Stx-Phage Integration and Multiple Lysogeny in *Escherichia coli*

Thesis submitted in accordance with the requirements of the University of Liverpool for the Degree of Doctor of Philosophy

Paul C. M. Fogg

September 2007

TEXT BOUND CLOSE TO THE SPINE IN THE ORIGINAL THESIS

Table of Contents

List of Figures

List of Tables

Page

Abbreviations

Φ	Bacteriophage
λ	Lambda
aa	amino acid
Amp	Ampicillin
BLAST	Basic Local Alignment Search Tool
CAT	Chloramphenicol acetyl transferase
CNS	Central Nervous System
Conc.	Concentration
CSPD	3-(4-methoxyspiro{1,2-dioxethane-3,2-(5'-chloro)tricycle[3.3.1.1^{3,7}]decan}-4-yl)
CTAB	Hexadecyltrimethylammonium Bromide
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EAEC	Enteroaggregative *Escherichia coli*
EDTA	Ethylene diamine tetra acetic acid
EHEC	Enterohaemorrhagic *Escherichia coli*
Ehx	Enterohaemolysin
EIEC	Enteroinvasive *Escherichia coli*
EPEC	Enteropathogenic *Escherichia coli*
Esp	*E. coli* secretory protein
EXPAND	Recombinant hi-fidelity DNA polymerase
Gb	Globotriaosyl ceramide

HC	Haemorrhagic colitis
HUS	Haemolytic Uraemic Syndrome
Int	Integrase
Kan	Kanamycin
LA	Luria Bertani Agar
LB	Luria Bertani
LEE	Locus for enterocyte effacement
LPS	Lipopolysaccharide
MOI	Multiplicity of infection
mRNA	Messenger RNA
NCBI	National Centre for Biotechnology Information
NFLX	Norfloxacin
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
Phage	Bacteriophage
PHLS	Public Health Laboratory Service
RNA	Ribonucleic acid
SDS	Sodium Dodecyl Sulphate
SE	Standard Error
STEC	Shigatoxigenic *E. coli*
Stx	Shiga toxin
TAE	Tris-acetate EDTA
Taq	Recombinant DNA polymerase from *Thermus aquaticus*
TE	Tris EDTA

Tir	Translocated Intimin receptor
Tris	Tris(hydroxymethyl)methylamine
UV	Ultra-violet
WT	Wild type

Units

b	Bases
bp	Base pair
cfu	Colony forming units
g	Grams
h	Hours
kb	Kilobases
kDa	Kilodaltons
kV	Kilovolts
lfu	Lysogen Forming Units
M	Moles per Litre
mg	Milligrams
min	Minutes
ml	Millilitres
mM	Millimoles per litre
mm	Millimetre
MW	Molecular Weight
ng	Nanograms
nm	Nanometres
°C	Degrees Centigrade
OD	Optical Density
rpm	Revolutions per minute
s	Seconds
U	Units
µg	Micrograms

µM	Micromoles per litre
V	Voltage
v/v	Volume/volume
vol.	Volume
w/v	Weight/Volume

Acknowledgements

I would like to thank my supervisors Alan McCarthy and Jon Saunders, as well as Heather Allison, for all their expert advice and assistance throughout the duration of my research. Their input and review of my work has allowed me to finally produce this body of work and to become a peer reviewed author. My thanks also go to BBSRC for their generous funding of my PhD and to DEFRA for alleviating my financial woes during the extensive writing up period.

Special thanks go to my mother to whom I owe a debt of gratitude for all her moral and financial support during my PhD, as well as many years before and, no doubt, still to come. All the members of lab H (past and present) also receive my appreciation. They have all been instrumental in my progress by way of practical assistance, camaraderie and, perhaps most importantly, regular trips to the AJ and beyond. Particular tribute goes to Amy for sharing the stresses of the hectic final months of our respective projects and for brightening the dismal office corner we had to suffer. Finally, credit must go to Elaine who arrived late but took on the invaluable task of distracting me from the hassles and tedium of thesis writing, allowing me to remain sane and pass the finish line unscathed.

Abstract

The characteristic virulence factor of shigatoxigenic *E. coli* strains (STEC) is the expression of Shiga-toxin encoded in the late gene region of temperate lambdoid phages. The use of differentially labelled isogenic recombinant Stx phage ($\Phi 24_B$) had demonstrated the production of true double lysogens in *E. coli*, contrary to the lambda immunity model, however, the insertion sites and underlying mechanisms had not been determined. Here, the bacteriophage $\Phi 24_B$ integrase gene and the preferred site of insertion in the *E. coli* genome have been identified. The integrase encoded by $\Phi 24_B$ differs significantly from the integrases of all other previously characterised phages, and the insertion site shares 24 bp of complimentary sequence with the phage *attP* site. Furthermore, an additional 3 integration sites in the *E. coli* genome were elucidated, each with a decreased level of identity with the *attP* site. Triple lysogens could be produced by means of isogenic phage in which the *stx* gene was interrupted with kanamycin, chloramphenicol and spectinomycin resistance genes as reporters. There were seven possible sites of integration identified, two of which lay within predicted essential genes in the *E. coli* genome, and thus could not be occupied, and integration at another site was never observed. The preferred integration site described above was close to an integrase gene on the *E. coli* chromosome but it was demonstrated unequivocally that this host integrase was not responsible for the multiple integration events.

The complete partially annotated genome sequence of $\Phi 24_B$ revealed the presence of a gene homologous to the putative antirepressor encoding genes of VT2-Sa and Lahn1 and similarity to the well-characterised phage P22 *ant* gene. Antirepressors bind to the lambdoid phage cI repressor responsible for superinfection immunity and it is therefore possible that the presence of an *ant* gene on $\Phi 24_B$ is involved in the absence of the classical lambda immunity response in $\Phi 24_B$. Application of reverse transcriptase PCR showed that the gene is transcribed in a lysogenised *E. coli* host as well as *E. coli* cells undergoing infection, with some evidence for expression levels correlating with multiplicity of infection.

An existing bank of 11 integrase primer sets was supplemented with specific $\Phi 24_B$ integrase primers, and applied in the identification of integrase types from 113 wild-type Stx phage isolates with the aim of examining integrase diversity within these populations of phage induced from a collection of STEC isolates. Despite the inherent mosaic nature of lambdoid phage, the diversity of integrases was limited and dominated by 2 groups, 933W-like and $\Phi 24_B$-like. Demonstration of the ability of Φ24B to form multiple lysogens, as well as the presence of different integrases within a population of Stx phages, has consequences for the evolution of pathogens, as novel Stx phages can emerge due to intracellular recombination events. Multiple lysogens may also produce more Stx-toxin and thus have increased virulence.

"If you would not be forgotten as soon as you are dead and rotten, either write something worth reading or do things worth writing."

-Benjamin Franklin

Chapter 1: Introduction

1.1 *Escherichia coli* Pathogenesis

Escherichia coli is a significant constituent of the natural human gut microflora. However *E. coli* does have the potential to become an opportunistic pathogen, usually in the immuno-compromised host or in the event of a breach of the gut epithelia. In addition, acquisition of various virulence factors can effectively facilitate the transition from commensal organism to disease causing organism in healthy individuals (Kaper *et al.*, 2004). Six distinct pathotypes have been defined to encompass all of the virulent strains of *E. coli* that have evolved thus far (Fig. 1).

1.1.1 Enteropathogenic *E. coli* (EPEC)

EPEC was the first pathotype to be described and is now generally absent from industrialised countries though still a major causative agent of potentially fatal diarrhoea in the developing world (Nataro & Kaper, 1998). Infection with the pathogen results in the formation of characteristic attaching and effacing lesions (A/E). These lesions are produced as a consequence of intimate adherence to the gut epithelia and subsequent induction of extensive cytoskeletal rearrangement to form a pedestal beneath the bacteria (Fig. 2) (Kaper *et al.*, 2004). The ability to effect this alteration of the host phenotype is contained within a 35 kb pathogenicity island (McDaniel *et al.*, 1995), and more specifically the type III secretory system and various effector proteins encoded within it (see section 1.3). The mechanisms underlying pathogenicity comprise a complex combination of factors acting in

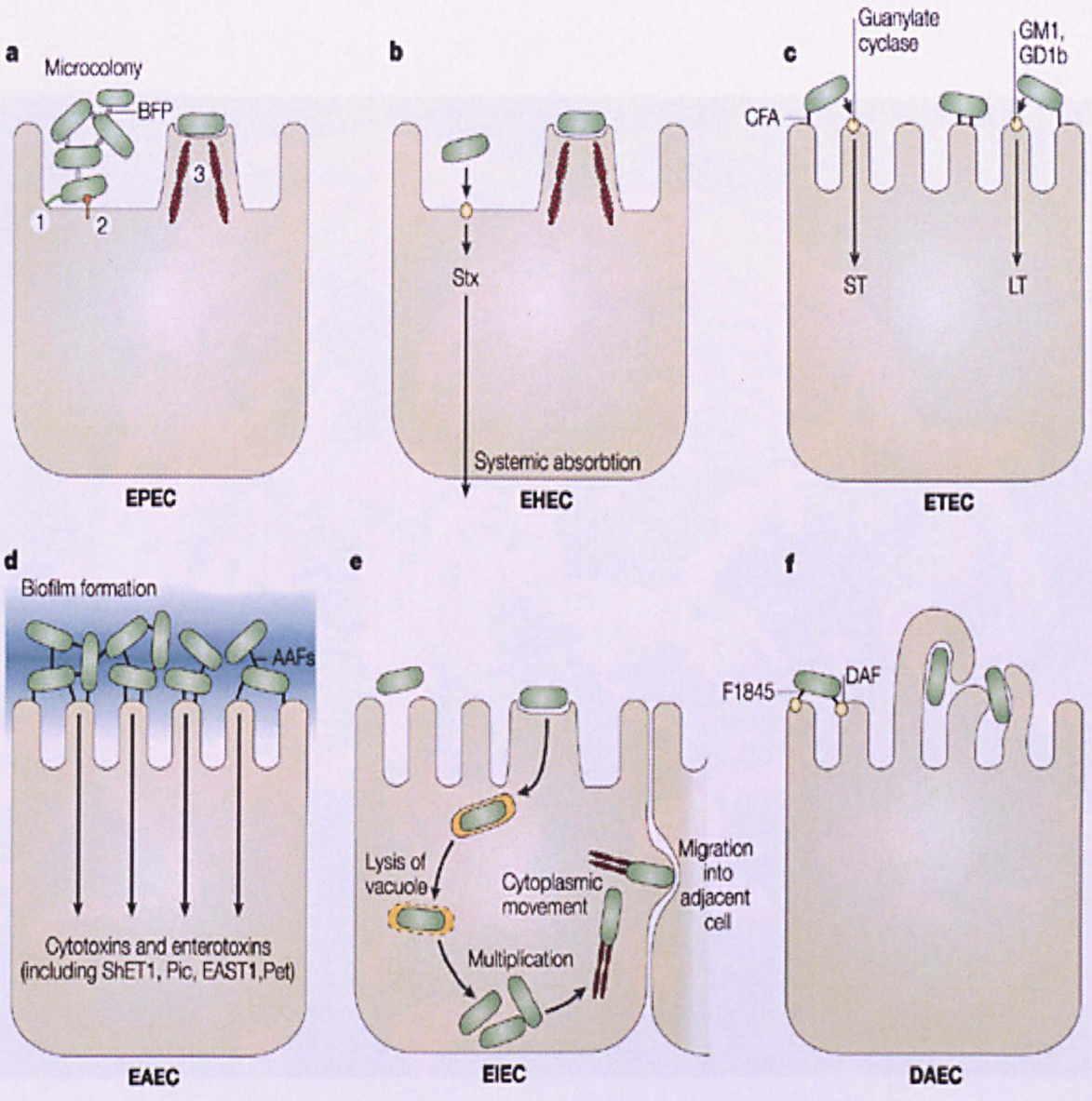


Figure 1. Pathogenic schema of diarrhoeagenic *E. coli* The six recognized categories of diarrhoeagenic *E. coli* each have unique features in their interaction with eukaryotic cells. Here, the interaction of each category with a typical target cell is schematically represented. These descriptions are largely the result of *in vitro* studies and might not completely reflect the phenomena that occurs in infected humans. **a** | EPEC adhere to small bowel enterocytes, but destroy the normal microvillar architecture, inducing the characteristic attaching and effacing lesion. Cytoskeletal derangements are accompanied by an inflammatory response and diarrhoea. 1. Initial adhesion, 2. Protein translocation by type III secretion, 3. Pedestal formation. **b** | EHEC also induce the attaching and effacing lesion, but in the colon. The distinguishing feature of EHEC is the elaboration of Shiga toxin (Stx), systemic absorption of which leads to potentially life-threatening complications. **c** | Similarly, ETEC adhere to small bowel enterocytes and induce watery diarrhoea by the secretion of heat-labile (LT) and/or heat-stable (ST) enterotoxins. **d** | EAEC adheres to small and large bowel epithelia in a thick biofilm and elaborates secretory enterotoxins and cytotoxins. **e** | EIEC invades the colonic epithelial cell, lyses the phagosome and moves through the cell by nucleating actin microfilaments. The bacteria might move laterally through the epithelium by direct cell-to-cell spread or might exit and re-enter the baso-lateral plasma membrane. **f** | DAEC elicits a characteristic signal transduction effect in small bowel enterocytes that manifests as the growth of long finger-like cellular projections, which wrap around the bacteria. AAF, aggregative adherence fimbriae; BFP, bundle-forming pilus; CFA, colonization factor antigen; DAF, decay-accelerating factor; EAST1, enteroaggregative *E. coli* ST1; LT, heat-labile enterotoxin; ShET1, *Shigella* enterotoxin 1; ST, heat-stable enterotoxin (Taken from Kaper *et al.*, 2004).

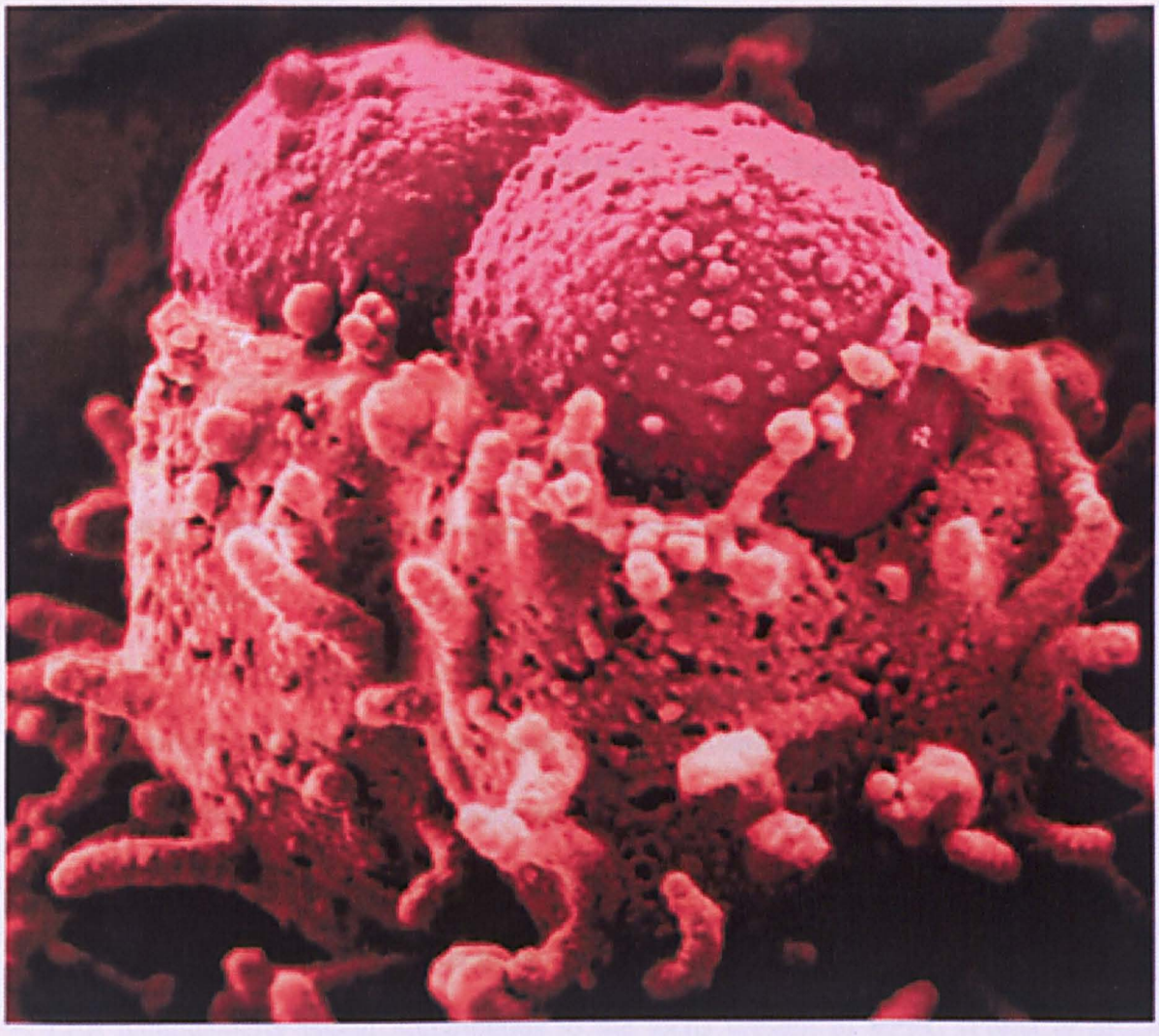


Figure 2. Attaching and effacing histopathology caused by EPEC and EHEC The attaching and effacing histopathology results in pedestal-like structures, which rise up from the epithelial cell on which the bacteria perch (Taken from Kaper *et al.*, 2004).

addition to attachment and toxin secretion. Contributory factors include active ion secretion, increase of intestinal permeability, inflammation and loss of epithelial surface area due to microvilli effacement.

1.1.2 Enterohaemorrhagic *E. coli* (EHEC) / Shiga-Toxin Producing *E. coli* (STEC)

The principal virulence determinant in EHEC is expression of Shiga-like toxins (Stx1 and Stx2). Stx1/2 are encoded on temperate lambdoid bacteriophages resident in the bacterial genome as either prophages or defective prophages. The toxin is produced in the colon where it causes localised damage including necrosis, haemorrhagic colitis (HC) and intestinal perforation, but it may also spread systemically and cause inflammation, vascular thromboses and endothelial damage (Andreoli *et al.*, 2002). Over 500 serotypes of *E. coli* have been identified as Stx positive, however most of these are not usually associated with expression of disease (Gyles, 2007; Kaper *et al.*, 2004). EHEC are a virulent sub-group of STEC strains that additionally possess a LEE pathogenicity island homologous to the locus present in EPEC (Perna *et al.*, 1998). EHEC was first described as the causative agent in cases of haemorrhagic colitis (HC) and haemolytic uraemic syndrome (HUS) in 1982 (Cohen & Giannella, 1992; Cordovez *et al.*, 1992; Riley *et al.*, 1983). The infective dose is extremely low, and it has been reported that less than 100 cells are required to produce manifestation of disease (Kaper *et al.*, 2004). Some strains of EHEC also carry the pO157 virulence plasmid, which encodes virulence factors such as enterohaemolysin (Ehx) and catalase-peroxidase (KatP) (Burland *et al.*, 1998).

1.1.3 Enterotoxigenic *E. coli* (ETEC)

ETEC are associated with mild to severe watery diarrhoea and these strains of *E. coli* are an important cause of childhood infections and travellers diarrhoea in developing countries. ETEC colonise the small intestine, a process mediated by proteinaceous fimbral colonisation factors, and secrete toxins into the lumen (Wolf, 1997). Two toxins, heat-labile (HL) and heat-stable (ST), are produced by members of the enterotoxigenic pathotype, and individual strains may produce one or both toxins (Kaper *et al.*, 2004). HL is an AB_5 holotoxin structurally similar to the cholera toxin of *Vibrio cholerae*, which initiates a regulatory cascade resulting in an increased secretion of chlorine ions by crypt cells and consequent diarrhoeal symptoms (Spangler, 1992). There are two ST variants, STa and STb; the former is a molecular mimic that imitates an endogenous ligand, guanylin, ultimately leading to increased secretion whilst the latter has not been linked to human disease (Nataro & Kaper, 1998). Interestingly STa is known to suppress the onset of colon cancer and this could therefore explain the correlation between the prevalence of ETEC infection in developing countries and the decreased incidence of cancer of the colon (Pitari, 2003).

1.1.4 Enteroaggregative *E. coli* (EAEC)

EAEC are the source of chronic persistent diarrhoea in children and adults worldwide and are characterised by auto-aggregative adherence morphology (Nataro & Kaper, 1998). Disease pathology comprises loose adherence to the intestinal mucosa, predominantly in the colon, and the release of toxins (Nataro *et al.*, 1998). Infection results in mild but widespread mucosal damage due in part to interleukin-8 stimulated

neutrophil transmigration (Steiner *et al.*, 2000). A number of EAEC toxins have been described though none have been conclusively linked to virulence.

1.1.5 Enteroinvasive *E. coli* (EIEC)

EIEC pathogenic profile is closely related to *Shigella* (Kaper *et al.*, 2004). Although EIEC can be responsible for invasive disease and dysentery, more often EIEC causes watery diarrhoea similar to other pathogenic *E. coli* strains (Nataro & Kaper, 1998). During the early stages of infection, EIEC penetrate the gut epithelia where they mediate lysis of the endocytic vacuole, multiply, migrate through the cytoplasm and ultimately extend into adjacent cells (Sansonetti, 2002). The genes essential for EIEC and *Shigella* pathogenic characteristics are encoded on a large 213 kb virulence plasmid, pWR100, and sequence data from *Shigella* indicates that the encoded elements were originally encoded on four distinct plasmids (Buchrieser, 2000). pWR100 encodes a type III secretory system, a 120-Kda outer membrane protein (IcsA) and various effector proteins involved with mediation of epithelial signalling events, cytoskeletal rearrangements, cellular uptake and lysis of the endocytic vacuole (Tran Van Nhieu *et al.*, 2000). In addition, various other non-essential accessory virulence have been identified *e.g.* secreted proteases and an iron acquisition system (Kaper *et al.*, 2004).

