R. A., (1980) Modified bases in bacteriophage DNAs. *Annu Rev Microbiol* **34**: 137-158

Watarai, M., T. Sato, M. Kobayashi, T. Shimizu, S. Yamasaki, T. Tobe, C. Sasakawa & Y. Takeda, (1998) Identification and characterization of a newly isolated Shiga toxin 2-converting phage from Shiga toxin-producing *Escherichia coli*. *Infect Immun* **66**: 4100-4107

Weigel, C. & H. Seitz, (2006) Bacteriophage replication modules. *FEMS Microbiol Rev* **30**: 321-381

Weigele, P. R., E. Scanlon & J. King, (2003) Homotrimeric, β-stranded viral adhesins and tail proteins. *J Bacteriol* **185**: 4022-4030

Wendel, A. M., D. H. Johnson, U. Sharapov, J. Grant, J. R. Archer, T. Monson, C. Koschmann & J. P. Davis, (2009) Multistate outbreak of *Escherichia coli*

O157:H7 infection associated with consumption of packaged spinach, August-September 2006: the Wisconsin investigation. *Clin Infect Dis* **48**: 1079-1086

Werber, D., B. W. Mason, M. R. Evans & R. L. Salmon, (2008) Preventing household transmission of Shiga toxin–producing *Escherichia coli* O157 infection: promptly separating siblings might be the key. *Clin Infect Dis* **46**: 1189-1196

Werts, C., V. Michel, M. Hofnung & A. Charbit, (1994) Adsorption of bacteriophage lambda on the LamB protein of *Escherichia coli* K-12: point mutations in gene J of lambda responsible for extended host range. *J Bacteriol* **176**: 941-947

Whitney, B. M., C. Mainero, E. Humes, S. Hurd, L. Niccolai & J. L. Hadler, (2015) Socioeconomic status and foodborne pathogens in Connecticut, USA, 2000-2011. *Emerg Infect Dis* **21**: 1617-1624

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

Wickner, S., (1984a) DNA-dependent ATPase activity associated with phage P22 gene 12 protein. *J Biol Chem* **259**: 14038-14043

Wickner, S., (1984b) Oligonucleotide synthesis by *Escherichia coli dnaG* primase in conjunction with phage P22 gene 12 protein. *J Biol Chem* **259**: 14044-14047

Willshaw, G. A., H. R. Smith, S. M. Scotland, A. M. Field & B. Rowe, (1987) Heterogeneity of *Escherichia coli* phages encoding Vero cytotoxins: comparison of cloned sequences determining VT1 and VT2 and development of specific gene probes. *J Gen Microbiol* **133**: 1309-1317

Wong, A. R., J. S. Pearson, M. D. Bright, D. Munera, K. S. Robinson, S. F. Lee, G. Frankel & E. L. Hartland, (2011) Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol Microbiol* **80**: 1420-1438

Xiang, Y., M. C. Morais, D. N. Cohen, V. D. Bowman, D. L. Anderson & M. G. Rossmann, (2008) Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage φ29 tail. *Proc Natl Acad Sci USA* **105**: 9552-9557

Xiong, Y., P. Wang, R. Lan, C. Ye, H. Wang, J. Ren, H. Jing, Y. Wang, Z. Zhou, X. Bai, Z. Cui, X. Luo, A. Zhao, Y. Wang, S. Zhang, H. Sun, L. Wang & J. Xu, (2012) A novel *Escherichia coli* O157:H7 clone causing a major hemolytic uremic syndrome outbreak in China. *PLoS One* **7**: e36144.doi:10.1371/journal.pone.0036144

Xu, J., R. W. Hendrix & R. L. Duda, (2004) Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell* **16**: 11-21

Xu, S.-y. & Y. K. Gupta, (2013) Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease. *Nucleic Acids Res* **41**: 378-390

Xu, X., S. P. McAteer, J. J. Tree, D. J. Shaw, E. B. Wolfson, S. A. Beatson, A. J. Roe, L. J. Allison, M. E. Chase-Topping, A. Mahajan, R. Tozzoli, M. E. Woolhouse, S. Morabito & D. L. Gally, (2012) Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic *Escherichia coli*. *PLoS Pathog* **8**: e1002672.doi:10.1371/journal.ppat.1002672.

Xue, Q. & J. B. Egan, (1995) DNA sequence of tail fiber genes of coliphage 186 and evidence for a common ancestor shared by dsDNA phage fiber genes. *Virology* **212**: 128-133

Yamaguchi, Y., J. H. Park & M. Inouye, (2011) Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet* **45**: 61-79

Yamamoto, T., S. Kojio, I. Taneike, S. Nakagawa, N. Iwakura & N. Wakisaka-Saito, (2003) ^{60}Co irradiation of Shiga toxin (Stx)-producing *Escherichia coli* induces Stx phage. *FEMS Microbiol Lett* **222**: 115-121

Yang, F., J. Yang, X. Zhang, L. Chen, Y. Jiang, Y. Yan, X. Tang, J. Wang, Z. Xiong, J. Dong, Y. Xue, Y. Zhu, X. Xu, L. Sun, S. Chen, H. Nie, J. Peng, J. Xu, Y. Wang, Z. Yuan, Y. Wen, Z. Yao, Y. Shen, B. Qiang, Y. Hou, J. Yu & Q. Jin, (2005) Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res* **33**: 6445-6458

Yang, Z., J. Kovar, J. Kim, J. Nietfeldt, D. R. Smith, R. A. Moxley, M. E. Olson, P. D. Fey & A. K. Benson, (2004) Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. *Appl Environ Microbiol* **70**: 6846-6854

Yokoyama, K., K. Makino, Y. Kubota, M. Watanabe, S. Kimura, C. H. Yutsudo, K. Kurokawa, K. Ishii, M. Hattori, I. Tatsuno, H. Abe, M. Yoh, T. Iida, M. Ohnishi, T. Hayashi, T. Yasunaga, T. Honda, C. Sasakawa & H. Shinagawa, (2000) Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. *Gene* **258**: 127-139

Young, R., (1992) Bacteriophage lysis: mechanism and regulation. *Microb Rev* **56**: 430-481

Yue, W. F., M. Du & M. J. Zhu, (2012) High temperature in combination with UV irradiation enhances horizontal transfer of stx_2 gene from *E. col*i O157:H7 to non-pathogenic *E. coli*. *PLoS One* **7**: e31308.doi:10.1371/journal.pone.0031308

Zaneveld, J. R., D. R. Nemergut & R. Knight, (2008) Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* **154**: 1-15.10.1099/mic.0.2007/011833-0

Zhang, X., Y. Cheng, Y. Xiong, C. Ye, H. Zheng, H. Sun, H. Zhao, Z. Ren & J. Xu, (2012) Enterohemorrhagic *Escherichia coli* specific enterohemolysin induced IL-1β in human macrophages and EHEC-induced IL-1β required activation of NLRP3 inflammasome. *PLoS ONE* **7**: e50288.doi:10.1371/journal.pone.0050288

Zhao, S. & K. P. Williams, (2002) Integrative genetic element that reverses the usual target gene orientation. *J Bacteriol* **184**: 859-860

Zhu, C., T. S. Agin, S. J. Elliott, L. A. Johnson, T. E. Thate, J. B. Kaper & E. C. Boedeker, (2001) Complete nucleotide sequence and analysis of the locus of enterocyte Effacement from rabbit diarrheagenic *Escherichia coli* RDEC-1. *Infect Immun* **69**: 2107-2115.10.1128/iai.69.4.2107-2115.2001

Zimmer, M., E. Sattelberger, R. B. Inman, R. Calendar & M. J. Loessner, (2003) Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed+1 translational frameshifting in structural protein synthesis. *Mol Microbiol* **50**: 303-317

Appendices

Appendix I: Functional annotation of AU5Stx1

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
AU5STX1_010	*int*	196	1365	1170	Integrase	389
AU5STX1_020	*xis*	1349	1531	183	Excisionase	60
AU5STX1_030		1592	1843	252	Hypothetical protein	83
AU5STX1_040		1831	2064	234	Hypothetical protein	77
AU5STX1_050		2208	2597	390	Hypothetical protein	129
AU5STX1_060		2633	2845	213	Hypothetical protein	70
AU5STX1_070		2805	3431	627	Adenine methylase	208
AU5STX1_080		3428	3859	432	Hypothetical protein	143
AU5STX1_090		3915	4553	639	Phage anti-repressor protein	212
AU5STX1_100		4818	4916	99	Hypothetical protein	32
AU5STX1_110		4909	5805	897	Hypothetical Protein	298
AU5STX1_120		5802	6317	516	Hypothetical protein	171
AU5STX1_130		6319	7266	948	Hypothetical protein	315
AU5STX1_140		7263	7484	222	Hypothetical protein	73
AU5STX1_150		7583	7864	282	Hypothetical protein	93
AU5STX1_160		7875	8066	192	Hypothetical protein	63
AU5STX1_170		8039	8227	189	Hypothetical protein	62
AU5STX1_180	*exo*	8218	8898	681	Exonuclease	226
AU5STX1_190	*bet*	8895	9680	786	Recombination protein bet	261
AU5STX1_200	*gam*	9686	9982	297	Host-nulcease inhibitor protein gam	98
AU5STX1_210	*kil*	10059	10327	269	Protein kil	89
AU5STX1_220	*cIII*	10171	10334	164	Protease inhibitor III	54
AU5STX1_230		10407	10775	369	Hypothetical Protein	122
AU5STX1_240		10926	11396	471	Hypothetical protein	156
AU5STX1_250	*N*	11455	11838	384	Antitermination protein N	127
AU5STX1_260		12254	12397	144	Hypothetical protein	47
AU5STX1_270	*cI*	12728	13432	705	Repressor protein CI	234
AU5STX1_280	*cro*	13546	13779	234	Regulatory protein Cro	77
AU5STX1_290	*cII*	13918	14214	297	Transcription activator CII	98
AU5STX1_300	*o*	14247	15185	939	Replication protein O	312
AU5STX1_310	*p*	15182	15883	702	Replication protein P	233
AU5STX1_320		15880	16170	291	Protein ren	96
AU5STX1_330		16167	16244	78	Putative membrane protein	25
AU5STX1_340		16241	16519	279	Hypothetical protein	92
AU5STX1_350		16652	16867	216	Hypothetical protein	71
AU5STX1_360		16878	17114	237	Restriction alleviation protein	78
AU5STX1_370	*ninB*	17071	17518	448	Recombinase NinB	148
AU5STX1_380		17514	18041	528	Adenine methyltransferase	175
AU5STX1_390	*ninE*	18038	18232	195	Protein NinE	64
AU5STX1_400		18229	18298	70	Hypothetical protein	22

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
AU5STX1_410		18489	19163	675	Phage anti-repressor protein	224
AU5STX1_420		19238	19489	252	Hypothetical protein	83
AU5STX1_430	*ninF*	19630	19806	177	Protein NinF	58
AU5STX1_440		19799	20521	723	Phage Regulatory Protein	240
AU5STX1_450	*ninG*	20521	21126	606	Protein NinG	201
AU5STX1_460	*ninH*	21123	21317	195	Protein NinH	64
AU5STX1_470	*q*	21310	21744	435	Antitermination protein Q	144
AU5STX1_480	*stxA1*	22251	23198	948	Shiga toxin I subunit A	315
AU5STX1_490	*stxB1*	23208	23477	270	Shiga toxin I subunit B	89
AU5STX1_500		23981	25921	1941	Acetylesterase	646
AU5STX1_510		26058	26237	180	Putative membrane protein	59
AU5STX1_520		26278	26550	273	Hypothetical protein	90
AU5STX1_530	*s*	26627	26842	216	Holin	71
AU5STX1_540	*r*	26847	27380	534	Endolysin	177
AU5STX1_550		27695	28237	543	Phage anti-repressor protein	180
AU5STX1_560		28252	28386	135	Putative membrane protein	44
AU5STX1_570	*rz*	28394	28858	465	Spanin, inner membrane subunit	154
AU5STX1_580	*rz1*	28614	28800	187	Spanin, outer membrane subunit	61
AU5STX1_590	*bor*	28890	29183	294	Lipoprotein bor	97
AU5STX1_600		29291	29536	246	Hypothetical protein	81
AU5STX1_610	*terS*	29592	30398	807	Small subunit terminase	268
AU5STX1_620	*terL*	30379	32085	1707	Large subunit terminase	568
AU5STX1_630	*porT*	32085	34229	2145	Portal protein	714
AU5STX1_640		34387	35394	1008	Hypothetical protein	335
AU5STX1_650		35418	36632	1215	Capsid protein	404
AU5STX1_660		36688	37077	390	Capsid protein	129
AU5STX1_670		37100	37588	489	Hypothetical protein	162
AU5STX1_680		37572	38135	564	Hypothetical protein	187
AU5STX1_690		38135	38785	651	Hypothetical protein	216
AU5STX1_700	*tfpS*	38782	40677	1896	Tail fibre of Podoviral Stx phages	631
AU5STX1_710	*atfT*	40559	40948	390	Alternate tail fibre tip	129
AU5STX1_720		41036	41278	243	Hypothetical protein	80
AU5STX1_730		41339	41458	120	Hypothetical protein	39
AU5STX1_740		41495	43198	1704	Hypothetical protein	567
AU5STX1_750		43195	44463	1269	Tail tip protein	422
AU5STX1_760		44478	44756	279	Hypothetical protein	92
AU5STX1_770		44762	45379	618	Hypothetical protein	205
AU5STX1_780	*lom*	45470	46204	735	Outer membrane protein Lom	244
AU5STX1_790		46437	46577	141	Prophage host toxic membrane protein	46
AU5STX1_800		46634	47035	402	Hypothetical protein	133
AU5STX1_810		47129	47785	657	Hypothetical protein	218
AU5STX1_820		47788	48234	447	Hypothetical protein	148
AU5STX1_830		48244	48495	252	Bacteriocin	83
AU5STX1_840		48506	49771	1266	Hypothetical protein	421