1.1.6 Diffusely Adherent *E. coli* (DAEC)

DAEC characterised by diffuse adherence to Hep-2 cells and have been implicated as a cause of children's diarrhoea in the over 12 months age bracket (Kaper *et al.*, 2004).

DAEC strains induce a cytopathic effect on the epithelial cells, characterised by the formation of long cellular extrusions, which envelop the attached bacteria. The process is mediated by the binding of a fimbral adhesin, produced by the majority of DAEC strains, to an enterocyte surface protein receptor, DAF (Bilge *et al.*, 1989). Receptor binding also initiates the decay-accelerating factor (CD55) signal transduction promoting cytoskeleton F-actin rearrangements (Peiffer *et al.*, 1998). Infection of intestinal cells also results in the impairment of enzyme activity and biosynthesis of brush border-associated hydrolases, independent of the adhesin mediated system (Peiffer *et al.*, 2001).

1.2 Shigatoxigenic *E. coli*: Animal Reservoir and Seasonal Shedding

The strains able to cause disease in humans are divided into several pathotypes (see Chapter 1.1), possibly the most important of which is the Shiga-like toxin producing *E. coli* and their enterohaemorrhagic subgroup (STEC and EHEC, respectively). Infection with STEC may be asymptomatic and severe complications in healthy adults are rare with those at risk being primarily the elderly, young or immuno-compromised. Haemolytic uraemic syndrome (HUS) is the major cause of acute renal failure in children and in excess of 90% of those HUS cases are caused by STEC infection (Banatvala et al., 2001; Fischer et al., 2001; Mahon et al., 1997).

There is a large spectrum of STEC serotypes isolated from human clinical samples from various countries and almost 500 serotypes have been isolated from humans with disease (Gyles, 2007). Despite this, less than 10 O-serogroups comprise the majority of disease cases, with O157 strains responsible for the main outbreaks in developed countries (Fig. 3). Non-O157 serotypes have been associated with sporadic outbreaks, although at present, the diagnostic procedures are inadequate to detect all such strains, in contrast to the more sensitive methods that have been developed for *E. coli* O157, and this may disguise the actual comparative prevalence of disease-causing non-O157 serotypes (Beutin, 2006; Fischer *et al.*, 2001). Ruminants, especially cattle, are the major reservoir of STEC, and human outbreaks of disease can often be traced back to bovine faecal contamination of foodstuffs or water supplies. STEC infections are therefore classified as zoonotic diseases (Gyles, 2007) and STEC will usually colonise the gastrointestinal (GI) tract of the adult animal without causing disease.

Over 435 serotypes of STEC have been identified in cattle and over 470 serotypes in human clinical isolates; the majority of the latter had been first identified in cattle (Gyles, 2007). The prevalence of STEC colonisation of animal hosts has been cited as up to 71% of individual animals and up to 100% of herds (Cerqueira *et al.*, 1999; Cobbold & Desmarchelier, 2000). The rate and duration of STEC shedding can vary greatly; detection of *E. coli* O157 in cattle faeces can range from 4 cfu 10 g^{-1} to > 1.1 x 10^5 cfu 10 g^{-1} over a period of less than one week to ten weeks within a single farm (Widiasih *et al.*, 2004). Despite this apparently large range in the magnitude of shedding, the usual rate is ca. 10-100 cfu g^{-1}. Naylor *et al.* (2003) found that experimentally infected calves shed STEC O157 for more than 14 days, with the clear majority being shed on the surface of the excreted faeces concomitant with significantly decreased levels detected in the GI lumen. In almost all persistently colonised hosts examined, the majority of tissue-associated *E. coli* O157 microcolonies were located within 5 cm of the recto-anal junction, an area typified by a high density of lymphoid follicles (Naylor *et al.*, 2003).

It has been reported that the most common human STEC pathogens are not the primary colonisers of the ruminant reservoir hosts (Roldgaard *et al.*, 2004). This is in contrast to an Australian study of three dairy farms, which demonstrated that calves tested both pre- and post-weaning, shed the largest number of putative EHEC strains and were the principle source of the serotypes most commonly associated with human disease (Cobbold & Desmarchelier, 2000). Human pathogenic strains may also occasionally come to predominate in the ruminant populations via a phenomenon known as "super-shedding". Mathematical modelling concluded that the top 20% most infective individual animals were responsible for approximately 80% of *E. coli*

O157 cattle to cattle transmission (Matthews *et al.*, 2006). Furthermore studies have shown that 3.7-9% of cattle studied shed in excess of 10^3 cfu g^{-1} faeces (Low *et al.*, 2005; Omisakin *et al.*, 2003). The relative abundance of these super-shedders could go some way to explain the seasonal and spatial variation in transmission and prevalence. A statistically significant correlation between cattle density and reported human infections has been described in Finland, Sweden and Canada (Eklund *et al.*, 2005; Kistemann *et al.*, 2004; Michel *et al.*, 1999). However this link was not universally observed suggesting additional contributory factors. A significant geographical disparity has been observed between *E. coli* O157 levels in beef cattle faeces, sampled immediately prior to harvest, in Eastern Colorado and Central Nebraska (Dewell *et al.*, 2005). The observed differences may be due to multiple variables, such as precipitation, soil content and regional management practices. Furthermore, cattle feed was also shown to effect shedding as those animals fed brewers grain were six times more likely to culture positive for *E. coli* O157 (Dewell *et al.*, 2005). Seasonal variation may also be of importance; STEC infections were found to be more prevalent in the summer months or early autumn (Banatvala *et al.*, 2001; Eklund *et al.*, 2005; Elliott *et al.*, 2001; Willshaw *et al.*, 2001), coinciding with peak STEC shedding from cattle (Barkocy-Gallagher *et al.*, 2003).

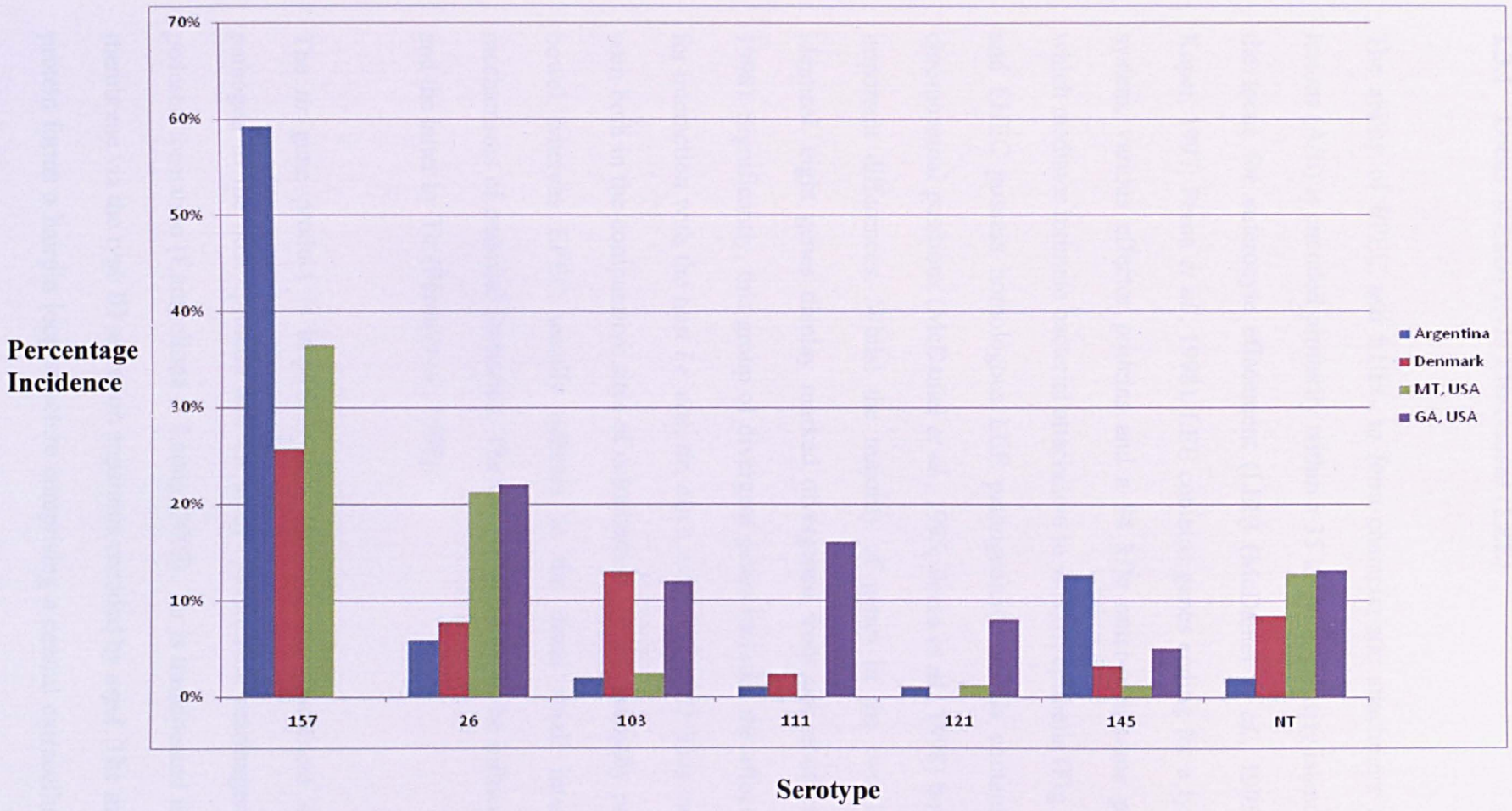


Figure 3. Serogroups of Shiga toxin-producing *Escherichia coli* frequently implicated in disease in humans Serogroup incidence represented by the percentages of total STEC isolates, from each study, belonging to relevant O-serotypes. NT = nontypable or partially typable (Adapted from Gyles, 2007). GA, USA study was of non-O157 serotypes only and as such O157 serotypes are not represented here.

1.3 *E. coli* Virulence Factors

1.3.1 Locus of Enterocyte Effacement (LEE)

The ability of EPEC and EHEC to form characteristic attachment and effacement lesions (A/E) is encoded primarily within a 35 kb pathogenicity island (PI) known as the locus for enterocyte effacement (LEE) (McDaniel *et al.*, 1995; McDaniel & Kaper, 1997; Perna *et al.*, 1998). LEE contains genes coding for a type III secretory system, various effector proteins and a 94 kDa outer-membrane protein, intimin, which mediates intimate bacterial attachment to the host epithelia (Fig. 4). Both EPEC and EHEC possess homologous LEE pathogenicity islands contained at identical chromosomal positions (McDaniel *et al.*, 1995; Perna *et al.*, 1998) however there are important differences. Whilst the majority of genes in the two loci are almost identical, eight genes display marked divergence from one another (Perna *et al.*, 1998). Significantly, this group of divergent genes includes the effectors responsible for interaction with the host *i.e. eae*, *tir*, *espA*, *espB* and *espD*. This variability can be seen both in the comparative sites of colonization (EHEC normally prefers the lower bowel whereas EPEC usually adheres to the distal small intestine) and the mechanisms of pedestal formation. The former is known to be influenced by intimin and the latter by Tir (Perna *et al.*, 1998).

The *tir* gene product is important both for intimate attachment of the bacterial pathogen to the host epithelia and for actin cytoskeleton rearrangement leading to pedestal formation (Campellone & Leong, 2003). Tir is translocated into the host cell membrane via the type III secretion apparatus encoded by *espA*. The membrane bound protein forms a hairpin loop structure comprising a central extracellular domain and

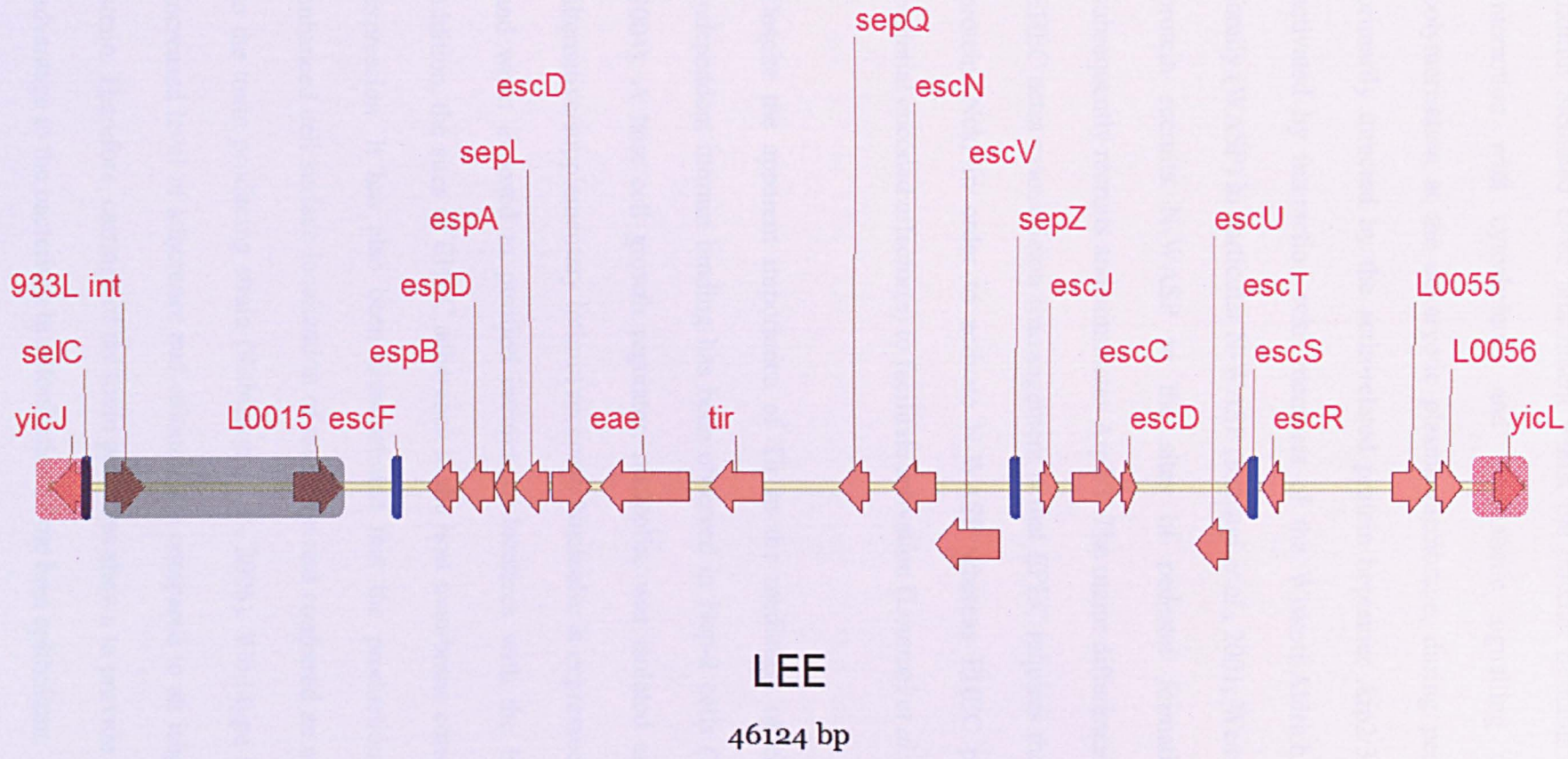


Figure 4. Diagram of the EDL933 LEE Putative prophage 933L is highlighted in grey. Genes homologous to those on the K-12 chromosome are highlighted red. Genes common to both th EDL933 (EHEC) and E2348/69 (EPEC) LEEs are unhighlighted.

intracellular amino- and carboxy- terminals. The external loop is the receptor for the intimin mediated close attachment whilst the internal domains are responsible for interaction with cytoskeleton and cytoplasmic signalling components. Actin polymerisation at the eukaryotic plasma membrane, during pedestal formation, is primarily directed by the actin-related protein heptamer Arp2/3. Arp2/3 in turn is activated by interaction with members of the Wiskott-Aldrich Syndrome Protein family (WASP) in particular N-WASP (Rohatgi *et al.*, 2001; Weaver *et al.*, 2002). Tir protein recruits N-WASP to the sites of pedestal formation and N-WASP subsequently recruits and stimulates Arp2/3. The major difference between EPEC and EHEC actin cytoskeleton rearrangement is that EPEC requires the presence of a host protein, Nck, in order to activate N-WASP whereas EHEC possesses additional bacterial encoded effector(s) to facilitate activation (Lommel *et al.*, 2004).

Despite the apparent importance of Tir in the mediation of close adhesion, Tir-independent intimin binding has been observed in Hep-2 cells (Sinclair & O'Brien, 2004). A host cell growth regulator, nucleolin, was isolated and identified as an alternative/supplementary intimin receptor. Nucleolin is expressed on the cell surface and when exposed to purified intimin, co-localizes with the bacterial protein. In addition, the sites of EHEC adherence to the host membrane correlate with nucleolin expression. It has also been demonstrated that the production of Stx2 mediated enhanced cell surface localization of nucleolin and conferred an adherence advantage to the toxin producing strain (Robinson *et al.*, 2006). Wild-type EHEC displayed an increased level of adherence and colonization compared to an isogenic Stx2 deficient strain. Therefore, carriage of the toxin gene was shown to provide a direct competitive advantage to the bacterium in colonisation of the host epithelium.

Effector proteins translocated by the type III secretory system are not limited to those contained within LEE. The first non-LEE encoded translocated effector protein, Cif (cycle inhibiting factor), was discovered by Marches *et al.* (2003), and is borne by a lambda-like prophage. Cif exerts an irreversible cytopathic effect on the host cell via the progressive recruitment of focal adhesins and the assembly of stress fibres. The mechanism of action is linked to the cyclin-dependent kinase Cdk-1, an important protein involved in driving mitosis, whereby Cdk-1 is maintained in a premitotic phosphorylated state. This interaction with Cdk-1 essentially leads to the termination of cell proliferation and cessation of mitosis. The carriage of this gene on a lambdoid bacteriophage further underlines the increasing recognition of the importance of phage in the evolution of virulence.

In addition the EAF (EPEC adherence factor) plasmid carried by EPEC strains contains the *bfp* gene cluster, encoding type IV bundle forming pili (Hicks *et al.*, 1998). The pili were originally thought to be involved in initial adherence; however they are now known to mediate inter-cellular adherence between bacteria and are vital in the formation of microcolonies. Accordingly, disruption of *bfp* leads to the production of abnormal colony morphologies.

1.3.2 pO157 Plasmid

The complete sequence of the 92kb virulence plasmid of *E. coli* O157:H7, pO157, revealed 100 open reading frames including a variety of genes with strong homology to known virulence factors (Burland *et al.*, 1998). The major factors contained on the plasmid include enterohaemolysin (Ehx), catalase/peroxidase (KatP), extracellular serine protease (EspP), *Clostridium difficile*-like toxin (ToxB) and a type II secretion pathway (Law, 2000). The pO157 plasmid is heterogeneous in nature with regard to gene composition and arrangement, and distinct serotype-related differences can be observed (Brunder *et al.*, 2006; Burland *et al.*, 1998; Makino *et al.*, 1998).

1.3.2a Enterohaemolysin

Ehx is part of the RTX family of toxins and has approximately 60% homology to *E. coli* α-haemolysin (Bauer & Welch, 1996). The production of Ehx is characterised by small zones of haemolysis due to a defective secretion system and, as such, complementation of *E. coli* 0157 with the α-haemolysin transport genes results in increased activity (Schmidt *et al.*, 1996). Production of the haemolysin is optimal *in vitro* under conditions of low oxygen availability approximating the conditions the gut environment (Chart *et al.*, 1998). Prevalence of *ehx* has been found to be strongly associated with the presence of *eae* and is more commonly associated with human isolates as opposed to animal isolates (Boerlin *et al.*, 1999; Sandhu *et al.*, 1996). Although *ehx* is frequently found in causal strains of HUS and HC, it is not essential for the presentation of symptoms. Its conserved association does suggest a role in virulence and selective pressure (Schmidt & Karch, 1996). The most likely role of Ehx is to increase haemoglobin availability and thus boost pathogen fitness.

1.3.2b Catalase-Peroxidase

KatP is an 82 kDa protein that exhibits dual functionality as both a catalase and a peroxidase (Brunder *et al.*, 1996; Varnado *et al.*, 2004). KatP forms part of a sub-group of periplasmic catalase-peroxidases, along with KatA (*Legionella pneumophila*) and KatY (*Yersinia pestis*), which share markedly more identity with one another than with their cytoplasmic counterparts *e.g. E. coli* KatG (Faguy & Doolittle, 2000; Klotz & Loewen, 2003). This sub-group has been shown to be associated with highly virulent pathogenic strains but not with mildly or non-pathogenic strains. In addition, up-regulation of KatY expression has been linked to an increase in temperature from 26 °C to 37 °C, corresponding to the shift encountered when *Y. pestis* is transferred from the vector to the human host (Garcia *et al.*, 1999). Despite the apparent importance of these enzymes, the mechanisms underlying their contribution to pathogenesis are somewhat unclear. However, the probable function of the catalase-peroxidases is to counteract certain aspects of the host neutrophil response. The presence of a highly active catalase within the periplasm allows hydrogen peroxide, or other host reactive oxygen species and their oxidant derivatives, *e.g.* HOCl, to be removed before components of the inner membrane are exposed and damaged (Varnado *et al.*, 2004). The peroxidase elements could also play a major part by reducing reactive peroxides such as peroxynitrite, $ONOO^-$, and peroxynitrous acid, ONOOH (Varnado *et al.*, 2004). This function is particularly important as the pKa of peroxynitrite dictates that under physiological conditions a significant proportion would be expected to exist in the ionic state. This abundant ionic form is unable to pass through the inner membrane, where it may be dealt with by cytosolic enzymes, and thus must be neutralized in the periplasm.