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
AU5STX1_850		49802	58225	8424	Hypothetical protein	2807

Appendix II: Functional annotation of AU6Stx1

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
AU6STX1_010	*int*	196	1365	1170	Integrase	389
AU6STX1_020	*xis*	1349	1531	183	Excisionase	60
AU6STX1_030		1592	1843	252	Hypothetical protein	83
AU6STX1_040		1896	2255	360	Truncated transposase (Pseudogene)	119
AU6STX1_050		2295	3240	946	Hypothetical protein (Pseudogene)	314
AU6STX1_060		3237	3458	222	Hypothetical protein	73
AU6STX1_070		3556	3837	282	Hypothetical protein	93
AU6STX1_080		3848	4039	192	Hypothetical protein	63
AU6STX1_090		4013	4200	188	Hypothetical protein	62
AU6STX1_100	*exo*	4191	4871	681	Exonuclease	226
AU6STX1_110	*bet*	4868	5653	786	Recombination protein bet	261
AU6STX1_120	*gam*	5659	5955	297	Host-nulcease inhibitor protein gam	98
AU6STX1_130	*kil*	6031	6300	270	Protein kil	89
AU6STX1_140	*cIII*	6143	6307	165	Protease inhibitor III	54
AU6STX1_150		6380	6748	369	Hypothetical protein	122
AU6STX1_160		6899	7369	471	Hypothetical protein	156
AU6STX1_170	*N*	7428	7811	384	Anti-termination protein N	127
AU6STX1_180		8300	8464	165	Hypothetical protein	54
AU6STX1_190	*stk*	8470	9513	1044	Serine/threonine-protein kinase	347
AU6STX1_200		9507	9968	462	Hypothetical protein	153
AU6STX1_210		10036	10377	342	Hypothetical protein	113
AU6STX1_220	*cI*	10438	11145	708	Repressor protein CI	235
AU6STX1_230	*cro*	11224	11451	228	Regulatory protein Cro	75
AU6STX1_240	*cII*	11590	11886	297	Transcription activator CII	98
AU6STX1_250	*o*	11919	12854	936	Replication protein O	311
AU6STX1_260	*p*	12854	13555	702	Replication protein P	233
AU6STX1_270		13552	13842	291	Protein ren	96
AU6STX1_280		13839	13917	79	Putative membrane protein	25
AU6STX1_290		13913	14191	279	Hypothetical protein	92
AU6STX1_300		14324	14539	216	Hypothetical protein	71
AU6STX1_310		14550	14786	237	Restriction alleviation protein	78
AU6STX1_320	*ninB*	14743	15190	448	Recombinase NinB	148
AU6STX1_330		15186	15713	528	Adenine methyltransferase	175
AU6STX1_340	*ninE*	15710	15893	184	Protein NinE	60
AU6STX1_350		16396	18168	1773	Hypothetical protein	590
AU6STX1_360		18687	19292	606	Putative endonuclease	201
AU6STX1_370	*ninG*	19267	19878	612	Protein NinG	203
AU6STX1_380	*ninH*	19875	20069	195	Protein NinH	64
AU6STX1_390	*q*	20062	20496	435	Anti-termination protein Q	144
AU6STX1_400	*stxA1*	21003	21950	948	Shiga toxin I subunit A	315
AU6STX1_410	*stxB1*	21960	22229	270	Shiga toxin I subunit B	89
AU6STX1_420		22733	24673	1941	Acetylesterase	646

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
AU6STX1_430		24810	24989	180	Putative membrane protein	59
AU6STX1_440		25030	25302	273	Hypothetical protein	90
AU6STX1_450	*s*	25379	25594	216	Holin	71
AU6STX1_460	*r*	25599	26132	534	Endolysin	177
AU6STX1_470		26447	26989	543	Phage anti-repressor protein	180
AU6STX1_480		27004	27138	135	Hypothetical protein	44
AU6STX1_490	*rz*	27146	27610	465	Spanin, inner membrane subunit	154
AU6STX1_500	*rz1*	27366	27551	186	Spanin, outer membrane subunit	61
AU6STX1_510	*bor*	27642	27935	294	Lipoprotein Bor	97
AU6STX1_520		28043	28288	246	Hypothetical protein	81
AU6STX1_530	*terS*	28344	29150	807	Small subunit terminase	268
AU6STX1_540	*terL*	29131	30837	1707	Large subunit terminase	568
AU6STX1_550	*porT*	30837	32981	2145	Portal protein	714
AU6STX1_560		33139	34146	1008	Hypothetical protein	335
AU6STX1_570		34170	35384	1215	Capsid protein	404
AU6STX1_580		35440	35829	390	Capsid protein	129
AU6STX1_590		35852	36340	489	Hypothetical protein	162
AU6STX1_600		36324	36887	564	Hypothetical protein	187
AU6STX1_610		36887	37537	651	Hypothetical protein	216
AU6STX1_620	*tfpS*	37534	39429	1896	Tail fibre of Podoviral Stx phages	631
AU6STX1_630	*atfT*	39311	39697	387	Alternate tail fibre tip	128
AU6STX1_640		39788	40030	243	Hypothetical protein	80
AU6STX1_650		40091	40210	120	Hypothetical protein	39
AU6STX1_660		40247	41950	1704	Hypothetical protein	567
AU6STX1_670		41947	43215	1269	Tail tip protein	422
AU6STX1_680		43230	43508	279	Hypothetical protein	92
AU6STX1_690		43514	44131	618	Hypothetical protein	205
AU6STX1_700	*lom*	44222	44956	735	Outer membrane protein Lom	244
AU6STX1_710		45189	45329	141	Prophage host toxic membrane protein	46
AU6STX1_720		45386	45787	402	Hypothetical protein	133
AU6STX1_730		45881	46537	657	Hypothetical protein	218
AU6STX1_740		46540	46986	447	Hypothetical protein	148
AU6STX1_750		46996	47247	252	Bacteriocin	83
AU6STX1_760		47258	48523	1266	Hypothetical protein	421
AU6STX1_770		48554	56974	8421	Hypothetical protein	2806

Appendix III: Functional annotation of pp14Stx1

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
14STX1_010	*int*	195	1364	1170	Integrase	389
14STX1_020	*xis*	1348	1530	183	Excisionase	60
14STX1_030		1591	1842	252	Hypothetical protein	83
14STX1_040		1895	2254	360	Transposase (Pseudogene)	119
14STX1_050		2294	3239	946	hypothetical protein	314
14STX1_060		3236	3457	222	Hypothetical protein	73
14STX1_070		3556	3837	282	Hypothetical protein	93
14STX1_080		3848	4039	192	Hypothetical protein	63
14STX1_090		4012	4200	189	Hypothetical protein	62
14STX1_100	*exo*	4191	4871	681	Exonuclease	226
14STX1_110	*bet*	4868	5653	786	Recombination protein bet	261
14STX1_120	*gam*	5659	5955	297	Host-nulcease inhibitor protein gam	98
14STX1_130	*kil*	6031	6300	270	Protein kil	89
14STX1_140	*cIII*	6143	6307	165	Protease inhibitor III	54
14STX1_150		6380	6748	369	Hypothetical protein	122
14STX1_160		6899	7369	471	Hypothetical protein	156
14STX1_170	*n*	7428	7811	384	Antitermination protein N	127
14STX1_180		8300	8464	165	Hypothetical protein	54
14STX1_190	*stk*	8467	9513	1047	Serine/threonine-protein kinase	348
14STX1_200		9507	9968	462	Hypothetical protein	153
14STX1_210		10035	10376	342	Hypothetical protein	113
14STX1_220	*cI*	10437	11144	708	Repressor protein CI	235
14STX1_230	*cro*	11223	11450	228	Regulatory protein Cro	75
14STX1_240	*cII*	11589	11885	297	Transcription activator CII	98
14STX1_250	*o*	11918	12856	939	Replication protein O	312
14STX1_260	*p*	12853	13554	702	Replication protein P	233
14STX1_270		13551	13841	291	Protein ren	96
14STX1_280		13838	13915	78	Putative membrane protein	25
14STX1_290		13912	14190	279	Hypothetical protein	92
14STX1_300		14323	14535	213	Hypothetical protein	70
14STX1_310		14549	14785	237	Restriction alleviation protein	78
14STX1_320	*ninB*	14742	15188	447	Recombinase NinB	148
14STX1_330		15185	15712	528	Phage adenine methyltransferase	175
14STX1_340	*ninE*	15709	15891	183	Protein NinE	60
14STX1_350		16395	18167	1773	Hypothetical protein	590
14STX1_360		18698	19303	606	Putative endonuclease	201
14STX1_370	*ninG*	19278	19889	612	Protein NinG	203
14STX1_380	*ninH*	19886	20080	195	Protein NinH	64
14STX1_390	*q*	20073	20507	435	Antitermination protein Q	144
14STX1_400	*stxA2*	21014	21961	948	Shiga toxin I subunit A	315
14STX1_410	*stxB2*	21971	22240	270	Shiga toxin I subunit B	89
14STX1_420		22744	24684	1941	Acetylesterase	646

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
14STX1_430		24821	25000	180	Hypothetical protein	59
14STX1_440		25041	25313	273	Hypothetical protein	90
14STX1_450	*s*	25390	25605	216	Holin	71
14STX1_460	*r*	25610	26143	534	Endolysin	177
14STX1_470		26458	27000	543	Phage anti-repressor protein	180
14STX1_480		27015	27149	135	putative membrane protein	44
14STX1_490	*rz*	27157	27621	465	Spanin, inner membrane subunit	154
14STX1_500	*rz2*	27377	27562	186	Spanin, outer membrane subunit	61
14STX1_510	*bor*	27653	27946	294	Lipoprotein bor	97
14STX1_520		28054	28299	246	Hypothetical protein	81
14STX1_530	*terS*	28355	29161	807	Small subunit terminase	268
14STX1_540	*terL*	29142	30848	1707	Large subunit terminase	568
14STX1_550	*porT*	30848	32992	2145	Portal protein	714
14STX1_560		33150	34157	1008	Hypothetical protein	335
14STX1_570		34181	35395	1215	Capsid protein	404
14STX1_580		35451	35840	390	Capsid protein	129
14STX1_590		35890	36351	462	Hypothetical protein	153
14STX1_600		36335	36898	564	Hypothetical protein	187
14STX1_610		36898	37548	651	Hypothetical protein	216
14STX1_620	*tfpS*	37545	39440	1896	Tail fibre of Podoviral Stx phages	631
14STX1_630	*atfT*	39322	39711	390	Alternate tail fibre tip	129
14STX1_640		39796	40992	1197	IS*ec20* transposase	398
14STX1_650		41250	41492	243	Hypothetical protein	80
14STX1_660		41559	41672	114	Hypothetical protein	37
14STX1_670		41712	43412	1701	Hypothetical protein	566
14STX1_680		43409	44677	1269	Tail tip protein	422
14STX1_690		44692	44970	279	Hypothetical protein	92
14STX1_700		44976	45593	618	Hypothetical protein	205
14STX1_710	*lom*	45684	46418	735	Outermembrane protein Lom	244
14STX1_720		46651	46791	141	Prophage host toxic membrane protein	46
14STX1_730		46848	47249	402	Hypothetical protein	133
14STX1_740		47343	47999	657	Hypothetical protein	218
14STX1_750		48002	48448	447	Hypothetical protein	148
14STX1_760		48458	48709	252	Hypothetical protein	83
14STX1_770		48720	49985	1266	Hypothetical protein	421
14STX1_780		50016	58439	8424	Hypothetical protein (Pseudogene)	2807