1.3.2c Extracellular Serine Protease

EspP is a member of the group of serine protease autotransporters of the *Enterobacteriaceae* (SPATE) subfamily of autotransporters that mediate their own secretion through the outer membrane (Velarde & Nataro, 2004). EspP is synthesised as a 1300 amino acid precursor protein that subsequently undergoes N- and C-terminal processing prior to secretion. The protein sequence is homologous to various secreted or surface proteins of pathogenic bacteria, including a significant identity to the EPEC autotransporter EspC (Brunder *et al.*, 1997). The secreted toxin is known to cleave a narrow range of substrates that include pepsin A and human coagulation factor V (Brunder *et al.*, 1997). This cleavage may contribute to the haemorrhagic complications associated with EHEC infection by compromising the blood-clotting cascade. Moreover, patients suffering from EHEC infection produce anti-EspP antibodies showing that the protein is expressed and secreted *in vivo* (Brunder *et al.*, 1997). Analysis of the prevalence of the *espP* gene by hybridisation and immunoblot analysis revealed widespread presence of the gene in *E. coli* O157 strains, with detection also noted in the O26 serogroup (Brunder *et al.*, 1997).

1.3.2d *Clostridium difficile*-like Toxin

Sequencing of the pO157 plasmid revealed the presence of an open reading frame encoding a large 3169 amino acid protein (ToxB) that showed strong sequence similarity to the catalytic N-terminal domain of the large clostridial family of toxins, LCT (Burland *et al.*, 1998; Law, 2000). For the characterised clostridial AB toxins, the C-terminal domain is responsible for entry into the host cell whilst the active N-terminal domain modifies Rho, Rac and Ras proteins resulting in altered cytoskeleton

regulation. The homology of ToxB to LCTs may indicate a shared mechanism of action. Furthermore, respective infections with EHEC and *Clostridium difficile* result in comparable pathologies (Nataro & Kaper, 1998). More recently, comparisons have been drawn between ToxB and two *E. coli* virulence factors, LifA and Efa-1. The former inhibits lymphocyte proliferation as well as interleukin-2 (IL-2), IL-4 and gamma-interferon expression (Klapproth *et al.*, 2000). The latter has been implicated as an adhesion factor (Nicholls *et al.*, 2000; Stevens *et al.*, 2002). EHEC 0157 strains do not contain the *lifA* gene and only a truncated version of *efa-1*. Curing the strains of the pO157 plasmid removes the lymphostatic activity of the cells, and mutational studies of *efa-1* and its truncated chromosomal copy result in reduced adherence to cultured epithelia and expression\secretion of LEE effectors (Stevens *et al.*, 2004).

1.3.2e Type II Secretion Pathway

The pO157 plasmid includes a type II secretion system encoded by the *etp* operon (Schmidt *et al.*, 1997). The system is required for the secretion of a metalloprotease, StcE, the expression of which is under the regulation of the global regulator Ler (Lathem *et al.*, 2002). StcE is not a general protease and the initial substrate described was a serine protease inhibitor, C1-Inh (Caliezi *et al.*, 2000). C1-Inh is a regulator of the complement system and is involved in the inflammatory response. Cleavage by StcE does not inhibit C1-Inh function and it also results in the localization of the inhibitor to the cell membranes, effectively enhancing its regulatory role. In effect, StcE mediates the reduction of the host inflammation and lytic response (Lathem *et al.*, 2004). Two further StcE substrates have been identified, glycoprotein 340 (gp340) and mucin 7 (MUC7). Both proteins play defensive roles in saliva and other tissues.

StcE's regulation by Ler suggests that the enzyme contributes to attachment and effacement via the breakdown of protective mucin and glycoproteins in the gut. Expression of the protein also decreases the viscosity of saliva leading to a proposed role in pathogen survival (Grys *et al.*, 2005).

1.4 Shiga-like Toxins

Shiga first discovered the causal agent of bacterial dysentery, *Shigella dysenteriae*, in 1898 but it was not until 1972 that the toxin itself was identified and subsequently characterised. The Stx toxins (Stx) are AB_5 holotoxins, structurally similar to cholera toxin and ricin and are comprised of a single A sub-unit (32 kDa) and 5 B sub-units (7.7 kDa each), which are involved in receptor recognition and binding (Sandvig, 2001). The A sub-unit is made up of 2 domains, A1 and A2, linked by a disulphide bridge. A1 is the enzymatically active domain while A2 is essential for toxin assembly. The pentameric B components are primarily composed of anti-parallel β sheets, and it is the cleft between adjacent sheets that interacts with the host receptor (Stein *et al.*, 1992). Shiga-like toxins are important contributory factors to the onset of the more serious complications associated with *E. coli* and *S. dysenteriae* pathogenesis, such as haemolytic uraemic syndrome (HUS) and haemorrhagic colitis (HC). Shiga toxin shares 98% sequence homology with bacteriophage-encoded Stx1 and are both antigenically susceptible to antisera raised against either toxin. In contrast, Stx2 is less than 60% homologous, antigenically distinct and displays a great deal of diversity within itself and to date seven variants labelled a to g have been identified (Garcia-Aljaro *et al.*, 2006; Sandvig, 2001). Stx2 variation is mainly limited to the B sub-unit, altering the specificity and affinity of the toxin for the host receptor and, as a consequence, overall toxicity. Shiga toxin was originally thought to be of bacterial origin, however studies of the *S. dysenteriae stx*–encoding region, as well as *stx*–encoding regions of STEC strains with no discernable bacteriophages, has demonstrated that the toxin genes are always linked to lambdoid phage sequences (McDonough & Butterton, 1999; Mizutani *et al.*, 1999; Unkmeir & Schmidt, 2000).

Several eukaryotic outer membrane proteins act as receptors for Stx toxins. The major receptor is globotriasoyl ceramide 3 (Gb3), though globotriasoyl ceramide 4 (Gb4) and P_1 (A blood group glycolipid antigen) are also recognised (Lingwood *et al.*, 1987; Mainil, 1999; O'Loughlin & Robins-Browne, 2001). The different toxins and toxin variants display differing affinities for these receptors and each possesses varying levels of tissue specificity based receptor prevalence (O'Loughlin & Robins-Browne, 2001). For example, Stx1-infected rabbits suffer from inflammation of the central nervous system and the gut, but not the renal system, consistent with relative Gb3 distribution in the tissue types in the rabbit; whereas Stx2e affects the cerebellum and colon of infected porcine hosts correlating with differential Gb4 expression levels in those tissues (Boyd *et al.*, 1993; Richardson *et al.*, 1992).

Once the toxin binds to the host receptor, it is rapidly internalised by host mediated endocytosis (Sandvig, 2001). Following up-take, toxin molecules are transported to the endoplasmic reticulum (ER) via the Golgi apparatus. At this point, the toxin is cleaved at the trypsin sensitive disulphide bond, by host enzymes in the ER and cytosol to yield a 27 kDa A1 product and a 5 kDa A2 protein (Garred *et al.*, 1995a; Olsnes *et al.*, 1981). Although cleavage markedly increases toxicity, undigested toxin still displays activity with a belated effect (Garred *et al.*, 1995b). Furin serine protease is the major proteolytic enzyme in the process of toxin activation (Garred *et al.*, 1995a) though other cytosolic enzymes play an important role for some Stx variants, *e.g.* Stx2e (O'Loughlin & Robins-Browne, 2001). The cellular target for activated A1 toxin is the ubiquitous eukaryotic 60S ribosome, which is comprised of the 5S, 5.8S and 28S sub-units. A1 cleaves the 28S sub-unit resulting in the inhibition of amino-acyl tRNA binding to the ribosome and, in turn, the prevention of protein elongation

(Paton & Paton, 1998; Reisbig *et al.*, 1981). The toxin also displays a degree of activity against some prokaryotic strains, thus providing an additional competitive advantage (Suh *et al.*, 1998).

Stx directly damages gut epithelium as characterised by mucosal inflammation and reduced fluid and nutrient absorption when purified toxin is inoculated into the rabbit ileal loop (Keusch *et al.*, 1972). Intestinal epithelia are able to translocate unmodified Stx from the lumen to the basolateral side of the cell whilst sustaining no cytotoxic effects in the process (Acheson *et al.*, 1996). This process is highly important in the spread of the toxin and the manifestation of systemic complications.

In the gastrointestinal tract, the major role of Stx is in the induction of vascular injury. The production of toxin results in more severe bloody diarrhoea and mucosal damage as compared to an equivalent bacterial strain lacking an Stx-phage in a rabbit model (Sjogren *et al.*, 1994). Histologically, the difference between the two infections is the presence of microvascular thrombosis in the submucosa of the colon and cecum (Sjogren *et al.*, 1994). Similar thromboses have been identified in human infection (Griffin *et al.*, 1990). Another important effect of toxin presence in the GI tract is the stimulation of a mucosal inflammatory response, involving the production of cytokines by epithelial and immune cells (Fig. 5).

It has been shown that Stx1 can be environmentally regulated by several factors but perhaps most significantly by the presence of iron. In the Stx-phage H-19B, an

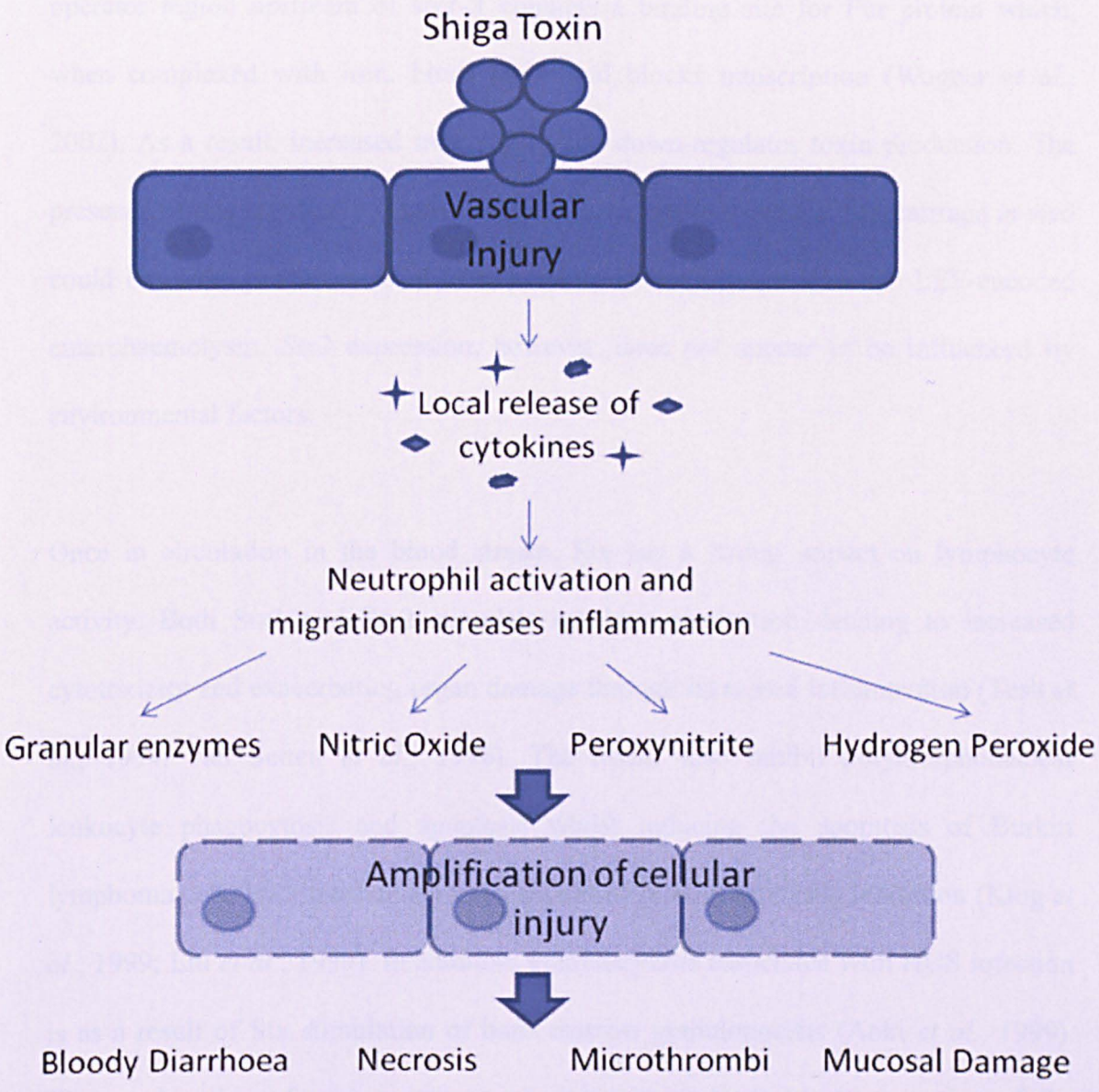


Figure 5. Mechanisms of endothelial cell injury caused by Shiga toxin Stx induced vascular injury stimulates the release of various cytokines, most notably interleukin-8, which activate host immune cells. The leukocytes secrete various enzymes and reactive oxygen metabolites resulting in an amplification of cellular injury and downstream sequelae (O'Loughlin & Robins-Browne, 2001).

operator region upstream of *stx1-A* contains a binding site for Fur protein which, when complexed with iron, binds DNA and blocks transcription (Wagner *et al.*, 2002). As a result, increased iron availability down-regulates toxin production. The presence of this regulatory system suggests that a major benefit of Stx carriage *in vivo* could be in the procurement of iron, possibly in conjunction with the LEE-encoded enterohaemolysin. *Stx2* expression, however, does not appear to be influenced by environmental factors.

Once in circulation in the blood stream, Stx has a strong impact on lymphocyte activity. Both Stx1 and Stx2 stimulate cytokine production, leading to increased cytotoxicity and exacerbating organ damage through increased inflammation (Tesh *et al.*, 1994; van Setten *et al.*, 1996). The toxins also inhibit polymorphonuclear leukocyte phagocytosis and apoptosis whilst inducing the apoptosis of Burkitt lymphoma cells by a mechanism independent of protein synthesis inhibition (King *et al.*, 1999; Liu *et al.*, 1999). In addition, granulocytosis associated with HUS infection is as a result of Stx stimulation of bone marrow granulopoeisis (Aoki *et al.*, 1999). The combination of wide-ranging toxin interactions leads to augmentation of the immune system to enhance toxicity.

The main site of organ damage caused by Stx in humans is the kidney, which expresses large quantities of the receptor Gb3 in the medulla and cortex (Boyd & Lingwood, 1989). Renal pathology includes capillary wall thickening, swollen endothelia, capillary thrombosis, cortical necrosis and extensive fibrin deposition (O'Loughlin & Robins-Browne, 2001). This variety of biological effects accounts for the clinical range of disease from anuria to uremia. Although the kidney is the major

site of organ damage, the majority of fatalities are associated with the toxin's impact upon the central nervous system. As with renal pathology, the range of manifestations is diverse in the CNS from cerebral oedema and diffuse hypoxic ischaemic changes to thrombosis and cerebral infarction (O'Loughlin & Robins-Browne, 2001). CNS complications are still observed in infections with non-Stx strains indicating the importance of additional factors, however the direct impact of the toxin has also been established (Ashkenazi *et al.*, 1990; Khan *et al.*, 1999). Purified toxin accumulates within affected tissues in animal models and histological data correlates tissue damage with toxin binding (Richardson *et al.*, 1992). It is thought that Stx targets microvascular endothelial cells where it triggers deposition of microthrombi and results in downstream sequelae such as paralysis, coma and death (Harrison *et al.*, 2005).

1.5 The Temperate Bacteriophage Life Cycle

Bacteriophage lambda is a temperate bacteriophage and as such is able to undergo lytic infection of the bacterial host, *Escherichia coli*, resulting in cell lysis and release of viral progeny, or alternatively it may establish a lysogenic state whereby the phage integrates into the host genome and the majority of viral expression is repressed (Court *et al.*, 2007). During the lysogenic state, superinfection with a subsequent identical or similar bacteriophage is precluded by immediate repression of gene expression from the incoming phage genome (Lwoff, 1953). The lytic and lysogenic life cycles are controlled by a complex system of pathways consisting of interconnected positive and negative regulators of expression (Fig. 6). The lambda regulatory and developmental pathways have been extensively studied for over 40 years, yet the intricacies of the lambda system, and that of the closely related lambdoid phage, continue to evolve.

After λ phage infection of *E. coli,* 2 genes (*N* and *cro*) are immediately transcribed from the pL and pR promoters (Court *et al.*, 2007; Ptashne, 2004). The former is an anti-terminator that acts as an alternative σ factor for the bacterial RNA polymerase. The N protein forms a complex in conjunction with host encoded Nus proteins and RNA polymerase. This complex is able to overcome termination sequences in the phage DNA at tL1 and tR1. The latter encodes a protein responsible for preventing repressor protein cI binding to the right operator binding regions, $O_{R1\text{-}3}$. The combination of the 2 gene products allows the transcription of the delayed early genes encoding cII and cIII lysogenic control proteins, O and P DNA replication initiation

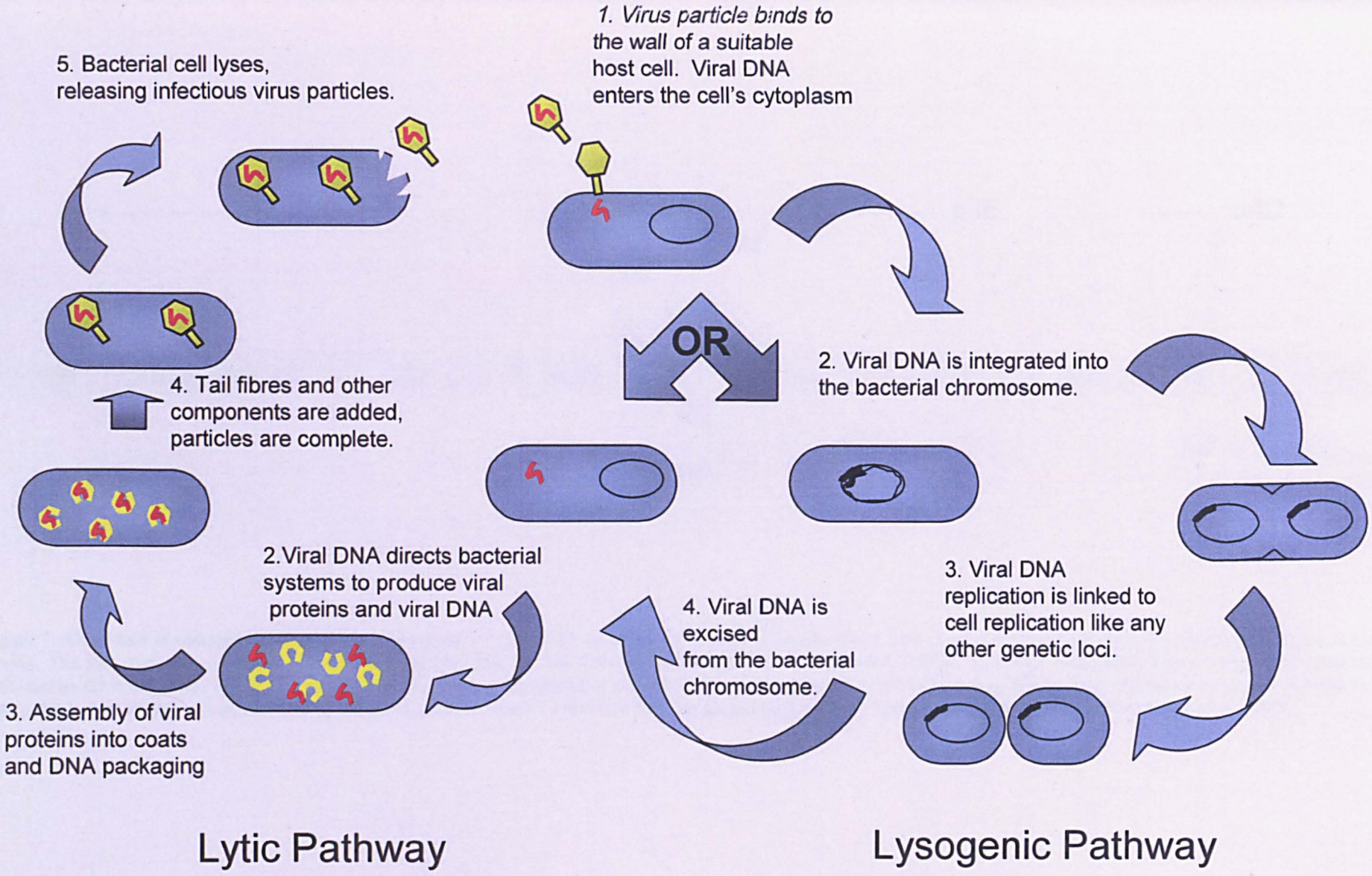


Figure 6. Temperate Bacteriophage Life Cycle Schematic of the genetic switch of temperate bacteriophages, highlighting the key stages of the lytic and lysogenic life cycles.

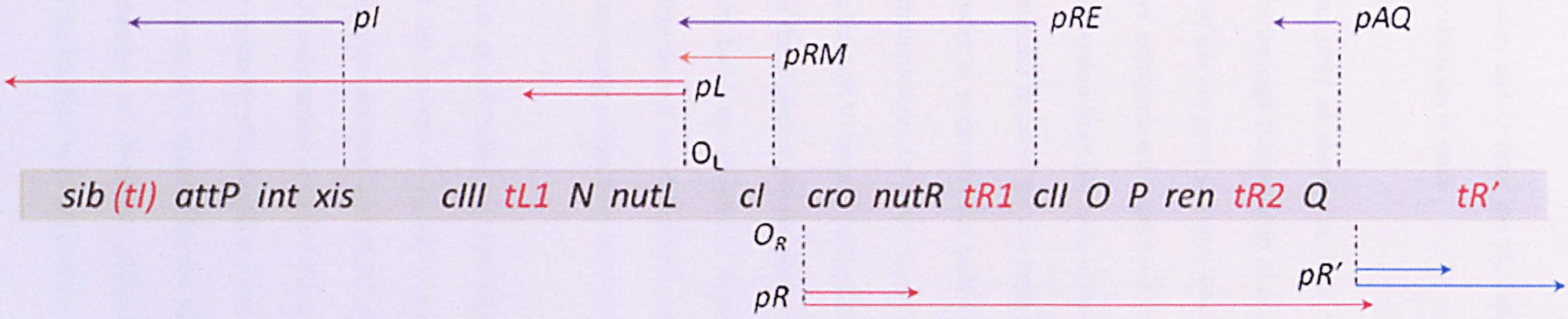


Figure 7. Gene and transcription map of bacteriophage λ Genes are shown in the shaded rectangle. Early transcription from the *pL* and *pR* promoters are shown as red arrows. The late transcript from *pR'* is indicated with blue arrows. The CII-activated *pI*, *pRE*, and *pAQ* transcripts are indicated with purple arrows. The *pRM* transcript activated by CI is an orange arrow. Transcription terminators, *t*, are shown as red letters among the genes. The *tI* terminator is indicated in parenthesis because it is contained within the larger *sib* processing site. The operators O_L and O_R where CI and Cro bind are shown next to the *pL* and *pR* promoters (Adapted from Court *et al.*, 2007).

control regulators and Q from the pL and pR promoters. It is at this point that the lysis/lysogeny decision is made.

In the vast majority of cases, the lytic life cycle is the preferred course for the infecting bacteriophage (Oppenheim *et al.*, 2005). During the lytic process, the early gene product of the cro gene acts as a weak repressor of the right operator region and as such allows uninterrupted expression from pR (Dodd *et al.*, 2005; Takeda *et al.*, 1977). The expression from pR leads to the build up of the regulator protein, Q, which is an anti-terminator required for the restructuring of the RNA polymerase elongation complex in order to overcome the termination sequence at tR' (Marr *et al.*, 2001). Bypassing termination results in the transcription of the late genes and commitment to lysis. Passage through to lysis is delayed in lambda by 2 distinct mechanisms. Firstly, the location of the Q gene at the periphery of the delayed early gene transcript leads to belated protein build up. Secondly, significant Q concentrations must accumulate prior to the detection of any discernable activity *in vivo* (Kobiler *et al.*, 2005; Little, 2005), thus suggesting a high dose threshold for effective function.

In a proportion of infections, the pathway leading to a repressed lysogenic state is favoured and this process is regulated initially by cII and cIII (Oppenheim *et al.*, 2005). Both proteins are produced as part of the delayed early gene product facilitated by the N-RNA polymerase complex (Greenblatt *et al.*, 1998). cII is a transcriptional regulator that exerts its effect at three sites within the page genome: pI, pRE and paQ. Transcription from pI is responsible for phage integration into the host chromosome via the production of integrase molecules; pRE activation permits the initial expression of the lambda repressor protein cI and paQ transcription regulates levels of

Q (Oppenheim *et al.*, 2005). The mechanism for cII regulation of Q is still unclear but it has been postulated that it may be due to an anti-sense translational inhibition effect (Kobiler *et al.*, 2005; Wagner & Simons, 1994). cII is also a highly unstable protein *in vivo* and its degradation is catalysed by the host-encoded protease FtsH (Shotland *et al.*, 1997). cIII is an alternative substrate to cII and therefore acts as a competitive inhibitor of FtsH (Kobiler *et al.*, 2007), and as a consequence provides a degree of protection.

cI comprises an N-terminal DNA binding domain and C-terminal oligomerisation domain (Sauer *et al.*, 1979) and, once produced, the cI repressor protein is vital for the establishment and maintenance of lysogeny (Ptashne, 2004). Primarily the protein binds to $O_L1/2$ and $O_R1/2$ of the lambda immunity region as co-operative tetramers (Dodd *et al.*, 2004; Ptashne, 2004). Binding at these 2 operators, located 2.3 kb apart, results in the formation of a DNA loop anchored by a repressor protein octomer (Fig. 8) (Oppenheim *et al.*, 2005). The formation of this loop simultaneously inhibits transcription from pL and pR and enhances transcription from pRM, which effectively shuts down transcription of all phage encoded genes except its own. Occupation of the final O_L and O_R sites further stabilises the loop structure and prevents transcription from pRM (Maurer *et al.*, 1980). This repressor-operator interaction means that the levels of cI are auto-regulated by binding of the 3 operator sites; occupation of operator sites 1 and 2 acts to positively regulate the expression of cI from the promoter pRM while binding to the third site down-regulates expression (Maurer *et al.*, 1980). cI therefore effectively auto-regulates its own expression. During the creation and preservation of the lysogenic state, the cI mediated repression of pR is crucial for the prevention of Q achieving the threshold required for lytic infection to

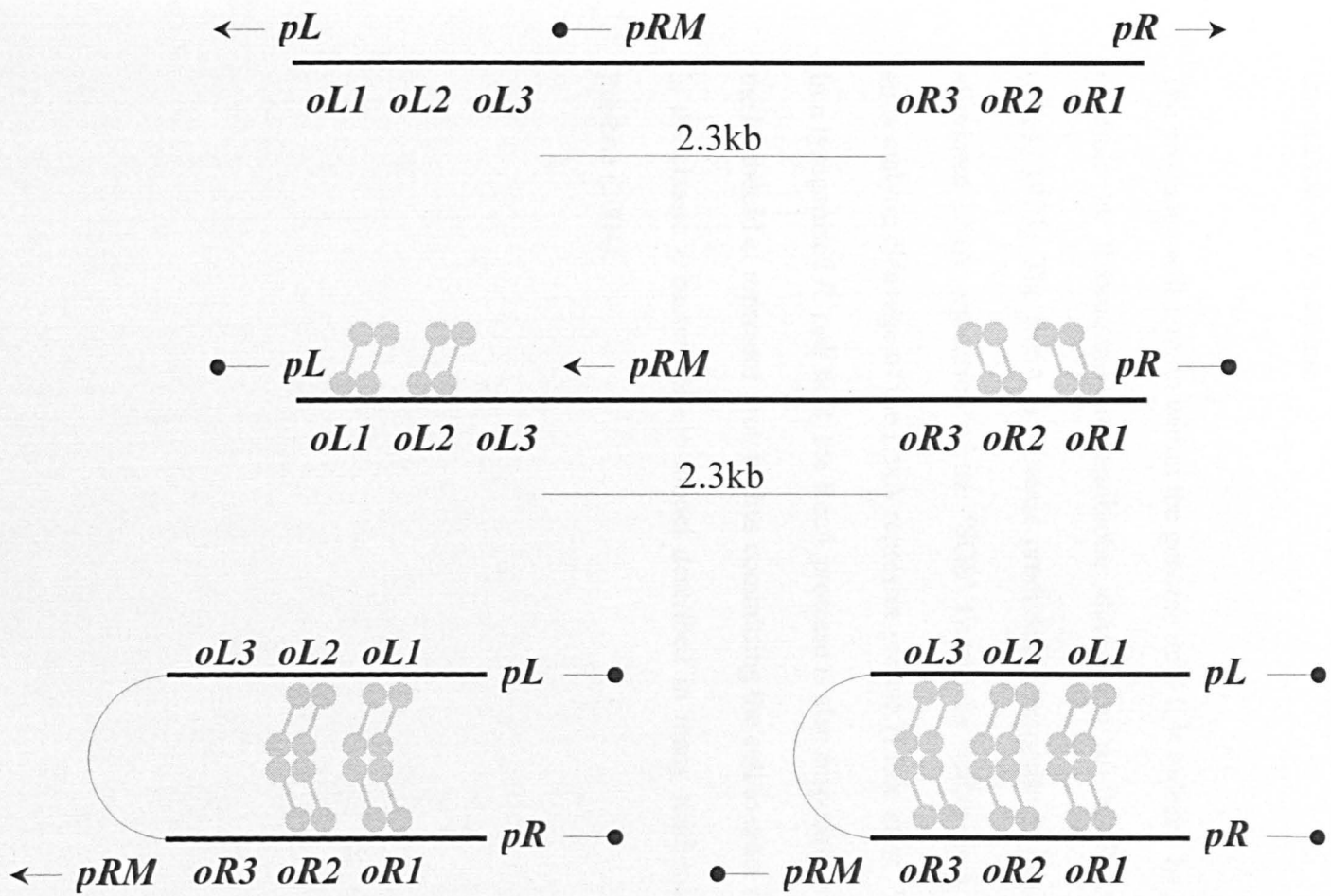


Figure 8. CI repressor protein regulation of transcription from *pRM* in bacteriophage lambda Configuration of bacteriophage lambda right and left operator regions are illustrated, with expression from the pRM promoter leading to CI synthesis indicated. Consecutive CI repressor dimer binding as well as subsequent dimer interactions leading to hairpin loop formation and termination of transcription from *pRM* is also shown (Adapted from Oppenheim *et al.*, 2005).

occur. The maintenance of cytoplasmic cI during lysogeny is also responsible for immunity to super-infecting homo-immune bacteriophage (Ptashne, 2004). Upon infection of an already lysogenised host, any phage sensitive to lambda immunity will be repressed by the immediate binding of cI to the operator region thus preventing integration or replication.

The prophage will remain within the genome until it is induced by DNA mutation, for example by fluoroquinalone antibiotic stimulation of the RecA 'SOS' response (Clark, 1973). The RecA is protease produced in response to DNA damage and this alleviates LexA repression of the 'SOS' system by acting as a co-protease in the autocatalytic cleavage of the LexA repressor protein (Little *et al.*, 1980; Little, 1991). In a lysogenised *E. coli* host, the RecA protease is also responsible for the cleavage of the lambdoid cI repressor protein thus committing the cell to enter the lytic cycle. This is the classic λ bacteriophage model described in many textbooks and in detail by Ptashne (2004).

1.6 Tyrosine Recombinases

Integrases belong to a family of tyrosine recombinases involved in site-specific recombination. The family has grown to include gene cassette mobilising enzymes of integrons, the yeast Flp enzyme responsible for insertion of DNA segments and bacterial XerC/D and P1 Cre that resolve dimeric circular replicons to monomers (Chen & Rice, 2003). All members of the family possess a highly conserved catalytic domain structure, closely resembling that of Topoisomerase type 1b, including the eponymous reactive tyrosine residue (Esposito & Scocca, 1997; Nunes-Duby *et al.*, 1998).

The first and most extensively studied tyrosine recombinase *in vitro* is the integrase (Int) of the temperate bacteriophage lambda. During bacteriophage genome integration into the host chromosome, λ integrase recognises two DNA attachment sites, *attB* and *attP* (Campbell, 1992); the former is present in the bacterial genome and the latter in the phage genome. λ Int aligns the complimentary *attP* and *attB* crossover sites (O) and recombines them to yield an integrated prophage flanked by *attL* and *attR* sequences (Groth & Calos, 2004) (Fig. 9). Upon induction into the lytic life cycle, the prophage is excised by the same integrase enzyme, which essentially regenerates the original recombination substrates (Landy, 1989). The mechanism for this involves a process of pair-wise DNA strand breakage and sequential exchange to create and subsequently resolve a four way branched Holliday junction intermediate (Biswas *et al.*, 2005). The reaction requires no high-energy co-factor and results in the net loss of no phosphodiester bonds (Grindley *et al.*, 2006). Non-integrase tyrosine recombinases such as Cre and Flp utilise two identical sites of up to 50 bp in length

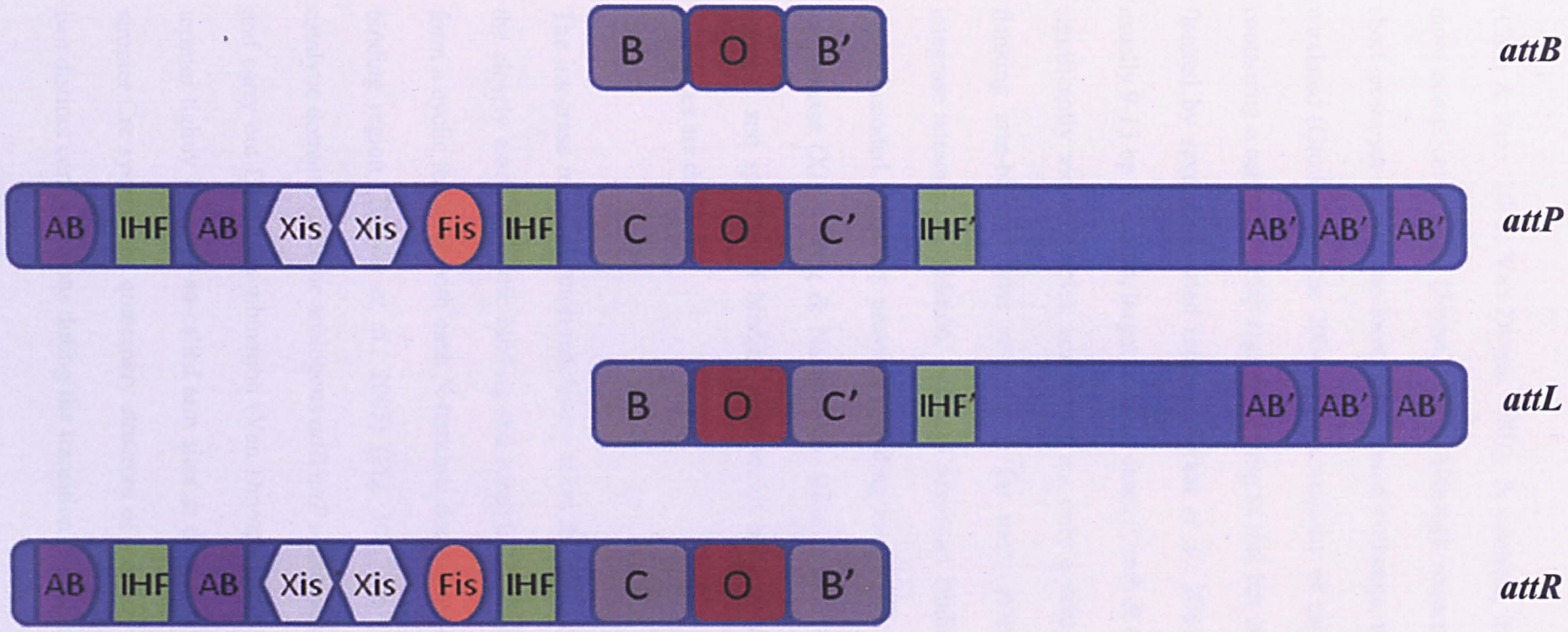


Figure 9. Lambda *Att* site structure A prime symbol (′) differentiates the right side of the *att* site from the left. Grey boxes represent core type binding sites (B and B′ in *attB*, C and C′ in *attP*). The overlap region where crossing-over occurs is represented by the letter O. Purple symbols labelled AB or AB′ represent arm-type binding sites for λ integrase. Binding sites for IHF and for cofactors Xis and Fis are represented by the appropriately labelled symbols. (Adapted from Groth & Calos, 2004).

(Chen & Rice, 2003; Van Duyne, 2001). In contrast, integrase DNA recognition is more complicated (Van Duyne, 2005). Although sequence homology is shared in a short crossover region, the location of strand exchange, the remainder of the sites are unrelated (Groth & Calos, 2004). In the majority of cases, *attB* is relatively simple containing a central overlap region, varying in size but usually approximately 6-8 bp, flanked by several inverted repeats (Sarkar *et al.*, 2001). This core-binding site is usually 9-13 bp in length, larger in some cases (Groth & Calos, 2004). The *attP* site is significantly more complex containing not only a core overlap but also numerous flanking arm-binding sites responsible for more extensive interactions with the integrase tetramer. In addition, *attP* also possesses binding sites for various host and phage-encoded accessory proteins including integration host factor (IHF), (Fis) and excisionase (Xis) (Craig & Nash, 1984; Ross *et al.*, 1979; Yin *et al.*, 1985). The quantity and spacing of binding sites varies between different integrases, and the sequences are distinct.

The integrase recombination machinery takes the form of a tetrad complex in which the closely associated core binding and catalytic domains of the four Int monomers form a cyclic tetramer with each N-terminal domain residing over each adjacent core binding region (Biswas *et al.*, 2005) (Fig. 10). The C-terminal core binding and catalytic domains hold the analogous *attB/attP* sequences in a square planar formation and carry out DNA recombination (Van Duyne, 2005). Meanwhile the integrase N-termini tightly bind the so-called arm sites at the peripheries of the *attP* site. In the simpler Cre system, the quaternary structure of the enzymatic core switches between two distinct conformations during the transition from first to second strand exchange,

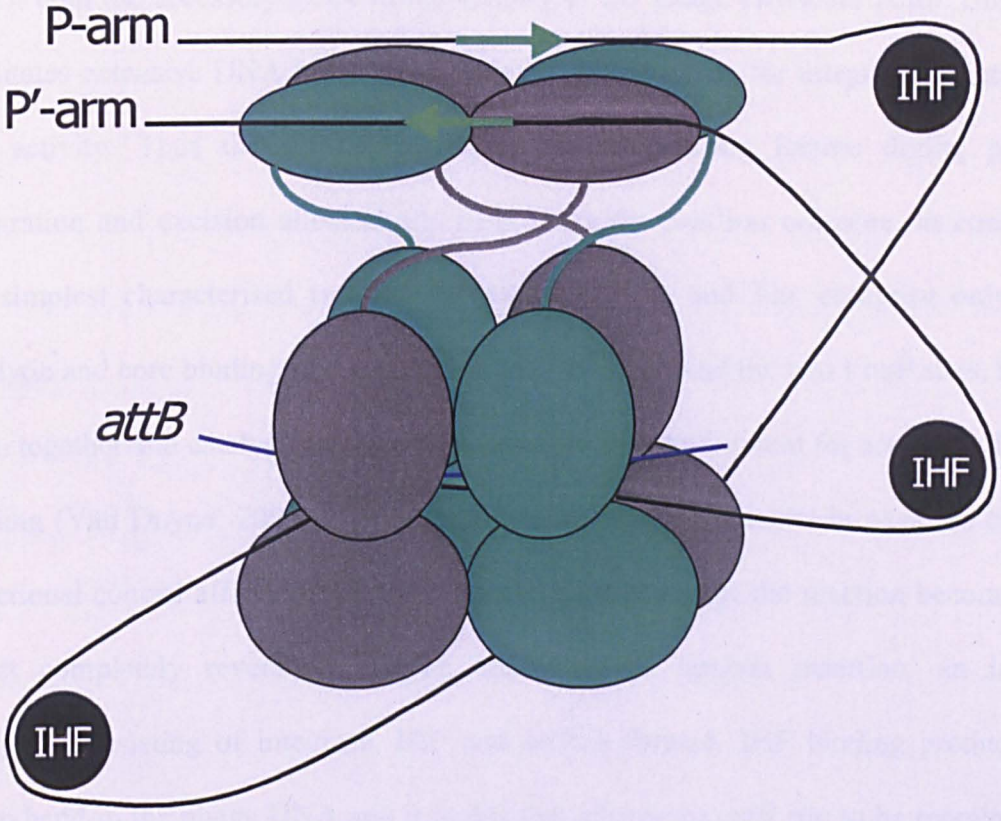


Figure 10. Protein-DNA interaction during bacteriophage lambda integration Extensive *attP* DNA bending required for intasome formation is facilitated by host encoded factors such as integration host factor (IHF) and Fis. A phage-encoded protein, excisionase (Xis), also helps to determine the conformation during prophage excison from the host chromosome. Core binding and catalytic domains hold the complimentary *attB/P* sites together whilst the arm-binding sites tightly bind peripheral *attP* sequences. (Adapted from Biswas *et al.*, 2005; Van Duyne, 2005)

resulting in only two Int molecules being simultaneously active at any one time (Gopaul *et al.*, 1998). The lambda integrase machinery, on the other hand, is far more rigid and thus is not able to switch between conformations once the arm-binding sites are bound (Biswas *et al.*, 2005). The lambda integrase intasome forms an elongated symmetrical parallelogram; the conformation of which is determined by the presence of accessory factors, most notably phage encoded Xis and bacterial encoded IHF, which bind the accessory DNA in the vicinity of the phage crossover point. Binding facilitates extensive DNA bending, an essential pre-requisite for integrase orientation and activity. Thus the differential expression of bending factors during phage integration and excision allosterically determines the reaction outcome. In contrast, the simplest characterised tyrosine recombinases, Cre and Flp, comprise only the catalytic and core binding domains which co-operate to bind the two LoxP sites, bring them together and catalyse recombination without the requirement for accessory DNA binding (Van Duyne, 2005). The lack of arm-binding sites results in negation of the directional control afforded by their presence. Consequently, the reaction becomes in effect completely reversible. During bacteriophage lambda insertion, an initial complex consisting of integrase, IHF and *attP* is formed. IHF binding produces a sharp bend in the phage DNA and it is this that allows the *attB* site to be recruited to the complex and enables recombination. Lambda integrase has been shown to bind with high affinity to the arm sites, but with low affinity to the core. Furthermore, λ Int and HK022 Int both recognise the same arm sites but require different overlap sequences to achieve integration. This suggests that although the core is essential for specificity and strand exchange, it is not vital for initial intasome formation.