Appendix IV: Functional annotation of pp571Stx1

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
571STX1_010	*int*	196	1365	1170	Integrase	389
571STX1_020	*xis*	1349	1531	183	Excisionase	60
571STX1_030		1592	1843	252	Hypothetical protein	83
571STX1_040		1899	2255	357	Truncated transposase (Pseudogene)	118
571STX1_050		2298	2642	345	Replication protein O (Pseudogene)	114
571STX1_060	*p*	2639	3340	702	Replication protein P	233
571STX1_070		3337	3627	291	Protein ren	96
571STX1_080		3624	3707	84	Hypothetical protein	27
571STX1_090		3742	3906	165	Hypothetical protein	54
571STX1_100	*ninB*	3914	4360	447	Recombinase NinB	148
571STX1_110		4357	4881	525	Phage adenine methyltransferase	174
571STX1_120	*ninE*	4878	5060	183	Protein NinE	60
571STX1_130		5564	7336	1773	Hypothetical protein	590
571STX1_140		7867	8472	606	Putative HNH endonuclease	201
571STX1_150	*ninG*	8447	9058	612	Protein NinG	203
571STX1_160	*ninH*	9055	9249	195	NinH protein	64
571STX1_170	*q*	9242	9676	435	Antitermination protein Q	144
571STX1_180	*stxA2*	10183	11130	948	Shiga toxin I subunit A	315
571STX1_190	*stxB2*	11140	11409	270	Shiga toxin I subunit B	89
571STX1_200		11913	13853	1941	Acetylesterase	646
571STX1_210		13990	14169	180	Hypothetical protein	59
571STX1_220		14210	14455	246	Hypothetical protein	81
571STX1_230	*s*	14533	14748	216	Holin	71
571STX1_240	*r*	14753	15286	534	Endolysin	177
571STX1_250		15601	16143	543	Phage anti-repressor protein	180
571STX1_260		16158	16292	135	putative membrane protein	44
571STX1_270	*rz*	16300	16764	465	Spanin, inner membrane subunit	154
571STX1_280	*rz2*	16520	16705	186	Spanin, outer membrane subunit	61
571STX1_290	*bor*	16796	17089	294	Lipoprotein bor	97
571STX1_300		17197	17442	246	Hypothetical protein	81
571STX1_310	*terS*	17498	18304	807	Small subunit terminase	268
571STX1_320	*terL*	18285	19991	1707	Large subunit terminase	568
571STX1_330	*porT*	19991	22135	2145	Portal protein	714
571STX1_340		22293	23300	1008	Hypothetical protein	335
571STX1_350		23324	24538	1215	Capsid protein	404
571STX1_360		24594	24983	390	Capsid protein	129
571STX1_370		25033	25494	462	Hypothetical protein	153
571STX1_380		25478	26041	564	Hypothetical protein	187
571STX1_390		26041	26691	651	Hypothetical protein	216
571STX1_400	*tfpS*	26688	28583	1896	Tail fibre of Podoviral Stx phages	631
571STX1_410	*atfT*	28465	28854	390	Alternate tail fibre tip	129
571STX1_420		28939	30135	1197	IS*ec20* transposase	398

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
571STX1_430		30393	30635	243	Hypothetical protein	80
571STX1_440		30696	30815	120	Hypothetical protein	39
571STX1_450		30855	32555	1701	Hypothetical protein	566
571STX1_460		32552	33820	1269	Tail tip protein	422
571STX1_470		33835	34113	279	Hypothetical protein	92
571STX1_480		34119	34736	618	Hypothetical protein	205
571STX1_490	*lom*	34827	35561	735	Outermembrane protein Lom	244
571STX1_500		35794	35934	141	Prophage host toxic membrane protein	46
571STX1_510		35991	36392	402	Hypothetical protein	133
571STX1_520		36486	37142	657	Hypothetical protein	218
571STX1_530		37145	37591	447	Hypothetical protein	148
571STX1_540		37601	37852	252	Hypothetical protein	83
571STX1_550		37863	39128	1266	Hypothetical protein	421
571STX1_560		39159	47582	8424	Hypothetical protein	2807

Appendix V: Functional annotation of pp14Stx2c

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
14STX2C_010	*int*	215	1393	1179	Integrase	392
14STX2C_020		1374	1565	192	Hypothetical protein	63
14STX2C_030		1647	1991	345	Hypothetical protein	114
14STX2C_040		2179	2529	351	Hypothetical protein	116
14STX2C_050		2526	2882	357	Hypothetical protein	118
14STX2C_060		2879	3394	516	Hypothetical protein	171
14STX2C_070		3396	4343	948	Hypothetical protein	315
14STX2C_080		4660	4941	282	Hypothetical protein	93
14STX2C_090		4952	5143	192	Hypothetical protein	63
14STX2C_100		5310	6197	888	IS*629* Transposase	295
14STX2C_110		6197	6523	327	IS*629* Transposase	108
14STX2C_120	*exo*	6608	7288	681	Exonuclease	226
14STX2C_130	*bet*	7285	8070	786	Recombination protein Bet	261
14STX2C_140	*gam*	8076	8492	417	Host-nuclease inhibitor protein gam	138
14STX2C_150	*kil*	8447	8716	270	Protein kil	89
14STX2C_160	*cIII*	8559	8723	165	Protease inhibitor III	54
14STX2C_170		8796	9164	369	Hypothetical protein	122
14STX2C_180		9347	9598	252	Hypothetical protein	83
14STX2C_190	*n*	9657	10001	345	Anti-termination protein N	114
14STX2C_200		10658	11179	522	Hypothetical protein	173
14STX2C_210		11402	11536	135	Hypothetical protein	44
14STX2C_220	*cI*	11681	12340	660	Repressor protein CI	219
14STX2C_230	*cro*	12452	12667	216	Regulatory protein Cro	71
14STX2C_240	*cII*	12809	13105	297	Regulatory protein CII	98
14STX2C_250		13138	13299	162	Hypothetical protein	53
14STX2C_260	*o*	13286	14107	822	Replication protein O	273
14STX2C_270	*dnaB*	14104	15480	1377	Helicase	458
14STX2C_280		15497	15829	333	Hypothetical protein	110
14STX2C_290		15999	16424	426	Restriction alleviation protein	141
14STX2C_300	*ninB*	16381	16827	447	Recombinase NinB	148
14STX2C_310		16824	17351	528	DNA N-6 adenine methyltransferase	175
14STX2C_320	*ninE*	17348	17530	183	Protein NinE	60
14STX2C_330		17805	18563	759	Phage antirepressor	252
14STX2C_340		18819	19499	681	Phage antirepressor	226
14STX2C_350		19574	20296	723	Phage regulatory protein	240
14STX2C_360	*ninG*	20296	20901	606	Protein NinG	201
14STX2C_370		20898	21569	672	Serine/threonine-protein phosphatase	223
14STX2C_380	*q*	21560	22048	489	Antitermination protein Q	162
14STX2C_390		22698	23657	960	Shiga toxin 2c A-subunit	319
14STX2C_400		23669	23938	270	Shiga toxin 2c B-subunit	89
14STX2C_410		24095	24982	888	IS*629* Transposase	295
14STX2C_420		24982	25308	327	IS*629* Transposase	108