1.7 Integrase Diversity

The N-terminal protein sequences of tyrosine recombinases are extremely diverse and it is very difficult to compare full-length sequences. The N-terminal is responsible for DNA binding and contributes to integration site specificity, which goes some way to explaining its divergent forms. In contrast, the C-terminal catalytic and core binding domains display a degree of homology and have therefore been the focus of investigation. Initially, the integrase family of tyrosine recombinases was defined on the basis of the alignment of seven sequences by Argos *et al.* (1986). This alignment resulted in the discovery of three highly conserved amino acid residues - a His-X-X-Arg motif and an invariant tyrosine. Subsequent analysis of an increased dataset of 28 integrases identified a fourth conserved arginine residue forming a RHRY formation (Abremski & Hoess, 1992). Furthermore, the alignment of 75 prokaryotic sequences identified 3 regions of heightened conservation surrounding the components of the RHRY motif- termed Box A, B and C; and allowed phylogentetic relationships to be established (Esposito & Scocca, 1997) (Fig. 11). Box A surrounds the first arginine (λ R^{212}) and increased homology is concentrated around the arginine residue, in excess of 80% for some amino acids. Box B contains the H-X-X-R motif although only the arginine is absolutely conserved in all known members of the family. In over 80% of sequences, the motif could be expanded to include 2 additional components, His-X-Leu-Arg-His. Box C flanks the catalytic tyrosine residue vital to enzymatic function and thus universally present in all tyrosine recombinases. Nunes-Duby *et al.* (1998) refined the regions of conservation to two areas termed Box I (λ A^{202}-G^{227}) and II (λ T^{306}-D^{344}) (Fig. 12A). Alignment of 105 sequences confirmed the prevalence of the RHRY motif but also noted the substitution of the histidine residue with one of four

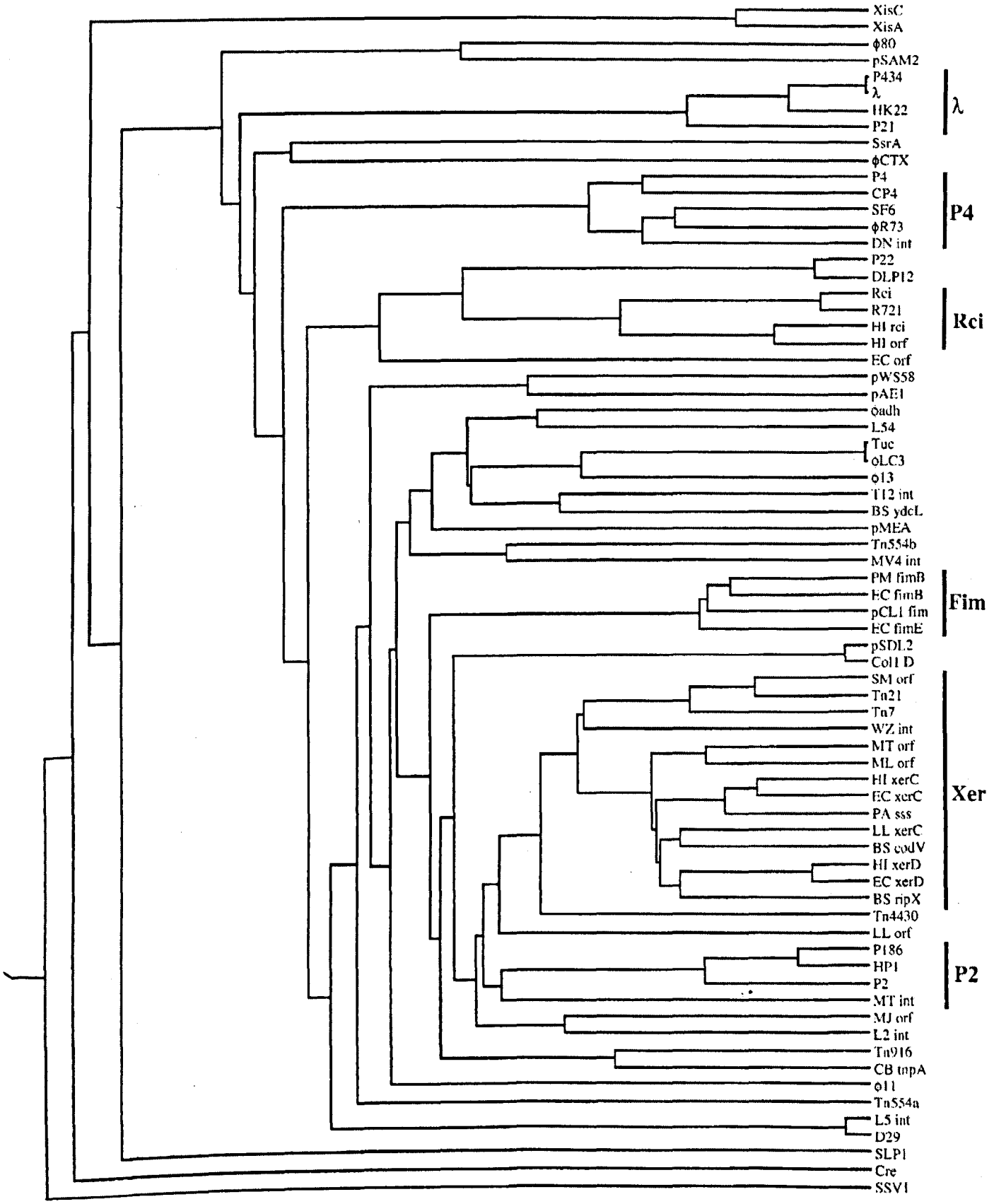


Figure 11. A phenogram depicting the evolutionary relationships between the active site domains of 71 prokaryotic tyrosine recombinases Constructed based on a best-fit alignment were calculated by PROTDIST and trees constructed with KITSCH. (Taken from Esposito & Scocca, 1997).

alternatives- arginine, lysine, asparagine or tyrosine, in eight integrases. Mutational studies supported the suggestion that the His residue is not a vital pre-requisite to catalytic activity by showing the retention of partial function after replacement (Nunes-Duby *et al.*, 1998). Comparison of the conserved regions with crystalline structure data allowed the positions of the amino acids of interest to be plotted onto the secondary structure of the extensively studied integrases Cre and lambda Int (Fig. 12B). Box I encompasses the λ integrase α-helices B and C that are the component elements of the protein core, whilst the conserved RHR residues are located on the protein surface at the site of DNA interaction (Nunes-Duby *et al.*, 1998).

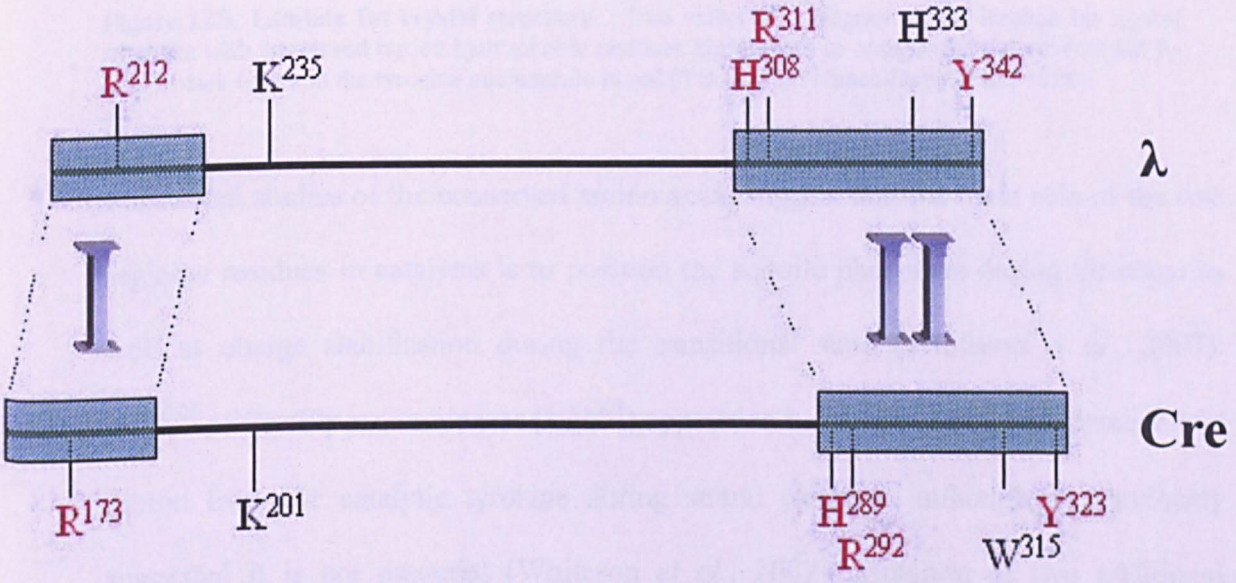


Figure 12A. Comparitive locations of conserved residues in lambda and Cre integrase catalytic domains (Nunes-Duby *et al.*, 1998)

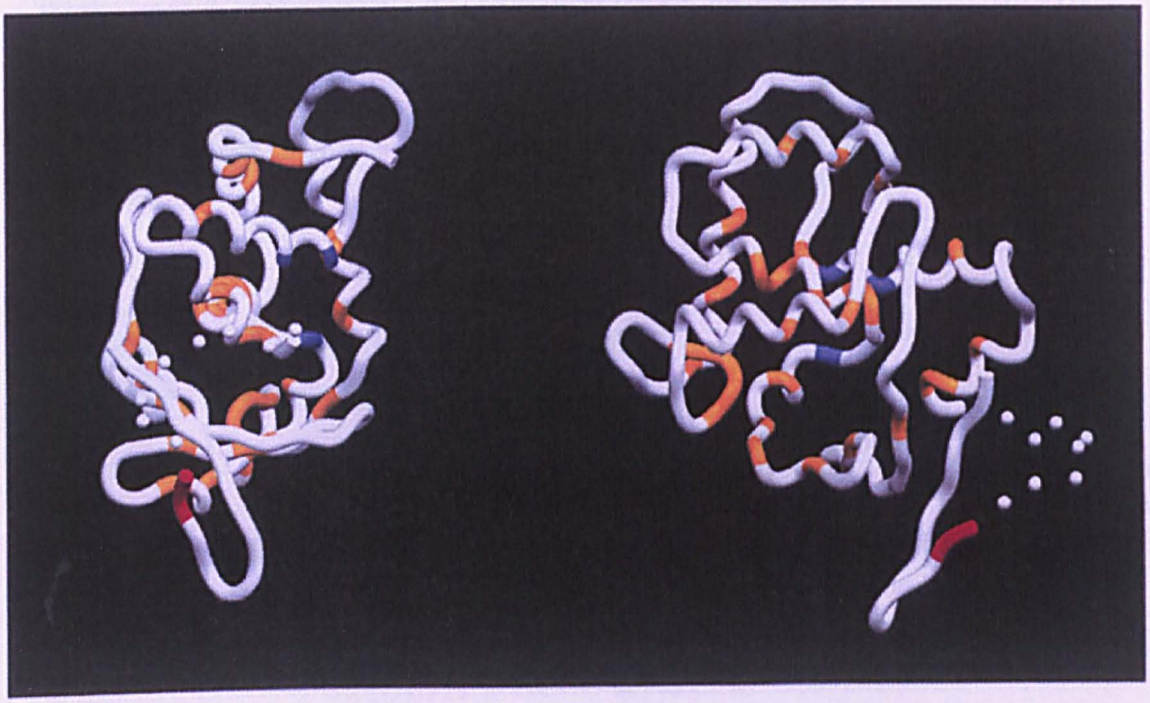


Figure 12B. Lambda Int crystal structure Two views of a diagram of the lambda Int crystal structure with conserved buried hydrophobic residues highlighted in orange, the conserved triad R-H-R in dark blue and the tyrosine nucleophile in red (Taken from Nunes-Duby *et al.*, 1998).

Functional studies of the conserved amino acids suggest that the main role of the two arginine residues in catalysis is to position the scissile phosphate during cleavage as well as charge stabilisation during the transitional state (Whiteson *et al.*, 2007). $H^{289/305}$ of Cre/Flp recombinases (λ H^{308}) appears to be involved in the acceptance of a proton from the catalytic tyrosine during strand cleavage, although as previously suggested it is not essential (Whiteson *et al.*, 2007). Mutation of two additional conserved residues that were identified has a far more profound effect on activity (Chen & Rice, 2003). Studies of *Vaccinia* topoisomerase have shown that a lysine (λ K^{235}) acts as a general acid during strand cleavage donating a proton to the 5' hydroxyl strand (Krogh & Shuman, 2000; Nagarajan *et al.*, 2005) (Fig. 13), whilst a

His/Trp (λ H^{333}/Cre W^{315}) acts to stabilise a reaction intermediate (Chen & Rice, 2003).

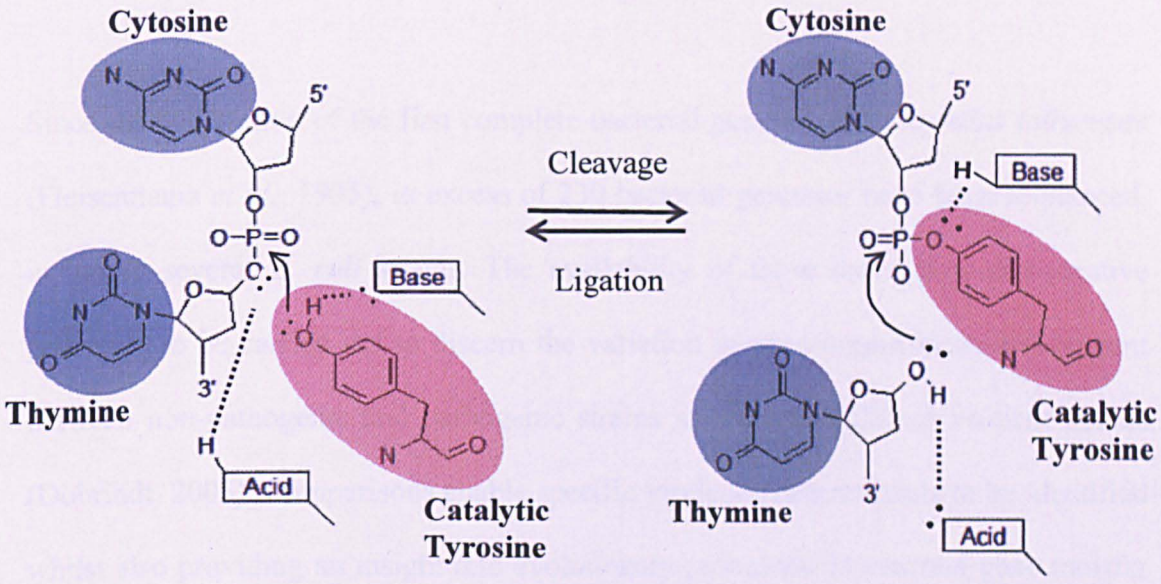


Figure 13. Schematic diagram of general acid and general base catalysis in DNA cleavage and ligation reactions In the DNA cleavage reaction catalyzed by tyrosine recombinases and type IB topoisomerases, a general base presumably accepts a proton from the attacking tyrosine's hydroxyl group, while a general acid donates a proton to the leaving 5'-bridging oxygen on the scissile phosphate, resulting in the formation of a 3'-phosphotyrosyl bond and a free 5'-OH group. This protonation and deprotonation process is reversed in the ligation reaction. Other active site residues may help stabilize the transition state (not shown). Curved arrows represent ligation reactions whilst dotted lines represent donation or receipt of a proton by a general acid or base (Adapted from Chen & Rice, 2003).

1.8 Bacteriophage Contribution to Bacterial Evolution and the Emergence of Novel Pathogens.

Since the completion of the first complete bacterial genome, *Haemophilus influenzae* (Fleischmann *et al.*, 1995), in excess of 230 bacterial genomes have been sequenced, including several *E. coli* strains. The availability of these data allow comparative genomics to be carried out to discern the variation in gene organisation and content between non-pathogenic and pathogenic strains as well as different virulent strains (Dobrindt, 2005). Comparisons enable specific virulence determinants to be identified whilst also providing an insight into evolutionary processes. Horizontal gene transfer has a major influence on bacterial evolution leading to the presence of distinct islands of DNA interspersed with the conserved genomic backbone of different bacterial lineages (Dobrindt, 2005). It has been estimated that 18% of the sequenced *E. coli* K-12 strain MG1655 genome is made up of horizontally acquired genetic material (Lawrence & Ochman, 1998). The major components involved in horizontal gene transfer are accessory genetic elements such as plasmids, transposons, insertion sequence elements, genomic islands and both inducible and remnant prophage. Comparison of the complete *E. coli* genomes reveal a core chromosome of approximately 4 Mbp, however overall genome size varies from 4.64 to 5.59 Mbp (Blattner *et al.*, 1997; Hayashi *et al.*, 2001). Indeed the pathogenic EHEC strains EDL933 and Sakai both contain over one thousand additional ORFs compared to *E. coli* K-12 MG1655 (Blattner *et al.*, 1997; Hayashi *et al.*, 2001; Perna *et al.*, 1998). A significant proportion of this may be attributed to the increased presence of resident prophage or remnant prophage, 24, 16 and 10 found in Sakai, EDL933 and MG1655 respectively (Table 1). Almost 40% of the O157 specific elements of EDL933 were

found to be located within the various remnant prophages or the inducible bacteriophage 933W, the latter of which carries the gene encoding Shiga-like toxin (Perna *et al.*, 1998).

Table 1. Bacteriophage-related chromosomal regions of the published *E. coli* K-12 MG1655 genome sequence plus the two genome sequences of *E. coli* O157 strains, EDL933 and Sakai (Dobrindt, 2005)

***E. coli* strain: Bacteriophage designation**	**Length [bp]**	**Chromosomal insertion site**	**Characteristics**
E. coli K-12 strain MG1655			
CP4-6	34,308	*thrW*	CP4 element
DLP12	21,302	*argU*	Lambdoid phage
e14	15,204	Between *icdA* and *mcrA*	—
Rac	23,060	*ydaO*	Lambdoid phage
Qin	16,381	Between b1543 and b1582	Lambdoid phage
CP4-44	12,873	Between b1955 and*yeeX*	CP4 element, contains *flu*
CPZ-55	6,782	Between *eutB* and *eutA*	—
CP4-57	22,030	*ssrA*	CP4 element
KpLE1	10,216	*argW*	Shares the same integrase gene with Sp16
KpLE2	40,071	*leuX*	Shares a part of the integrase gene with SpLE6
E. coli O157:H7 strain EDL933			
CP-933H	10,586	*thrW*	Lambdoid phage
CP-933I	12,895	*thrW*	P4-like phage, integrated in tandem with CP-933H
CP-933K	38,588	Between *ybhC* and *ybhB*	Lambdoid phage
CP-933M	45,244	*serT*	Lambdoid phage
BP-933W	61,663	*wrbA*	Lambdoid phage, contains *stx*2 and three tRNA loci
CP-933N	47,315	*potB*	Lambdoid phage, contains two tRNA loci
CP-933C	15,227	Between *ycfD* and *phoQ*	Similar to Sp7
CP-933X	54,000	Between *icdA* and *minE*	Lambdoid phage
CP-933O	80,826	*yciD*	Lambdoid phage, contains two tRNA loci
CP-933R	49,797	*ydaO*	Lambdoid phage, contains two tRNA loci
CP-933P	57,984	b1582	Lambdoid phage, contains three tRNA loci
CP-933T	21,120	*leuZ*	P2-like phage
CP-933U	45,177	*serU*	Lambdoid phage, contains three tRNA loci
CP-933V	48,916	*yehV*	Lambdoid phage, contains *stx*1
CP-933Y	21,681	*ssrA*	Lambdoid phage
E. coli O157:H7 strain Sakai			
Sp1	10,586	*thrW*	Lambdoid phage
Sp2	12,887	*thrW*	P4-like phage, integrated in tandem with Sp1
Sp3	38,586	Between *ybhC* and *ybhB*	Lambdoid phage
Sp4	49,650	*serT*	Lambdoid phage, contains three tRNA loci
Sp5	62,708	Between *yccJ* and *ycdG*	Lambdoid phage, contains *stx*2 and

			three tRNA loci
Sp6	48,423	*potB*	Lambdoid phage
Sp7	15,463	Between *ycfD* and *phoQ*	—
Sp8	46,897	Between *icdA* and *minE*	Lambdoid phage
Sp9	58,175	*yciD*	Lambdoid phage, contains three tRNA loci
Sp10	51,112	*ydaO*	Lambdoid phage, contains three tRNA loci
Sp11	45,778	Between b1543 and b1582	Lambdoid phage, contains three tRNA loci
Sp12	46,142	Between b1543 and b1582	Lambdoid phage, integrated in tandem with Sp11, contains three tRNA loci
Sp13	21,120	*leuZ*	P2-like phage
Sp14	44,029	*serU*	Lambdoid phage, contains three tRNA loci
Sp15	47,879	*yehV*	Lambdoid phage, contains *stx*1
Sp16	8,551	*argW*	P22-like phage
Sp17	24,199	*ssrA*	Lambdoid phage
Sp18	38,759	Between ECs4941 and ECs4999	Mu-like phage
SpLE1	86,249	*serX*	CP4-like element
SpLE2	13,459	Between b1955 and *yeeX*	Corresponds to CP4-44 of K-12
SpLE3	23,454	*pheV*	—
SpLE4	43,450	*selC*	CP4-like element, contains the LEE locus
SpLE5	10,235	*leuX*	—
SpLE6	34,148	*leuX*	Integrated in tandem with SpLE5

Although clearly temperate bacteriophages can play a pivotal role in the transfer of virulence determinants throughout bacterial populations, the potential is also available for recombination between phage resident in a single host to yield novel pathogenicity factors or novel phage with an altered host range. Within bacteriophage populations, the modular theory of evolution describes the exchange of regions of the genome, usually of comparable or similar function, between integrated phages, whether they are inducible or remnant (Botstein, 1980). This process of modular substitution provides phages with substantial genomic diversity and in turn can lead to the emergence of novel pathogens (Saunders *et al.*, 2001; Brüssow *et al.*, 2004).

For various pathogenic bacteria the principal virulence determinant is encoded on a resident prophage whilst closely related non-pathogenic strains lack the bacteriophage genes in question. Moreover, in several cases subsequent curing of the prophage from

the pathogen has been shown to alleviate virulence completely (Brussow *et al.*, 2004). Therefore the ability of bacteriophages to convert the phenotype of the host bacterium is one the major factors driving the evolution of novel pathogens (Saunders *et al.*, 2001) (Table 2). Lysogenic infection of non-toxigenic *Corynebacterium diptheriae* with corynebacteriophage results in the production of diphtheria toxin under low iron conditions and thus increases the pathogenicity of the bacterium (Freeman, 1951). Likewise, infection of endogenous gut bacteria by Stx-phage can lead to conversion of commensal intestinal microflora to virulent pathotypes able to cause severe disease (Brabban *et al.*, 2005; Brussow *et al.*, 2004). Similar infection of environmental *Vibrio* spp. with CTXΦ, from *Vibrio cholerae*, yielded strains able to produce and secrete toxin (Faruque *et al.*, 2003a; Faruque *et al.*, 2003b; Michel *et al.*, 1995). Other pathogenic bacteria possess an arsenal of diverse virulence factors, all of which are important for clinical manifestation of infection, and thus it is more difficult to ascertain the specific impact of phage-encoded factors. *Streptococcus pyrogenes* can be converted to a toxin-producing strain in the presence of a toxin-encoding phage, however no significant increase in virulence is observed despite the fact that direct injection of toxin from culture supernatant resulted in typical symptoms in laboratory animals (Brussow *et al.*, 2004). Generally, it can be stated that bacteriophage conversion of a host organism can significantly alter its virulence characteristics, although the overall impact of phage-encoded factors varies greatly from strain to strain.

In addition to toxins, bacteriophages encode a number of other virulence factors beneficial to the host (Table 2). For example, it was first shown in 1971 that O-antigen genes were encoded by phage epsilon in *Salmonella* spp. (Wright, 1971) and

O-antigen modification systems have been detailed on a variety of bacteriophages subsequently. Modification of the antigens leads to greater diversity and aids in evasion of the host immune response (Boyd & Brussow, 2002).

Table 2. Common themes among phage-encoded virulence factors (Brussow *et al.*, 2004) and updated from Boyd & Brussow (2002)

Protein	Gene	Bacterial host
Extracellular toxins		
Diphtheria toxin	*tox*	*Cory. diphtheriae*
Neurotoxin	C1	*Clost. botulinum*
Shiga toxins	*stx1, stx2*	*E. coli*
Enterohaemolysin	*hly2*	*E. coli*
Cytotoxin	*ctx*	*P. aeruginosa*
Enterotoxin	*see, sel*	*Staph. aureus*
Enterotoxin P	*sep*	*Staph. aureus*
Enterotoxin A	*entA*	*Staph. aureus*
Enterotoxin A	*sea*	*Staph. aureus*
Exfoliative toxin A	*eta*	*Staph. aureus*
Toxin type A	*speA*	*Staph. pyogenes*
Toxin type C	*speC*	*Strep. pyogenes*
Cholera toxin	*ctxAB*	*V. cholerae*
Leukocidin	*pvl*	*Staph. aureus*
Superantigens	*speA1, speA3, speC, speI, speH, speM, speL, speK, ssa*	*Strep. pyogenes*
Cytolethal distending toxin	*cdt*	*E. coli*
Proteins altering antigenicity		
Membrane proteins	Mu-like	*N. meningitidis*
Glucosylation	*rfb*	*Salm. enterica*
Glucosylation	*gtr*	*Salm. enterica*
O-antigen acetylase	*oac*	*Shig. flexneri*
Glucosyl transferase	*gtrII*	*Shig. flexneri*
Effector proteins involved in invasion		
Type III effector	*sopE*	*Sal. enterica*
Type III effector	*sseI (gtgB)*	*Sal. enterica*
Type III effector	*sspH1*	*Sal. enterica*
Enzymes		
Superoxide dismutase	*sodC*	*E. coli* O157
Superoxide dismutase	*sodC-I*	*Sal. enterica*
Superoxide dismutase	*sodC-III*	*Sal. enterica*
Neuraminidase	*nanH*	*Sal. enterica*
Hyaluronidase	*hylP*	*Strep. pyogenes*
Leukocidin	*pvl*	*Staph. aureus*

Staphylokinase	*sak*	*Staph. aureus*
Phospholipase	*sla*	*Strep. pyogenes*
DNase/streptodornase	*sdn, sda*	*Strep. pyogenes*
Serum resistance		
OMP[b]	*bor*	*E. coli*
OMP	*eib*	*E. coli*
Adhesions for bacterial host attachment		
Vir	*vir*	*M. arthritidis*
Phage coat proteins	*pblA, pblB*	*Strep. mitis*
Others		
Mitogenic factors	*mf2, mf3, mf4*	*Strep. pyogenes*
Mitogenic factor	*toxA*	*P. multocida*
Mitogenic factor	Unnamed	*Strep. canis*
Virulence	*gtgE*	*Sal. enterica*
Antivirulence	*grvA*	*Sal. enterica*

1.9 Aims

Clearly lambdoid bacteriophages are extremely heterogeneous and are able to contribute significantly to bacterial fitness and pathogenicity through the carriage of a multitude of genes encoding a diverse range of virulence determinants. In addition, the extensive host range of lambdoid phage means that the potential for dissemination of these genes throughout heterogeneous bacterial populations is immense. In particular, Stx-phages are a major contributory factor in *E. coli* pathogenesis, and the subsequent sporadic outbreaks of STEC infections since 1982 (Riley *et al.*, 1983). The rapid emergence of STEC as an important pathogen highlights the potential for bacteriophage driven evolution. This thesis primarily addresses elucidation of the primary mechanisms underpinning aspects of Stx-phage biology that contribute to heterogeneity and evolution. Allison *et al.* (2003) found that, contrary to the lambda immunity model, it is possible for two isogenic phages ($\Phi24_B$) to sequentially infect and integrate into the same host. This thesis aims to address the underlying mechanisms of this novel observation:

- The $\Phi24_B$ integrase gene will be identified and sequenced.
- The diversity of integrases, within phage populations induced from a collection of clinical and environmental STEC isolates, will be determined.
- The chromosomal insertion sites of $\Phi24_B$ in a laboratory *E. coli* strain, MC1061, will be elucidated.
- The principal mechanisms underlying the phenomenon of multiple lysogeny will be investigated.

Chapter 2: General Materials and Methods

Table 3. Growth media, supplements and buffers

	Ingredients
Agar Plates	3.8% (w/v) Luria Burtani Agar
LB Growth Media	2.5% (w/v) Luria Burtani Broth
Phage Buffer	2.5% (w/v) Luria Burtani Broth, 0.01 M $CaCl_2$
TAE (50x)	2 M Tris base, 1 M glacial acetic acid, 0.05 M EDTA
TE	10 mM Tris-HCl, 1 mM EDTA, pH 8.0
EDTA (pH 8.0)	0.5 M Di-sodium ethylenediamine tetra-acetate, adjust pH with NaOH
SM Buffer	100 mM NaCl, 8 mM $MgSO_4$, 50mM Tris-HCl, pH 7.5
Spectinomycin (1000x)	100 mg ml^{-1} Spectinomycin dihydrochloride
Kanamycin (1000x)	50 mg ml^{-1} Kanamycin disulfate
Streptomycin (1000x)	50 mg ml^{-1} Streptomycin sulphate
Tetracycline (1000x)	20 mg ml^{-1} Tetracycline
Chloramphenicol (1000x)	50 mg ml^{-1} Chloramphenicol
Ampicillin (1000x)	100 mg ml^{-1} Ampicillin
Norfloxacin (1000x)	1 mg ml^{-1} Norfloxacin
Rifampicin	34 mg ml^{-1} Rifampicin
Depurination Solution	250 mM HCl
Denaturation Solution	0.5 M NaOH, 1.5 M NaCl
Neutralization Solution	0.5 M Tris-HCl, pH 7.5; 3 M NaCl
20 x SSC	3 M NaCl, 300 mM sodium citrate, pH 7.0
2 x SSC	300 mM NaCl, 30 mM sodium citrate, pH 7.0
Pre-Hybridization Solution	5 x SSC, 50% (v/v) deionised formamide, 0.1% (w/v) N-lauroylsarcosine, 0.02% SDS, 2% (w/v) Blocking Reagant
Hybridization Solution	500 ng DIG-labelled probe, 20 ml pre-hybridization solution
2 x Wash Solution	2 x SSC, 0.1% (w/v) SDS
0.5 x Wash Solution	0.5 x SSC, 0.1% (w/v) SDS
Washing Buffer	100 mM maleic acid, 150 mM NaCl, pH 7.5; 0.3% (v/v) Tween 20
Maleic Acid Buffer	100 mM maleic acid, 150 mM NaCl, pH 7.5
Blocking Solution	1% (w/v) blocking reagent (Roche diagnostics ltd, Lewes, UK), maleic acid buffer
Antibody Solution	Anti-DIG-AP fragments diluted 1:20 000 in blocking solution
Detection Buffer	100 mM Tris, 100mM EDTA; pH 8.0
Substrate Solution	CDP-Star diluted 1:100 in detection buffer

NB All Solutions made up in distilled H_2O except Chloramphenicol and Rifampicin (both ethanol).

2.1 Bacterial strains, bacteriophages and media

Routinely used media, antibiotics and buffers are listed in Table 3. Bacterial strains, bacteriophages and plasmids used in this research are listed in Table 4. *E. coli* K12 strain MC1061-Rif was the host chosen for routine bacteriophage propagation and lysogen production. JM109-λ pir was the propagation strain for suicide plasmids, while Invitrogen One Shot TOP10 cells were used for all other plasmids. The bacterial strains were routinely cultured in phage buffer (2.5% (w/v) Luria Bertani Broth (Merck KGaA, Darmstadt, Germany) plus 0.01M $CaCl_2$) with or without 1.5% (w/v) agar (Merck KGaA, Darmstadt, Germany). Bacteriophage suspensions were routinely stored in phage buffer at 4°C. Where appropriate, the following antibiotics were used: rifampicin (150 μg ml^{-1}), norfloxacin (1 μg ml^{-1}), ampicillin (100 μg ml^{-1}), kanamycin (50 μg ml^{-1}), chloramphenicol (50 μg ml^{-1}), spectinomycin (100 μg ml^{-1}) and tetracycline (10 μg ml^{-1}).

Table 4. Bacterial Strains, bacteriophages and plasmids

Strain/Plasmid	Relevant Characteristics	Reference
MC1061-Rif	K12 derivative, Rif^R	(James *et al.*, 2001)
MC1061Δ*intS*	MC1061 *intS*	This research
JM109-λ pir	K12 derivative, λ pir^+	(Penfold & Pemberton, 1992)
TOP10 One Shot Cells	Chemically Competent Cells	Invitrogen Ltd, Paisley, UK
DM1187	K12 derivative, constitutive RecA expression	
WG5-Rif	*E. coli* C derivative, constitutive RecA expression, Rif^R	(Grabow & Coubrough, 1986)
$\Phi 24_B$::Cat	Stx phage, $stxA_2$ΔCAT	(Allison *et al.*, 2003)
$\Phi 24_B$::Kan	Stx phage, $stxA_2$Δ*aph3*	(Allison *et al.*, 2003)
$\Phi 24_B$::Spec	Stx phage, $stxA_2$Δ*aadA*	This research
NTP707	Φ933W *stx* operon	(Willshaw *et al.*, 1987)
pKT230	Kan^R, low copy number	(Bagdasarian *et al.*, 1981)
pPCMF*intS*comp	pKT230 carrying *intS*	This research
pJP5603	Kan^R, R6K origin	(Penfold & Pemberton, 1992)
pPCMF*intS*	pJP5603 carrying *intS* interrupted by *aadA*	This research
pKNG101	Sm^R, *sacB*, R6K origin	(Kaniga *et al.*, 1991)
pCR-Blunt	Blunt ended cloning vector, lacZ α, Kan^R, *ccdB*	Invitrogen Ltd, Paisley, UK
pCR2.1-TOPO	TA overhang cloning vector, lacZ α, Kan^R, Amp^R	Invitrogen Ltd, Paisley, UK

2.2 Microbiological Methods

2.2.1 Lysogen Production

Host bacteria were grown to an optical density of OD_{600} 0.5 in phage buffer at 37°C with shaking. Aliquots of these cultures were mixed with phage suspensions to a multiplicity of infection (MOI) of 0.1 in a final volume of 100 µl prior to incubation at 37 °C. After 25 min, the mixture was spread plated onto selective media (2.5% (w/v) Luria Burtani Agar), containing the appropriate antibiotic, and incubated overnight at 37°C before the colonies were identified and counted.

2.2.2 Norfloxacin Induction of Lysis

Phage stocks were produced by the addition of 10 µl of the fluoroquinolone antibiotic Norfloxacin (1 mg ml^{-1}) to a 10 ml mid-exponential growth phase lysogen culture and incubation for 1 h at 37 °C. 1 ml of the culture was sub-cultured into 10 ml LB broth and allowed to recover for 2 h. The bacterial debris and whole cells were removed by filtration through a 0.2µm pore diameter filter (Millipore UK Ltd, Watford, UK). The phage titre was determined by plaque assay.

2.2.3 Bacteriophage enumeration

Phage lysate titres were calculated by plaque assay using a *recA*441 mutant of laboratory *E. coli* strain K-12, DM1187, which constitutively expresses RecA. 100 µl of mid-exponential DM1187 culture was added to 500 µl serial dilutions of phage

stock (10^{-1} to 10^{-9}) and incubated for 25 min at 37 °C. The infection mix was then added to 5 ml of molten top agar (0.4% w/v Lab M Agar No.1, 2.5% LB Broth), pre-warmed to 50°C, and poured onto bottom agar plates (0.7% Agar No.1, 2.5% LB Agar). After overnight incubation at 37°C dilutions containing between 20 and 200 plaques were counted and used to estimate the total number of viable phage particles in the original phage suspension.

2.3 Molecular Biology Methods

2.3.1 Agarose Gel Electrophoresis

DNA was routinely analysed and quantified by electrophoresis on a 0.75% TAE agarose gel (40 mM Tris base, 20 mM glacial acetic acid, 1 mM EDTA, 0.75% w/v agarose, 0.2 μg ml^{-1} ethidium bromide) at 100 V and 230 mA, unless otherwise stated.

2.3.2 Preparation of Bacterial Genomic DNA

1.5 ml of bacterial culture was harvested by centrifugation at 5 000 g for 5 min and the pellet resuspended in 567 μl of TE buffer. 30 μl of 10% SDS and 3μl of proteinase K (20 mg ml^{-1}) were added prior to incubation at 37 °C for 1h. 100 μl of 5 M NaCl and 80 μl of CTAB/NaCl (1.5 M NaCl, 10% w/v hexadecyltrimethyl ammonium bromide) were added and the solution incubated for 10 min at 65°C. An equal volume of 25:24:1 phenol/chloroform/isoamyl-alcohol was thoroughly mixed with the solution and centrifuged for 5 min at 14 500 g. The aqueous supernatant was transferred to a fresh tube. Phenol treatment was repeated until no white interface was detectable. A 60% volume of isopropanol was added to precipitate the DNA and the solution was centrifuged at 14 500 g for 30 min. The resulting DNA pellet was washed with 70% ethanol and resuspended in 100μl of distilled water.

2.3.3 Preparation of Bacteriophage DNA

Filtered phage suspensions were subjected to DNA extraction according to a method modified from Sambrook *et al.* (1989). Phage stock solutions were plaque assayed at

10^{-1}-10^{-2} dilutions to produce semi-confluent lysis. After overnight incubation at 37 °C, the plates were overlaid with 3 ml SM (Table 3) buffer and shaken gently overnight at 4 °C. The SM buffer was removed and the top agar was scraped into a sterile glass centrifuge tube. The agar was vortexed with a 10% (v/v) ratio of SM to top agar, centrifuged at 10 000 g for 10 min and the supernatant pooled with the recovered SM buffer. Chloroform (30 µl) was added to each 10 ml aliquot of phage suspension to lyse any remaining bacteria. RQ1 DNase (5 µg ml^{-1}) and RNase (1 µg ml^{-1}) were added to remove any bacterial contamination, and the solution incubated for 1 h at 37°C. Bacteriophage were precipitated by incubation with 33% (w/v) PEG 8000 on ice for 30 min and harvested by centrifugation for 10 min at 10000 g. The supernatant was discarded and the phage pellets resuspended in 500 µl SM buffer per 30 ml original volume. An additional DNase and RNase step was carried out as described previously. Phage protein was removed and DNA purified by the addition of an equal volume of equilibrated 25:24:1 phenol/chloroform/isoamyl-alcohol (pH 8.0). The mixture was centrifuged at 14,500 g for 5 min and the resulting aqueous phase transferred to a fresh tube. The process was repeated twice and the DNA precipitated by the addition of 0.6 volume of isopropanol, flash chilling in liquid nitrogen for 2 min and subsequent centrifugation at 14,500 g for 30 min. The supernatant was discarded and the pellet washed with 70% (v/v) ethanol prior to resuspension in 100 µl distilled water.

2.3.4 PCR Amplification Parameters

Routine DNA amplification was carried out using 1.25 U of MBI fermentas recombinant Taq DNA polymerase (Fermentas UK, York, UK) in a 50 µl reaction

mix consisting of 1 mM dNTP mix, 1.5 mM $MgCl_2$, proprietary buffer, 200 nM oligonucleotide primers and 100 ng - 1 µg template DNA. DNA for applications requiring a minimal error rate was amplified using 1.25 U of EXPAND high fidelity DNA polymerase (Roche Diagnostics Ltd, Lewes, UK), as indicated, in a reaction mix containing 1 mM dNTP mix, proprietary $MgCl_2$ containing buffer, 200 nM oligonucleotide primers and 100 ng - 1 µg template DNA. Cycling conditions comprised an initial denaturation at 94 °C for 5 min, 35 cycles consisting of denaturation at 94 °C for 1 min, annealing for 1 min at various temperatures (Table 5) and extension at 72 °C/68 °C for 1 min kb^{-1} (Taq/EXPAND respectively), followed by a final extension at 72 °C/68 °C for 7 min.

Table 5. DNA amplification Primers

Primer	Sequence (5'-3')	Annealing
Int Probe Fwd	CTGGAAGTAATCCGCAGG	50 °C
Int Probe Rev	AGCTCTTTCGTTCTTAGG	50 °C
pGEM Fwd	ACTCACTATAGGGCGAATTCG	55 °C
pGEM Rev	CTGCAGGTCGACTCTAGAGG	55 °C
Integ Fwd	GTGTCACTCAGTGCCAATGC	55 °C
intS Fwd *Age* I	TCGCTTCAGACCGGTGACAGCCGCACTCC	55 °C
intS Rev *Sal* I	GATGAGTCGACACTTATACACACAAAGC	55 °C
aadA Fwd *Acc* I	CGAAACCTTGCGGTCGACCGCCAGCC	60 °C
aadA Rev *Acc* I	GCCTTTCATGATATGTCGACCAATTTGTGTAGGGC	60 °C
aadA Fwd *Pst* I	CGAAACCCTGCAGTCGTTCGCCAGCC	60 °C
aadA Rev *Pst* I	GCCTTTCATGATATATCTGCAGATTTGTGTAGGGC	60 °C
intS Inverse Fwd *Acc* I	CAGCACGATTGTCGACGAGCAC	65 °C
intS Inverse Rev *Acc* I	CCTCGCAGTCGACGCTTTATCC	65 °C
argW Fwd	TTTAAGCAATCGAGCGGCAGCG	55 °C
24_B IR F1	GTATCTTCATCATTCGCTCC	50 °C
24_B IR F2	CGTAGCACGCCTTTCC	50 °C
24_B IR F3	GCATAAGACCTTTGAGTG	50 °C
24_B IR F4	GGCGGTCAATGCTAAGATTATC	50 °C
24_B IR F5	CGGCCTTATGCTTTCAATG	50 °C
24_B IR R1	GATAGATGGCGTTCAATGAC	50 °C
24_B IR R2	GCTTGCGAAATTTCTAAACG	50 °C
24_B IR R3	GCTGCCATACGCGTTAC	50 °C
24_B IR R4	CAAGCAGGGAGATAATACGG	50 °C
24_B IR R5	GTCTGGGGACGCTATAC	50 °C
PIA Fwd	AACAGCCTTACTCCTGATGC	55 °C
PIA Rev	ATCGAATGTGATTACAATCG	55 °C
PIB Fwd	ATCAGATTATGTGATGACTCG	50 °C
PIB Rev	AAGATAGTGTGAAAATGACCC	50 °C
PIE Fwd	GCTCCAACAAGAAATAGACG	55 °C
PIE Rev	GAGAATGCTTTAGCGATCG	55 °C
PIF Fwd	AACAAACCTGCGCTGAGTGG	60 °C
PIF Rev	AGTACCTGCAGCGAGTGACG	60 °C
24_B *int* Probe Fwd	CACATCAAATCCGATATTAAGC	55 °C
24_B *int* Probe Rev	CTCAACAAGAAAAAGGCTCG	55 °C
24_B *int* Seq F	CCCGATTACAGCAGATGC	50 °C
24_B *int* Seq R	ACTGTAAATCTACGGAATGC	50 °C
24_B *int* Seq F2	GAAAGCCGCAAACATCAGC	60 °C
24_B *int* Seq R2	CTCAACAAGAAAAAGGCTCG	55 °C
24_B *int* Seq F RC	CTGCATCTGCTGTAATCG	55 °C
24_B *int* Seq F2 RC	CTGATGTTTGCGGCTTTTCG	60 °C
24_B *int* Seq R2 RC	AGCCTTTTTCTTGTTGAGGC	55 °C
24_B *int* Seq F3	AACCGAGCTGTTTAGCGTGC	60 °C
24_B *int* Seq F3 RC	CGTGTATAATAAACACGACTGG	55 °C

2.3.5 Southern Blotting and Chemi-luminescent Detection

Chromosomal DNA from naïve MC1061-Rif and $\Phi 24_B$-infected MC1061-Rif was digested with either *Eco*R I, *Bam*H I or *Hin*c II (New England Biolabs, Hitchin, UK) restriction endonucleases according to the manufacturer's recommendations in a final volume of 100 µl. DNA (1 µg) from each digest plus 5 µl DIG-labelled DNA marker IV or VII (Roche Diagnostics Ltd, Lewes, UK) were separated on an agarose gel for 4 h at 70 V. The DNA was prepared for transfer to a nylon membrane by successive submersed washes, once in depurination solution and twice in denaturation solution and neutralisation solution (Table 3) according to manufacturer's protocols (Roche Diagnostics Ltd, Lewes, UK). The gel was allowed to equilibrate in washing buffer (Table 3) and the DNA was transferred to a nylon membrane (Hybond N, Amersham Pharmacia Biotech, NJ) by capillary transfer for 16 h, using 20x SSC (Table 3) and Whatman 3MM paper (Whatman plc, Brentford, UK) as a wick. Transferred DNA was UV cross-linked to the membrane using a UV Stratalinker 2400 (Stratagene, La Jolla, CA) at 120 mJ. The membranes were incubated in proprietary pre-hybridization solution (Table 3) for 2 h at 42°C. The proprietary hybridization solution was then applied containing specific DIG-labelled probes (20 ng ml^{-1}), which were allowed to hybridize overnight at 42 °C. Excess probe was removed by incubation of the membrane twice with 2 x wash solution (Table 3) at room temperature for 5 min and twice with 0.5 x wash solution at 68°C. After equilibration in washing buffer, non-specific binding sites were blocked by gentle agitation in blocking solution for 1 h at room temperature. This was subsequently removed and replaced by antibody solution, and incubated for 30 min at room temperature with gentle agitation. Unbound antibodies were removed via two 15 min rinses in washing buffer. After equilibration

of the membrane in detection buffer (Table 3) for 2 min, 0.5 ml of CDP-Star substrate solution (Table 3) was applied directly to the membrane and luminescence allowed to develop for 1-2 mins. Target bands were identified following chemiluminescent detection via exposure to X-ray film (Kodak Limited, Hemel Hempstead, UK). Exposure times varied up to 10 min depending upon signal strength.