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
14STX2C_430	*iraM*	25548	25871	324	Anti-adapter protein IraM	107
14STX2C_440		26115	28052	1938	Acetylesterase	645
14STX2C_450		28190	28369	180	Hypothetical protein	59
14STX2C_460		28410	28682	273	Hypothetical protein	90
14STX2C_470	*s*	28759	28974	216	Lysis protein S	71
14STX2C_480	*lyzP*	28974	29471	498	Lysozyme	165
14STX2C_490	*rz*	29468	29905	438	Spanin, inner membrane subunit	145
14STX2C_500	*rz2*	29667	29852	186	Spannin, outer membrane protein	61
14STX2C_510		30108	30605	498	DNA binding protein	165
14STX2C_520		30602	30859	258	Hypothetical protein	85
14STX2C_530		31322	31549	228	Hypothetical protein	75
14STX2C_540		31591	31956	366	Endonuclease	121
14STX2C_550	*terS*	32249	32812	564	Terminase, small subunit	187
14STX2C_560	*terL*	32809	34470	1662	Terminase, large subunit	553
14STX2C_570		34534	36470	1937	Major capsid protein	645
14STX2C_580		36683	39269	-	Portal protein (Pseudogene)	-
14STX2C_590		39266	39592	327	Head-tail connector protein	108
14STX2C_600		39602	39952	351	Head-tail adaptor	116
14STX2C_610		39949	40395	447	Phage tail protein	148
14STX2C_620		40392	40736	345	Hypothetical protein	114
14STX2C_630		40705	41511	807	Phage tail protein	268
14STX2C_640		41517	41891	375	Tail protein	124
14STX2C_650		41888	42196	309	Tail protein	102
14STX2C_660		42249	45491	3243	Tape measure protein	1080
14STX2C_670		45484	45825	342	Minor tail protein	113
14STX2C_680		45825	46523	699	Minor tail protein	232
14STX2C_690	*mnt*	46540	46860	321	Regulatory protien Mnt	106
14STX2C_700	*arc*	46968	47141	174	Transcriptional repressor Arc	57
14STX2C_710	*ant*	47212	48135	924	Antirepressor protein	307
14STX2C_720		48190	48927	738	Tail assembly protein	245
14STX2C_730		48825	49502	678	Tail assembly protein	225
14STX2C_740		49743	53225	3483	Host specificity protein	1160
14STX2C_750		53292	53891	600	Outer membrane protein Lom	199
14STX2C_760	*tfsS*	53956	55269	1314	Tail fibre of Siphoviral Stx phages	437
14STX2C_770	*atfT*	55232	55540	309	Alternate tail fibre tip	102
14STX2C_780	*ospB*	55681	56556	876	T3SS secreted effector OspB-like protein	291
14STX2C_790	*nleG*	56781	57431	651	Non-LEE-encoded type III effector, NleG	216
14STX2C_800		57416	57700	285	Hypothetical protein	94
14STX2C_810		58151	58453	303	Hypothetical protein	100

Appendix VI: Functional annotation of pp571Stx2c

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
571STX2C_010	*int*	215	1393	1179	Integrase	392
571STX2C_020		1377	1565	189	Hypothetical protein	62
571STX2C_030		1647	1991	345	Hypothetical protein	114
571STX2C_040		2179	2529	351	Hypothetical protein	116
571STX2C_050		2526	2882	357	Hypothetical protein	118
571STX2C_060		2879	3394	516	Hypothetical protein	171
571STX2C_070		3396	4343	948	Hypothetical protein	315
571STX2C_080		4343	5874	-	Phage regulatory protein (Pseudogene)	-
571STX2C_090		4498	4824	327	Transposase	108
571STX2C_100		4824	5711	888	Transposase	295
571STX2C_110		5976	6254	279	Hypothetical protein	92
571STX2C_120		6268	6456	189	Hypothetical protein	62
571STX2C_130		6432	6611	180	Hypothetical protein	59
571STX2C_140	*exo*	6608	7288	681	Exonuclease	226
571STX2C_150	*bet*	7285	8070	786	Recombination protein Bet	261
571STX2C_160	*gam*	8076	8492	417	Host-nuclease inhibitor protein gam	138
571STX2C_170	*kil*	8450	8716	267	Protein kil	88
571STX2C_180	*cIII*	8562	8723	162	Protease inhibitor III	53
571STX2C_190		8796	9164	369	Hypothetical protein	122
571STX2C_200		9350	9598	249	Hypothetical protein	82
571STX2C_210	*n*	9657	10001	345	Anti-termination protein N	114
571STX2C_220		10658	11179	522	Hypothetical protein	173
571STX2C_230		11405	11536	132	Hypothetical protein	43
571STX2C_240	*cI*	11681	12340	660	Repressor protein CI	219
571STX2C_250	*cro*	12452	12667	216	Regulatory protein Cro	71
571STX2C_260	*cII*	12809	13102	294	Regulatory protein CII	97
571STX2C_270		13138	13296	159	Hypothetical protein	52
571STX2C_280	*o*	13286	14107	822	Replication protein O	273
571STX2C_290	*dnaB*	14104	15480	1377	Helicase	458
571STX2C_300		15497	15829	333	Hypothetical protein	110
571STX2C_310		15999	16424	426	Restriction alleviation protein	141
571STX2C_320	*ninB*	16381	16827	447	Recombinase NinB	148
571STX2C_330		16824	17351	528	DNA N-6-adenine-methyltransferase	175
571STX2C_340	*ninE*	17348	17542	195	Protein NinE	64
571STX2C_350		17799	18473	675	Phage antirepressor	224
571STX2C_360		18548	18949	402	Hypothetical protein	133
571STX2C_370		19111	19833	723	Phage regulatory protein	240
571STX2C_380	*ninG*	19833	20438	606	Protein NinG	201
571STX2C_390	*ninH*	20435	20629	195	Protein NinH	64
571STX2C_400	*q*	20583	21056	474	Antitermination protein Q	157
571STX2C_410	*stx2cA*	21839	22798	960	Shiga toxin 2c A-subunit	319
571STX2C_420	*stx2cB*	22810	23076	267	Shiga toxin 2c B-subunit	88

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
571STX2C_430	*iraM*	23376	23699	324	Anti-adapter protein IraM	107
571STX2C_440		23943	25880	1938	Acetylesterase	645
571STX2C_450		26018	26197	180	Hypothetical protein	59
571STX2C_460		26238	26510	273	Hypothetical protein	90
571STX2C_470	*s*	26587	26799	213	Lysis protein S	70
571STX2C_480	*lyzP*	26802	27299	498	Lysozyme	165
571STX2C_490	*rz*	27296	27733	438	Spanin, inner membrane subunit	145
571STX2C_500	*rz1*	27495	27677	183	Spannin, outer membrane protein	60
571STX2C_510		27936	28433	498	DNA binding protein	165
571STX2C_520		28430	28684	255	Hypothetical protein	84
571STX2C_530		29149	29377	229	Hypothetical protein	75
571STX2C_540		29419	29784	366	Endonuclease	121
571STX2C_550	*terS*	30076	30639	564	Terminase, small subunit	187
571STX2C_560	*terL*	30636	32297	1662	Terminase, large subunit	553
571STX2C_570		32361	34298	1938	Major capsid protein	645
571STX2C_580		34510	37094	-	Portal protein (Pseudogene)	-
571STX2C_590		37091	37417	327	Head-tail connector protein	108
571STX2C_600		37427	37777	351	Head-tail adaptor	116
571STX2C_610		37774	38220	447	Phage tail protein	148
571STX2C_620		38217	38561	345	Hypothetical protein	114
571STX2C_630		38620	39336	717	Phage tail protein	238
571STX2C_640		39342	39716	375	Tail protein	124
571STX2C_650		39713	40021	309	Tail protein	102
571STX2C_660		40074	43316	3243	Tape measure protein	1080
571STX2C_670		43309	43650	342	Minor tail protein	113
571STX2C_680		43650	44348	699	Minor tail protein	232
571STX2C_690	*mnt*	44365	44685	321	Regulatory protien Mnt	106
571STX2C_700	*arc*	44793	44963	171	Transcriptional repressor Arc	56
571STX2C_710	*ant*	45037	45960	924	Antirepressor protein	307
571STX2C_720		46015	46752	738	Tail assembly protein	245
571STX2C_730		46650	47330	681	Tail assembly protein	226
571STX2C_740		47568	51050	3483	Host specificity protein	1160
571STX2C_750		51117	51713	597	Outer membrane protein Lom	198
571STX2C_760	*tfsS*	51781	53094	1314	Tail fibre of Siphoviral Stx phages	437
571STX2C_770	*atfT*	53057	53365	309	Alternate tail fibre tip	102
571STX2C_780	*ospB*	53506	54381	876	T3SS secreted effector OspB-like protein	291
571STX2C_790	*nleG*	54606	55256	651	Non-LEE-encoded type III effector, NleG	216
571STX2C_800		55241	55525	285	Hypothetical protein	94
571STX2C_810		55976	56278	303	Hypothetical protein	100