2.3.6 Preparation of Chemically Competent Cells

Overnight culture (0.5 ml) was used to inoculated 20 ml of TYM broth (2% Bacto Tryptone, 0.5% yeast extract, 0.1 M NaCl, 10 mM $MgSO_4$) in a 250 ml flask which was then incubated at 37 °C until it reached an OD_{600} of 0.5-0.8. The culture was then transferred to a 2 L flask containing 100 ml TYM broth and incubated at 37 °C until it reached an OD_{600} of 0.6-0.9. A further 380 ml TYM broth was added and the culture was again returned to 37 °C. Once the culture reached OD_{600} 0.6, it was rapidly cooled by shaking in ice water and cells harvested by centrifugation at 5,000 g for 15 min. The supernatant was discarded and the bacterial pellet gently resuspended in 100 ml ice cold TfB I (30 mM KOAc, 50 mM $MnCl_2$, 100 mM KCl, 10 mM $CaCl_2$, 15% v/v glycerol). The cells were harvested by centrifugation at 5 000 g for 8 min, the supernatant discarded and the pellet resuspended in 20 ml cold TfB II (10 mM Na MOPS; pH 7, 75 mM $CaCl_2$, 10 mM KCl, 15% v/v glycerol). 100 µl aliquots were apportioned into pre-chilled tubes, flash frozen in liquid nitrogen and stored at -80 °C.

2.3.7 Transformation of Chemically Competent Cells

Plasmid DNA (200-400 μg) was added to 100 μl of ice-thawed competent cells incubated on ice for 30 min. The cells were subsequently heat-shocked at 42 °C for 30 s, immediately diluted 1:10 in Luria Bertani broth and recovered at 37 °C for 1 h. Cells (100 μl) were spread plated onto selective media.

2.3.8 Preparation of Electro-Competent Cells

Bacterial cells were cultured at 37 °C in LB until they reached mid-exponential growth phase (~OD_{600} 0.6). The cells were washed via repeated harvesting by centrifugation at 5,000 g for 5 min and resuspension in 1/10 vol. ice cold distilled water. The washing process was routinely repeated 3 times before final resuspension in ice cold 10% glycerol (0.001% of the original culture volume). Aliquots (100μl) were stored at -80 °C.

2.3.9 Electroporation of Competent Cells

Chilled competent cells (100 μl) were transferred to a disposable 1 mm aperture electrocuvette containing 200-400 μg of plasmid DNA. Ligation products were desalinated prior to transformation. Desalination was carried out by drop dialysis whereby the ligation mix was transferred to a filter disc, floating on distilled water, for 1 h. The cells were electroporated using a BIO-RAD Gene Pulser electroporator (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK) at 2.5 V, 25 μF, 200 Ω. The cells

were recovered by the addition of 1 ml LB and incubation at 37 °C for 60 min. After recovery, the cells were harvested by centrifugation 5000 g for 10 min, resuspended in 100µl LB and spread plated onto selective agar plates. Colonies resistant to the selection antibiotic(s) were picked and analysed.

2.3.10 Colony Blotting

Fresh transformant colony plates were incubated at 4°C for 1 h, and nylon membrane discs (Hybond N, Amersham Pharmacia Biotech) were then overlaid onto the agar plates and removed promptly after 1 min. The membranes were then placed sequentially, colony side up, on Whatman 3MM paper soaked in denaturing solution for 15 min, neutralization solution for 15 min, and finally in 2 x SSC for 10 min. The DNA was UV cross-linked to the membrane using a UV Stratalinker 2400 at 120 mJ. Cell debris was removed mechanically between 2 sheets of Whatman 3MM paper soaked in dH_2O. Ensuing hybridization through to chemiluminescent detection was carried out as for Southern blotting.

2.4 Sequence Analysis

GenBank database searches were carried out using various Blast programs on the NCBI website (http://www.ncbi.nlm.nih.gov). In addition, the Colibri web server was used to locate specific short sequence patterns in the *E. coli* K-12 genome (http://genolist.pasteur.fr/Colibri). Sequence alignments were carried out with GeneDoc Multiple Sequence Alignment Editor and Shading Utility (http://www.psc.edu/biomed/genedoc), BioEdit Sequence Alignment Editor (http://www.mbio.ncsu.edu/**BioEdit/bioedit**.html) and ClustalW Multiple Sequence Alignment Program (http://www.ebi.ac.uk/**clustal**w/). Primer binding sites, restriction mapping and sequence manipulation were carried out using pDRAW32 DNA Analysis Software (www.acaclone.com).

Chapter 3: Identification of the $\Phi 24_B$ Integrase, the Primary Integration Site and 3 Additional Integration Sites

3.1 Background

Bacteriophages can integrate into their host's genome via either transposition, *e.g.* phage Mu (Shapiro, 1979), or site specific recombination, *e.g.* bacteriophage lambda (Ptashne, 2004). The former can insert in any number of locations throughout the genome whereas the latter is highly specific. Although temperate lambdoid bacteriophages integrate at a diverse variety of sites within the *E. coli* genome, each individual phage has a distinct preference for one specific site. Lambda, for example, despite its strong affinity for a single preferred site of integration, can occupy secondary locations in the absence of the primary site (Shimada *et al.*, 1972). These supplementary sites share a degree of identity with the primary site in the overlap region, the core binding region or both (Campbell, 1992). Utilisation of additional insertion sites usually occurs at a decreased frequency in comparison to the preferred integration site (Mizuuchi *et al.*, 1981). The specificity of integration is dependent upon DNA/DNA interaction between the corresponding overlap regions and DNA/protein interaction of the integrase molecule and the core binding sites.

P2 is a non-lambdoid, temperate bacteriophage of the *Enterobacteriaciae* for which lysogeny with a superinfecting phage is not precluded by immunity (Bertani & Six, 1958). The primary integration site for P2, locI, possesses 27 bp of complimentary overlap between *attP* and *attB* (Yu *et al.*, 1989). When the preferred site is either absent or occupied by a resident prophage, at least nine further integration sites have been reported to be functional (Barreiro & Haggard-Ljungquist, 1992) (Table 6). The

extra sites do not share the same degree of consensus sequence with the *attP* overlap region. LocII matches 20 out of 27 (74%) nucleotides whilst locIII and locH only possess 17 out of 27 (63%) base pair identity exposing a tolerance for at least a 37% mismatch in overlap sequence.

Table 6. Location of the multiple P4 insertion sites

Insertion Site	Chromosomal Position	Strain
locI	Between Histidine operon and *metG*	*E. coli* C
locII	Between *metE* and *rha*	*E. coli* C & K12
locIII	Between *trp* and *terC*	*E. coli* C & K12
locH	Between *shiA* and *his*	*E. coli* K12
locIV-IX	Unknown specific location	-
locE	Unknown specific location	-

Temperate bacteriophages frequently insert preferentially within tRNA genes. Generally, there are two main types of bacteriophage with respect to insertion preferences, lambda-like and P4-like. Lambda-like phages usually prefer to integrate into tRNA anti-codon loops whilst P4-like phages favour the downstream TΨC ribosome recognition site (Campbell, 2003). Host *attB* sites typically range from 20-30 bp, including a central overlap region comprised of seven to fifteen nucleotides of identical sequence to the viral *attP*. The regions flanking the overlap are usually similar if not identical in the bacterial and phage DNA partners, often comprised of imperfect inverted repeat sequences. Strand swapping during Holliday junction resolution is highly dependent upon complimentary base pairing, therefore sequence identity is vital for efficient reaction culmination (Campbell, 2003). The requirement for complimentary sequence is illustrated by the lambdoid phage 21 and the defective element e14 of *E. coli* K-12 (Wang *et al.*, 1997). Both phages integrate into the same *attB* site within the *E. coli* K-12 *icd* gene and both encode integrase, excisionase and

repressor protein homologues. However integration and excision function of each phage on its counterpart are either absent or reduced due in part to differing arm-type *attP* binding sites but also a lack of sufficient sequence identity between *attP* and *attB* when undergoing recombination with the respective integrases.

Among lambdoid phages some are prone to insert within host genes, *e.g.* λ, whilst others prefer intragenic regions, *e.g.* phage 21. In the latter case, the bacteriophage will generally encode a duplicate copy of the interrupted gene to avoid any disruption to function. Jacob & Wollman (1961) reported that of 13 independently isolated coliphage, seven were found to be UV inducible, lambdoid phage that were all clustered within a specific region of the genome. Subsequently Bachmann (1990) localised the lambdoid *attB* sites as well as the loci of the defective lambdoid bacteriophages to between 6 and 44 minutes of the *E. coli* genome.

It has been demonstrated unequivocally, through the use of differentially labelled, isogenic recombinant phages ($\Phi 24_B$::Cat and $\Phi 24_B$::Kan) that double lysogens of a lambdoid Stx-phage are possible (Allison *et al.*, 2003). Previously this phenomenon had only been observed in lambda via immunity mutant experiments or homologous recombination events (Calef, 1967; Freifelder & Kirschner, 1971). The phage used for these experiments, originally named ΦE86654-Stx2, was one of two phages induced from a clinical STEC isolate (Allison *et al.*, 2003). The STEC isolate was an outbreak associated O157:H7 strain obtained from the Central Public Health Laboratory, Colindale, London. Clearly, the occurrence of multiple lysogens in a single host is likely to enhance the evolution and dissemination of bacteriophage-encoded genes throughout bacterial populations, with particular applied relevance for

Stx-phages responsible for increasing the pathogenic potential of *E. coli* hosts. The underlying mechanisms responsible for multiple lysogeny in *E. coli* infected with Stx-phage $\Phi 24_B$ are unknown. As the genome sequence of $\Phi 24_B$ was not available, it was the objective of this chapter to specifically identify the $\Phi 24_B$ integrase and the locations of the *attB* and the *attP* sites that are required for site specific recombination.

3.2 Specific Methods

3.2.1 Creation of *intS* Mutant

The CPS-53 (KpLE1) putative prophage integrase *intS*, present in the *E. coli* K12 strain MC1061, was amplified with *intS* Fwd *Age* I and *intS* Rev *Sal* I primers and the relevant parameters (Table 5) using EXPAND high fidelity polymerase and purified using the Eppendorf Perfect Prep Gel Extraction Kit (Eppendorf UK Ltd, Cambridge, UK) and cloned into Zero Blunt cloning vector (Invitrogen Ltd, Paisley, UK) according to manufacturers' protocols. Plasmid from transformants was purified using a QIAGEN mini-prep kit (QIAGEN, Hilden, Germany) and subjected to PCR using *intS* Inverse Fwd *Acc* I and *intS* Inverse Rev *Acc* I oligonucleotides with the relevant parameters (Table 5). The resulting amplification product was purified and 1 μg digested with the restriction endonuclease *Acc* I for 1 h at 37 °C (New England Biolabs, Hitchin, UK). The antibiotic resistance cassette, *aadA*, encoding resistance to spectinomycin, was amplified with *aadA* Fwd acc I and *aadA* Rev acc I oligonucleotides and EXPAND High Fidelity Polymerase with the relevant parameters (Table 5). After subsequent purification, 1 μg of the amplification product was digested with the restriction enzyme *Acc* I for 1 h at 37 °C. The spectinomycin resistance cassette possessing *Acc* I 3' overhangs was then cloned into the Zero Blunt backbone containing the *intS* flanking regions (268bp and 224bp) and compatible *Acc*I cohesive ends. The resulting plasmid was purified using QIAGEN's mini prep kit and the mutated *intS* was excised from the plasmid backbone by sequential digest using the restriction endonucleases *Age* I and *Sal* I. Following heat inactivation of the restriction endonucleases, the construct was ligated into *Age* I/*Sal* I digested suicide vector, pJP5603. This plasmid is dependent upon λ pir for

replication and was routinely propagated in the λ pir^+ strain JM109. The constructed pPCMF*intS* (Fig. 14) was introduced into MC1061 via electroporation (2.5 V, 25 μF, 200 Ω) by means of the Bio-Rad Gene Pulser (Bio-Rad Laboratories Ltd, Hemel Hempstead, UK). Selection for kanamycin and spectinomycin resistance allowed the identification of clones produced by a single recombination event, which were confirmed by PCR and Southern hybridisation (data not shown). Passage of a single recombinant and a 30 s heat shock treatment at 42 °C was followed by screening for the loss of kanamycin resistance facilitated the identification of an *intS* knock-out in *E. coli* MC1061 and ablation of the gene was confirmed by PCR (Fig. 16).

3.2.2 Complementation of *intS*

The putative prophage integrase, *intS*, was amplified with *intS* Fwd Age I and *intS* Rev Sal I primers and EXPAND high fidelity DNA polymerase with the relevant parameters (Table 5) and cloned into the low copy number vector pKT230 (Table 4) according to standard cloning protocols (Sambrook *et al.*, 1989). MC1061 *intS*⁻ was used as the *E. coli* K-12 host.

3.2.3 Recombinant Mutant Phage Construction

The plasmid NTP707 was digested with the restriction endonuclease *Pst* I whilst the spectinomycin antibiotic resistance cassette, *aadA*, was amplified with *aadA* Fwd *Pst* I and *aadA* Rev *Pst* I using the relevant conditions (Table 5) and digested with *Pst* I. The two fragments were ligated together and the resultant plasmid, NTP707-*aadA*, was transformed into the *E. coli* strain MC1061, which was subsequently infected

with $\Phi 24_B$::Kan. Bacteriophage resulting from this infection were isolated by filtration and used to infect naïve MC1061. Lysogens produced by the infection were plated onto LB agar containing spectinomycin and replica plated onto both LB agar plus kanamycin and LB agar containing tetracycline. Colonies displaying spectinomycin resistance but kanamycin and tetracycline sensitivity were deemed putative $\Phi 24_B$::Spec recombinant phage lysogens and further examined for the ability to produce infective phage by plaque assay.

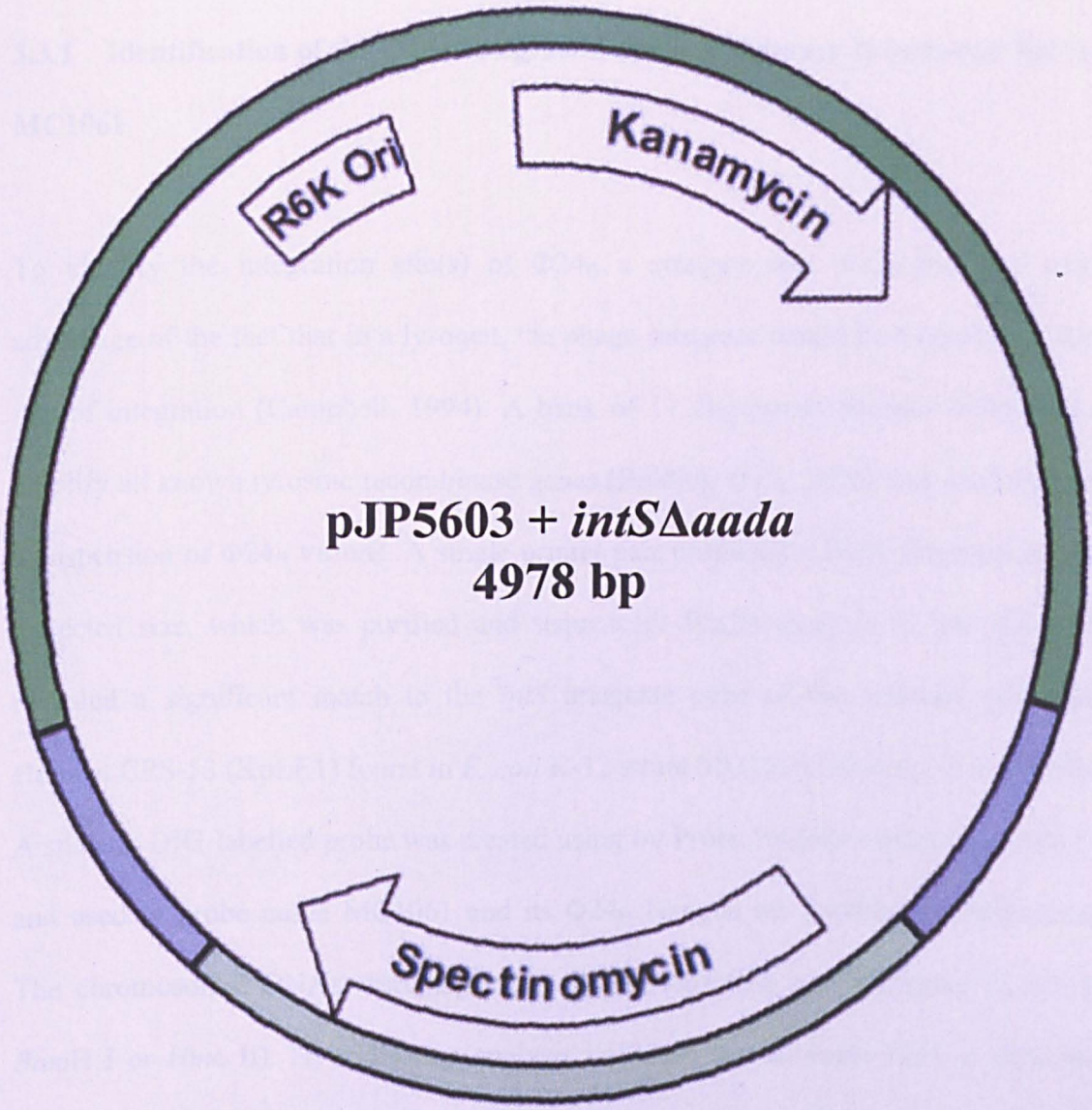
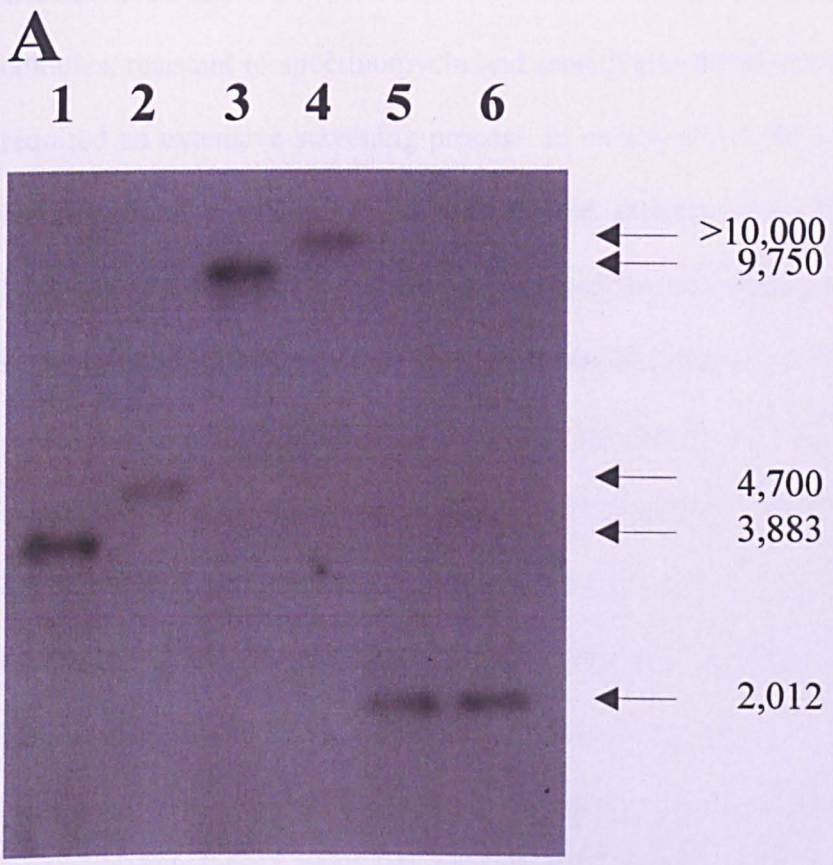


Figure 14. Suicide plasmid construct for ablation of *intS* pJP5603 contains a lambda pir dependent R6K origin of replication and resistance to kanamycin. *intS* (blue) is interrupted by spectinomycin resistance (light green).