Appendix VII: Functional annotation of pp2441Stx2c

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
2441STX2C_010	*int*	215	1393	1179	Integrase	392
2441STX2C_020		1377	1565	189	Hypothetical protein	62
2441STX2C_030		1647	1991	345	Hypothetical protein	114
2441STX2C_040		2228	2860	633	Antirepressor	210
2441STX2C_050		3184	3912	729	Hypothetical protein	242
2441STX2C_060		4008	4373	366	Hypothetical protein	121
2441STX2C_070		4370	4726	-	Hypothetical protein (Pseudogene)	-
2441STX2C_080		4723	5233	-	Hypothetical protein (Pseudogene)	-
2441STX2C_090		5235	6182	948	Hypothetical protein	315
2441STX2C_100		6179	6400	222	Hypothetical protein	73
2441STX2C_110		6499	6780	282	Hypothetical protein	93
2441STX2C_120		6791	6982	192	Hypothetical protein	63
2441STX2C_130	*exo*	7134	7814	681	Exonuclease	226
2441STX2C_140	*bet*	7811	8596	786	Recombination protein Bet	261
2441STX2C_150		8602	8706	105	Host-nuclease inhibitor protein gam(Pseudogene)	34
2441STX2C_160		8751	9638	888	IS*629* Transposase	295
2441STX2C_170		9638	9964	327	IS*629* Transposase	108
2441STX2C_180		10018	10278	261	Hypothetical protein	86
2441STX2C_190		10461	10712	252	Hypothetical protein	83
2441STX2C_200	*n*	10771	11115	345	Anti-termination protein N	114
2441STX2C_210		11773	12294	522	Hypothetical protein	173
2441STX2C_220		12517	12651	135	Hypothetical protein	44
2441STX2C_230	*cI*	12796	13455	660	Repressor protein CI	219
2441STX2C_240	*cro*	13566	13781	216	Regulatory protein Cro	71
2441STX2C_250	*cII*	13923	14219	297	Regulatory protein CII	98
2441STX2C_260		14252	14413	162	Hypothetical protein	53
2441STX2C_270	*o*	14400	15221	822	Replication protein O	273
2441STX2C_280	*dnaB*	15218	16594	1377	Helicase	458
2441STX2C_290		16673	17284	612	HNH endonuclease	203
2441STX2C_300		17531	17746	216	Hypothetical protein	71
2441STX2C_310		17748	18080	333	Hypothetical protein	110
2441STX2C_320		18077	18487	411	Hypothetical protein	136
2441STX2C_330		18771	18944	174	Hypothetical protein	57
2441STX2C_340		18941	19276	336	Phage antirepressor (Partial)	111
2441STX2C_350		19538	20212	675	Phage antirepressor	224
2441STX2C_360		20281	21011	-	DNA binding protein (Pseudogene)	-
2441STX2C_370	*ninG*	21011	21616	606	Protein NinG	201
2441STX2C_380		21613	22284	672	Serine/threonine-protein phosphatase	223
2441STX2C_390	*q*	22275	22763	489	Antitermination protein Q	162
2441STX2C_400	*stx2cA*	23419	24378	960	Shiga toxin 2c A-subunit	319
2441STX2C_410	*stx2cB*	24390	24659	270	Shiga toxin 2c B-subunit	89

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
2441STX2C_420	*iraM*	24956	25279	324	Anti-adapter protein IraM	107
2441STX2C_430		25523	27460	1938	Acetylesterase	645
2441STX2C_440		27598	27777	180	Hypothetical protein	59
2441STX2C_450		27818	28090	273	Hypothetical protein	90
2441STX2C_460	*s*	28167	28382	216	Lysis protein S	71
2441STX2C_470	*lyzP*	28382	28879	498	Lysozyme	165
2441STX2C_480	*rz*	28876	29313	438	Spanin, inner membrane subunit	145
2441STX2C_490	*rz1*	29075	29260	186	Spannin, outer membrane protein	61
2441STX2C_500		29516	30013	498	DNA binding protein	165
2441STX2C_510		30010	30267	258	Hypothetical protein	85
2441STX2C_520		30730	30957	228	Hypothetical protein	75
2441STX2C_530		30999	31364	366	Endonuclease	121
2441STX2C_540	*terS*	31654	32217	564	Terminase, small subunit	187
2441STX2C_550	*terL*	32214	33875	1662	Terminase, large subunit	553
2441STX2C_560		33939	35876	1938	Major capsid protein	645
2441STX2C_570		36088	38666	-	Portal protein (Pseudogene)	-
2441STX2C_580		38669	38995	327	Head-tail connector protein	108
2441STX2C_590		39005	39355	351	Head-tail adaptor	116
2441STX2C_600		39352	39798	447	Phage tail protein	148
2441STX2C_610		39795	40139	345	Hypothetical protein	114
2441STX2C_620		40108	40914	807	Phage tail protein	268
2441STX2C_630		40920	41294	375	Tail protein	124
2441STX2C_640		41276	41599	324	Tail protein	107
2441STX2C_650		41652	44732	3081	Tape measure protein	1026
2441STX2C_660		44725	45066	342	Minor tail protein	113
2441STX2C_670		45066	45764	699	Minor tail protein	232
2441STX2C_680	*mnt*	45781	46101	321	Regulatory protien Mnt	106
2441STX2C_690	*arc*	46209	46382	174	Transcriptional repressor Arc	57
2441STX2C_700	*ant*	46453	47376	924	Antirepressor protein	307
2441STX2C_710		47431	48168	738	Tail assembly protein	245
2441STX2C_720		48066	48743	678	Tail assembly protein	225
2441STX2C_730		48984	52478	3495	Host specificity protein	1164
2441STX2C_740		52549	53148	600	Outer membrane protein Lom	199
2441STX2C_750	*tfsS*	53213	54526	1314	Tail fibre of Siphoviral Stx phages	437
2441STX2C_760	*atfT*	54408	54797	390	Alternate tail fibre tip	129
2441STX2C_770	*ospB*	54938	55813	876	T3SS secreted effector OspB-like protein	291
2441STX2C_780	*nleG*	56038	56688	651	Non-LEE-encoded type III effector, NleG	216
2441STX2C_790		56673	56957	285	Hypothetical protein	94
2441STX2C_800		57408	57710	303	Hypothetical protein	100

Appendix VIII: Functional annotation of pp1810Stx2k

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
1810STX2_010	*int*	196	1365	1170	Integrase	389
1810STX2_020	*xis*	1349	1531	183	Excisionase	60
1810STX2_030		1610	1987	378	Hypothetical protein	125
1810STX2_040		2023	2235	213	Hypothetical protein	70
1810STX2_050		2195	2821	627	Adenine methylase	208
1810STX2_060		2818	3249	432	Hypothetical protein	143
1810STX2_070		3305	3982	678	Hypothetical protein	225
1810STX2_080		4307	4564	258	Addiction module antidote protein	85
1810STX2_090		4609	4890	282	Hypothetical protein	93
1810STX2_100		4979	5284	306	Addiction module antidote protein	101
1810STX2_110		5327	5983	657	Hypothetical protein	218
1810STX2_120		6159	6782	624	Hypothetical protein	207
1810STX2_130		6786	7073	288	Hypothetical protein	95
1810STX2_140		7075	7293	219	Hypothetical protein	72
1810STX2_150		7295	7510	216	Hypothetical protein	71
1810STX2_160		7470	7976	507	Hypothetical protein	168
1810STX2_170		7978	8925	948	Hypothetical protein	315
1810STX2_180		8922	9161	240	Hypothetical protein	79
1810STX2_190		9154	9357	204	Hypothetical protein	67
1810STX2_200		9354	10232	879	Phosphoadenosine phosphosulphate reductase	292
1810STX2_210		10340	10783	444	Hypothetical protein	147
1810STX2_220		10860	11681	822	Hypothetical protein	273
1810STX2_230		11745	12092	348	Hypothetical protein	115
1810STX2_240		12168	12755	588	Hypothetical protein	195
1810STX2_250	*exo*	12755	13444	690	Exonuclease	229
1810STX2_260	*recT*	13441	14391	951	Recombination DNA repair protein RecT	316
1810STX2_270		14410	14691	282	Hypothetical protein	93
1810STX2_280		14712	14993	282	Hypothetical protein	93
1810STX2_290		15005	15217	213	Killing protein	70
1810STX2_300		15288	15935	648	Phage regulatory protein	215
1810STX2_310		16796	17749	954	Restriction endonuclease	317
1810STX2_320		17746	19215	1470	DNA methyltransferase	489
1810STX2_330	*cI*	19310	20023	714	Repressor protein CI	237
1810STX2_340	*cro*	20095	20322	228	Regulatory protein cro	75
1810STX2_350		20377	20646	270	Hypothetical protein	89
1810STX2_360		20854	21231	378	Endonulcease	125
1810STX2_370		21225	22745	1521	DNA helicase	506
1810STX2_380		22735	23706	972	DNA primase	323
1810STX2_390		23706	24155	450	Hypothetical protein	149
1810STX2_400	*ninG*	24163	24726	564	Protein NinG	187

Locus_tag	gene	Start position	End position	Gene size (bp)	Product	Product size (aa)
1810STX2_410	*ninH*	24723	24917	195	NinH protein	64
1810STX2_420	*q*	24910	25344	435	Antitermination protein Q	144
1810STX2_430	*kan*	26128	26166	1518	Tn5 neomycin phosphotransferase	506
1810STX2_440		28188	30125	1938	Acetylesterase	645
1810STX2_450		30260	30439	180	Putative membrane protein	59
1810STX2_460		30480	30725	246	Hypothetical protein	81
1810STX2_470	*s*	30803	31018	216	Holin	71
1810STX2_480	*r*	31023	31556	534	Lysozyme	177
1810STX2_490		31831	32400	570	Antirepressor	189
1810STX2_500		32415	32549	135	Putative membrane protein	44
1810STX2_510	*rz*	32557	33021	465	Spanin, inner membrane subunit	154
1810STX2_520	*rz2*	32777	32962	186	Spanin, outer membrane subunit	61
1810STX2_530	*bor*	33053	33346	294	Lipoprotein Bor	94
1810STX2_540		33454	33699	246	Hypothetical protein	81
1810STX2_550	*terS*	33755	34561	807	Small subunit terminase	268
1810STX2_560	*terL*	34542	36248	1702	Large subunit terminase	568
1810STX2_570	*porT*	36248	38392	2145	Portal protein	714
1810STX2_580		38550	39557	1008	Hypothetical protein	335
1810STX2_590		39581	40795	1215	Capsid protein	404
1810STX2_600		40851	41240	390	Capsid protein	129
1810STX2_610		41263	41751	489	Hypothetical protein	162
1810STX2_620		41735	42298	564	Hypothetical protein	187
1810STX2_630		42298	42948	651	Hypothetical protein	216
1810STX2_640	*tfpS*	42945	44882	1938	Tail fibre of Podoviral Stx phage	645
1810STX2_650	*atfT*	44764	45153	390	Alternate tail fibre tip	129
1810STX2_660		45239	45481	243	Hypothetical protein	80
1810STX2_670		45548	45661	114	Hypothetical protein	37
1810STX2_680		45698	47401	1704	Hypothetical protein	567
1810STX2_690		47398	48666	1269	Tail tip protein	422
1810STX2_700		48681	48959	279	Hypothetical protein	92
1810STX2_710		48965	49582	618	Hypothetical protein	205
1810STX2_720		49673	52857	-	Outer membrane protein Lom (Pseudogene)	-
1810STX2_730		49871	51409	1539	IS*ec8* Transposase	513
1810STX2_740		51459	51806	348	IS*ec8* Transposase	116
1810STX2_750		53090	53230	141	Prophage host toxic membrane protein	46
1810STX2_760		53287	53688	402	Hypothetical protein	133
1810STX2_770		53782	54438	657	Hypothetical protein	218
1810STX2_780		54441	54887	447	Hypothetical protein	148
1810STX2_790		54897	55148	252	Bacteriocin	83
1810STX2_800		55159	56424	1266	Hypothetical protein	421
1810STX2_810		56455	64875	8421	Hypothetical protein	2806