3.3 Results and Discussion

3.3.1 Identification of the $\Phi24_B$ Integrase Gene and Primary Integration Site in MC1061

To identify the integration site(s) of $\Phi24_B$ a strategy was employed that took advantage of the fact that in a lysogen, the phage integrase would be located near the site of integration (Campbell, 1994). A bank of 11 degenerate primers designed to amplify all known tyrosine recombinase genes (Balding *et al.*, 2005) was used against a suspension of $\Phi24_B$ virions. A single primer pair produced a DNA fragment of the expected size, which was purified and sequenced. Blastn analysis of the sequence revealed a significant match to the *intS* integrase gene of the remnant prophage element CPS-53 (KpLE1) found in *E. coli* K-12 strain MG1655 (Blattner *et al.*, 1997). A specific DIG-labelled probe was created using *int* Probe Fwd/Rev primers (Table 5) and used to probe naïve MC1061 and its $\Phi24_B$ lysogen by Southern hybridisation. The chromosomal DNAs were digested with 3 restriction endonucleases (*Eco*R I, *Bam*H I or *Hin*c II). Hybridisation analysis indicated that a single band in both the naive MC1061 lanes and the $\Phi24_B$ lysogen lanes, however the latter were ~1 kb larger in size for 2 out of 3 digests (Fig. 15A). This discrepancy in size suggested that the integration of $\Phi24_B$ occurred in the vicinity of the host *intS* gene (Fig. 15B). This presented the possibility that one or more of the integration events that had been previously reported for $\Phi24_B$ (Allison *et al.*, 2003) might in fact have been directed by the host cell encoded *intS* gene. In order to rule out any host-mediated effect on $\Phi24_B$ integration, the host *intS* gene was replaced by a spectinomycin resistance cassette, *aadA*. The process of *intS* replacement was complicated by various experimental difficulties including PCR optimisation to account for large amplicon size during



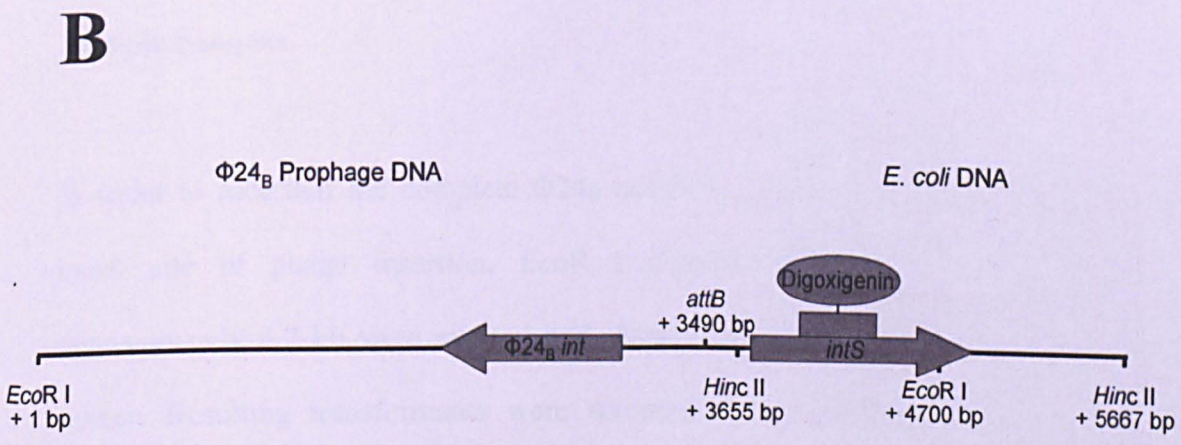


Figure 15. Identification of the $\Phi 24_B$ integrase and integration site **Panel A.** Southern Blot of Naïve and $\Phi 24_B$ lysogen DNA digested with *Eco*R I, *Bam*H I or *Hin*c II and hybridised with DIG-labelled integrase probe. Odd numbered lanes contain DNA of Naïve MC1061 cells. Even numbered lanes contain DNA from $\Phi 24_B$ lysogens. The DNA was digested with restriction endonucleases as follows: lanes 1 & 2 with *Eco*R I, lanes 3 & 4 with *Bam*H I and lanes 5 & 6 with *Hin*c II. Arrows to the right of the blot indicate the deduced size of the DNA bands in bp. **Panel B.** Map of the $\Phi 24_B$ DNA sequence depicting the location of the restriction endonuclease sites, location of the $\Phi 24_B$ integrase, location of the *intS* gene and the location of the *attB* sequence. The numbers report the number of base pairs that a given location is from the upstream *Eco*R I site. The *Bam*H I sites are located beyond the boundaries of the map.

PAGE MISSING IN ORIGINAL

inverse PCR reactions and primer mismatches when introducing restriction endonuclease sites. In addition, the identification of putative double recombinant colonies, resistant to spectinomycin and sensitive to the plasmid encoded kanamycin, required an extensive screening process. In excess of 12 500 colonies were screened before a putative *intS* knock-out was detected. Deletion of the gene was confirmed by PCR amplification using primers that spanned the *intS* gene, *argW* Fwd/ *intS* Rev *Sal* I, and internal to the deleted portion, *int* Probe Fwd/Rev (Fig. 16). The rate of lysogen production in MC1061 *intS*⁻ was lower than the rate observed in wild-type MC1061 in a parallel infection assay (Appendix 2, $p<0.05$), however the wild-type phenotype could not be complemented by re-introduction of *intS* on a low copy number plasmid (Appendix 2, $p>0.05$), pKT230 (Table 4). This signifies that the effects on infection rate are most likely due to polar effects brought about by the abrogated *intS* gene. As such, these data indicate that the host *intS* gene is neither required nor involved in the integration of $\Phi24_B$, and therefore it does not appear to be a factor in the formation of multiple lysogens.

In order to ascertain the complete $\Phi24_B$ integrase gene sequence and to define the exact site of phage insertion, *Eco*R I digested lysogen DNA fragments of approximately 4.7 kb were excised and cloned into *Eco*R I pre-cut pGEM cloning vector. Resulting transformants were screened for ampicillin and spectinomycin resistance, and the identity of resistant colonies confirmed by PCR amplification with *aadA* Fwd/Rev *Acc* I (Table 5). The insert from a PCR-confirmed resistant colony was sequenced. The resulting 4.7 kb of sequence data contained 2 ORFs corresponding to *aadA* and putative $\Phi24_B$ *int*. This integrase ($\Phi24_B$ *int*, Accession number EF397940) has only been identified once previously (Stx-phage-86, 96%

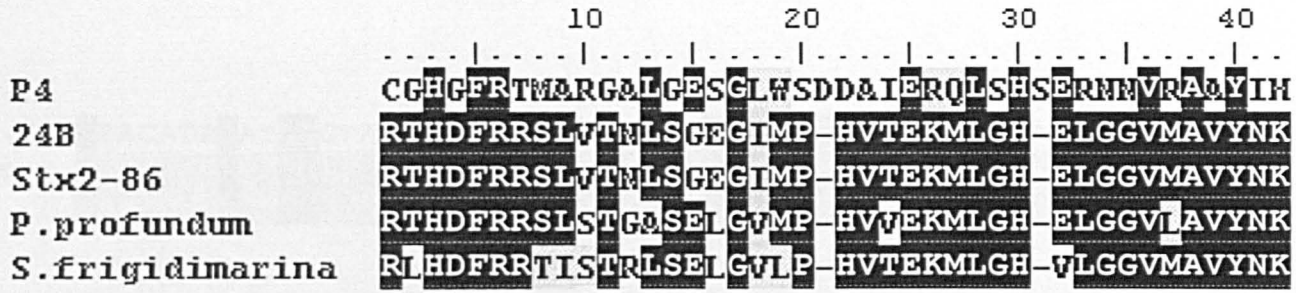
identity, Accession number: AB255463), and that phage has yet to be described in the scientific literature. Sequences flanking the integrase genes are homologous between $\Phi 24_B$ and phage-86, most notably the region stretching from the integrase gene to *attP* (98% nucleotide identity), however, significant regions of disparity are noticeable at the whole genome level. Furthermore, the amount of overall amino acid sequence identity and similarity shared between the $\Phi 24_B$ *int* gene and *intS* is only 14% and 24%, respectively. Blastx analysis of the integrase of $\Phi 24_B$ revealed a match to the conserved domain database for P4 integrase whilst the closest matches in the GenBank database were to prophages of *Photobacterium profundum* and *Shewanella frigidimarina* (46% and 38% identity, respectively; Fig. 17). DNA alignment of the insertion sequence to naïve *E. coli* MC1061 host DNA enabled the identification of the integration point. As expected from the original Southern Blot analysis, the insertion occurred at a point almost 250 bp 5' of *intS* and within 75 bp of the site preferred by the bacteriophages Sf6 and HK620, both of which possess an *intS* like integrase (Casjens *et al.*, 2004; Clark *et al.*, 2001). Alignment of the *E. coli* MC1061 *attB*, $\Phi 24_B$ *attP* and integrated prophage sequences enabled the extent of complimentary sequence overlap to be determined as 24 bp (Fig. 18A).

3.3.2 Identification of 3 Additional Integration Sites

Having identified one, and possibly the primary, integration site, it was clear that there would be at least one other integration site present in order to explain the double lysogeny observed by Allison *et al.* (2003). The identification of the $\Phi 24_B$ DNA integrase sequence facilitated the production of a DIG-labelled $\Phi 24_B$ *int* specific

Figure 17

Box I:



Box II:

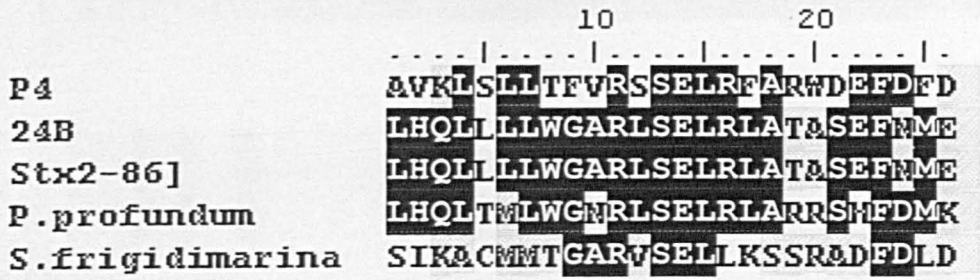


Figure 17. Comparison of $\Phi 24_B$ integrase to the most closely related integrases ClustalW multiple amino acid sequence alignments of the conserved box I and box II regions of the integrases of $\Phi 24_B$, P4 and the three closest Blast matches.

Figure 18

Figure 18. Analysis of the $\Phi 24_B$ *attB* sites **Panel A.** Alignment of $\Phi 24_B$ *attP*, Naïve MC1061 *attB* and integrated lysogen sequences, containing the 24 bp overlap region highlighted in black. **Panel B.** Alignment of the previously defined 24 bp overlap region of all four *E. coli* MC1061 integration sites (labelled I-IV).

probe, which was allowed to hybridise to DNA from two single lysogens (MC1061 carrying $\Phi24_B$::Cat or $\Phi24_B$::Kan) and two double lysogens (MC1061 infected first with $\Phi24_B$::Cat then $\Phi24_B$::Kan and *vice versa*) that were derived from the single lysogens digested with either *Eco*R I or *Ava* I. In total, from the four different DNA preparations, the *int* probe bound to DNA fragments that indicated the presence of four distinct integration sites (Fig. 19). The DNA fragments corresponding to the three novel insertion sites were excised and cloned to produce 3 subgenomic libraries. The presence of the *int*-containing sequence in each of these libraries was confirmed by PCR using $\Phi24_B$ *int* Fwd/Rev primers (Table 5). An aliquot from each plasmid library was then used as template for a subsequent PCR amplification using one primer specific for phage sequence, Integ Fwd, and one specific to pGEM sequence, pGEM Fwd/Rev (Table 5). Amplification using this primer pair was problematic, presumably due to the fact that the plasmid specific primer would bind non-discriminatorily to all plasmids present. PCR amplifications, therefore, required optimisation to minimise the number of bands produced. Furthermore, the distance from the integration site to the plasmid specific primer, for the larger cloned DNA fragments, would be too large to easily amplify. Ultimately a 350 bp product was amplified from the library prepared from *Ava* I fragments of approximately 3 kb (Fig. 19, fragment II*), and the amplified product was sequenced. Approximately 150 bp were found to correspond to a second distinctive location in the host genome identified by a 100% Blastn match to an *E. coli* K-12 MG1655 hypothetical protein, *yfbO*. The degree of homology between the core *attP* sequence and the second insertion site was less than the 24 bp overlap identified at the first site, but there was a consensus sequence motif, (TATCYUTTTAWMTAA). The sequence

Figure 19

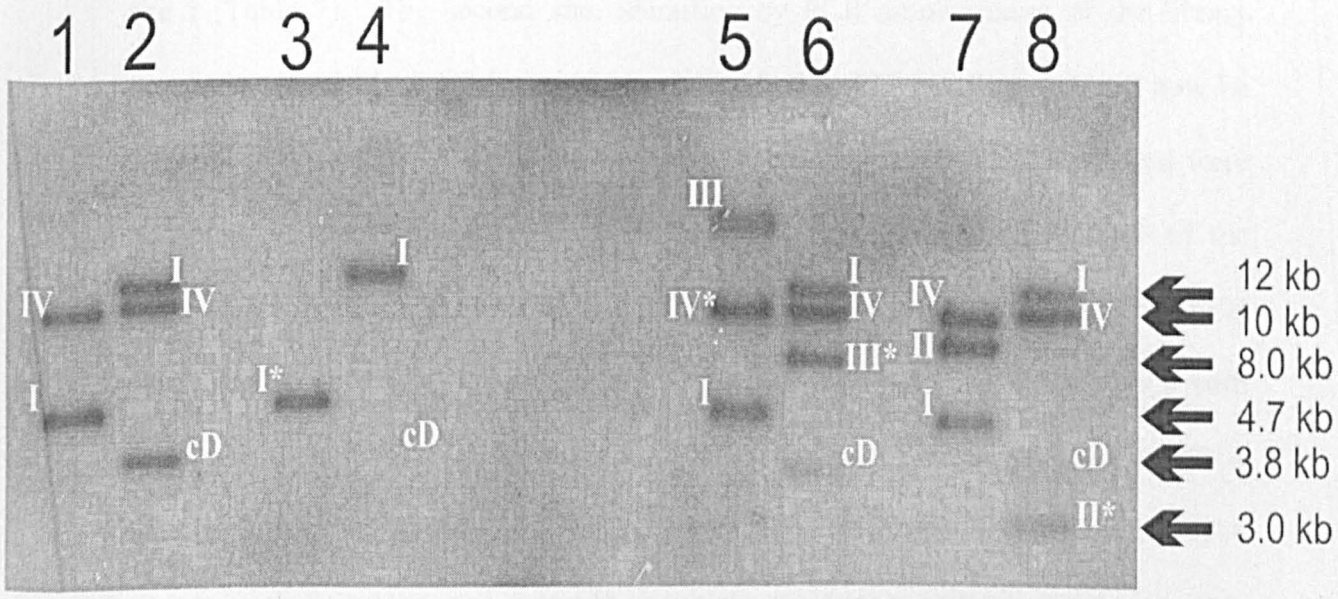


Figure 19. Identification of additional $\Phi 24_B$ integration sites Southern blot analysis of the DNA from two single and two double $\Phi 24_B$ lysogens. The source of the DNA is as follows: lanes 1 & 2, $\Phi 24_B$::Cat lysogen; lanes 3 & 4, $\Phi 24_B$::Kan lysogen; lanes 5 & 6, $\Phi 24_B$::Cat lysogen subsequently infected by $\Phi 24_B$::Kan; lanes 7 & 8, $\Phi 24_B$::Kan lysogen subsequently infected by $\Phi 24_B$::Cat. DNA in lanes 1,3,5 and 7 was digested with *Eco* RI and lanes 2,4,6 and 8 with *Ava* I. Bands corresponding to the identified integration sites are identified with Roman numerals at either the left (odd numbered lanes) or right (even numbered lanes). "*" indicates those bands used to produce the subgenomic libraries. Arrows to the right of the figure indicate approximate molecular size. The bands corresponding in size to circular, intracellular $\Phi 24_B$ DNA (3.8 kb) are labelled cD. DNA bands corresponding to the PIA-F identified *in silico* are as follows: I = PID, II = PIB, III = PIC. PIA and PIF lie within essential genes (Gerdes *et al.*, 2003).

TATC--TTTA--TAA, based upon this consensus sequence, was used to search the Colibri *E. coli* database for potential additional sites in the published *E. coli* K-12 genome sequence. This analysis yielded six possible sites, including the two already identified above (Fig. 20). The first integration site, near the MC1061 *intS* gene, was identified *in silico* as putative integration site D (PID) and will now be referred to as site I (Table 7). The second site, identified by PCR amplification of the library possessing the 3 kb *Ava* I fragment, was identified as PIC *in silico* and will now be referred to as site II. To investigate if any of the other sites identified *in silico* were occupied, PCR primers were designed to enable amplification across each of the putative integration sites - PIA, PIB, PIE and PIF (Table 5). Of these, only PCR amplification of DNA prepared from a double lysogen (MC1061 infected first with $\Phi 24_B$::Cat then with $\Phi 24_B$::Kan) using the PIB-specific primer pair produced no DNA product, suggesting that this site was occupied. PCR amplification of the same DNA sample with PIB Fwd and Integ Fwd initially produced multiple bands, possibly originating from PIB Fwd recognising sites located in the plasmid and in the chromosome. A 400 bp band was selected, correlating to the predicted band size determined *in silico*. Sequence analysis of this band revealed 100% identity (Blastn) to a region of the *E. coli* K-12 MG1655 genome that lies 377 bp upstream of *yfaL*, which encodes an adhesion protein, and 265 bp downstream of *nrdA*, a gene that encodes a reductase α-subunit (Blattner *et al.*, 1997). This site will now be referred to as site III (Fig. 19; Table 7).

In order to identify the final integration site (site IV) indicated by the data in Fig. 19, transformants produced from the library created with 10kb *Eco*R I fragments from a double lysogen (MC1061 infected first with $\Phi 24_B$::Kan then $\Phi 24_B$::Cat, Fig. 19,

6 pattern(s) found with 0 mismatch(s)

Pattern (position, strand)		bp from start codon of gene		Inside gene	Pattern sequence
390809	+	-2876	*yaiV*: Function unknown	*yaiT*: Function unknown; interrupted by IS3B	tctaaattaacttcagatactatcgatttaactaacggttcttcgctgaacattg
2342586	+	-299	*nrdA*: Ribonucleoside diphosphate reductase subunit B1; class I enzyme, aerobic, physiologically active		aaagattaataagccatctatatcaatttatctaacctattatgccgttcaagaa
2387097	-	-651	*yfbN*: Function unknown		aatttatgtaaacatccgcttatctgtttatataaatgttttaatcccaccaaca
2464349	-	-1322	*vacJ*: Shigella vacJ homolog		cccttaggaggggctcgttatatccatttaactaagaggacaatgcggcatgagt
2710061	-	-2029	*rpoE*: RNA polymerase, sigma E-subunit, high-temperature transcription (heat shock and oxidative stress)	*yfiC*: Function unknown	cggcgctgacccaggcttttatatctgtttatgtaatgattgccgggggaaaggat
2752185	-	-3456	*grpE*: Heat shock protein; mutant survives induction of prophage lambda; stimulates DnaK ATPase; nucleotide exchange function	*yfiF*: Function unknown	ttgctggaattgcgtaccgatatcgatttaactaaaaacaaagtcggcatttaca

Figure 20. Colibri Database putative insertion site output Details of the locations of six potential integration sites identified *in silico*. The red box outlines the consensus sequences of the insertion sites.

Table 7. NCBI Blast putative insertion sites output Gene descriptions and chromosomal positions of each of the potential $\Phi 24_B$ phage insertion sites indicated by the Colibri server output. PIA-F and I-IV integration site designations are also included. When phage integration is intergenic, bracketed numbers represent the distance of the integration site upstream or downstream of the flanking genes

		Chromosome Position	5'	3'
PIA		390809	Hypothetical Protein (1450 bp)	IS3 element protein *insF* (120 bp)
PIB	III	2342586	*yfaL* adhesin protein (377 bp)	*nrdA* reductase α-subunit (265 bp)
PIC	II	2387097	*yfbO* hypothetical protein	
PID	I	2464349	*yfdC* inner membrane protein (62 bp)	*intS* CPS-53 prophage integrase (196 bp)
PIE		2710061	*yfiC* methyl transferase	
PIF		2752185	*yfjF* hypothetical protein	
	IV	4525974	*sgcA* KpLE2 phage-like element predicted phosphotransferase component	

fragment IV*) were subjected to colony blot hybridisation with a $\Phi 24_B$ *int* specific DIG-labelled probe. The colony blotting technique is limited by the impurity of the DNA template used for hybridisation. Measures are taken to remove as much cellular debris as possible; however, some contamination inevitably remains, leading to background signals and frequent false positives. The labelled *int* probe bound to several colonies, which were picked and confirmed to possess the target sequence by PCR using $\Phi 24_B$ *int* specific primers (Table 5). The plasmids from three transformant clones were purified and sequenced. Blastn analysis revealed that the final integration site, revealed in Fig. 19, fell within the *sgcA* gene, which encodes a putative transport protein encoded on the KpLE2 phage-like element in *E. coli* K-12 MG1655. Alignment of the sequence to the previous three insertion sites reduced the consensus from a motif containing eleven conserved residues to only seven, however the percentage similarity of site IV to site I (the original 24 bp overlap) was 54% while site III and site II were 50% and 62.5% similar, respectively with site I (Fig. 18B). Site IV was not identified *in silico*. This tolerance for base pair mismatches in the secondary insertion sites has also been reported in other phages. For example, the non-lambdoid phage P2 possesses a 27 bp consensus sequence between phage the *attP* and the primary *attB* site, whilst its secondary *attB* sites contain only 63% nucleotide identity to *attP* (Barreiro & Haggard-Ljungquist, 1992). As is the case for $\Phi 24_B$, these sites are only utilised when the primary site is unavailable.

The sequences of all insertion sites in the naïve *E. coli* MC1061 genome were determined and compared to those found in the sequenced *E. coli* strain MG1655 using primer pairs Naïve 1-4 Fwd/Rev (Table 5) and found to possess 100% identity to those sequences in strain MG1655. The putative integration sites, PIA and PIF, are

located within regions thought to encode essential functions for *E. coli* and therefore if used will result in the death of the lysogen. A third recombinant phage, $\Phi24_B$::Spec, was constructed and used to infect a double lysogen culture (MC1061 infected first with $\Phi24_B$::Cat then $\Phi24_B$::Kan). Putative triple lysogens were confirmed by screening for resistance to kanamycin, chloramphenicol and spectinomycin, as well as the ability to release all three marked phage upon induction. The $\Phi24_B$ triple lysogen was subjected to further phage infection using all three of the labelled recombinant phages, in an attempt to identify an occupied PIE site, the only *in silico*-identified, putative integration site for which actual $\Phi24_B$ integration has not been demonstrated. This site does not share any known association with genes predicted to provide an essential function. However, this site was never found to be occupied. Lysogen production by $\Phi24_B$ has therefore been demonstrably extended to the ability to form triple lysogens, and four different sites of bacteriophage integration have been formally identified within the *E. coli* host genome (Fig. 18B; Table 7). However, the question of how the lambda super-infection immunity model is overcome remained.

3.4 Conclusions

- $\Phi 24_B$ possesses an integrase gene which is homologous to only one, as yet uncharacterised, integrase gene in the database.
- Four distinct integration sites have been identified in an *E. coli* K-12 host strain, MC1061.
- A further three sites have been hypothesised in silico, however, two are within predicted essential genes and the third has never been found to be occupied.
- A triple lysogen of a single host, with three differentially-labelled bacteriophage, was produced thus highlighting increased potential for intracellular recombination and the resultant implications for bacteriophage evolution.

Chapter 4: Screening for Integrases in Phages Induced From Wild Type STEC Strains

4.1 Background

Prophage and remnant prophage DNA accounts for a significant proportion of bacterial pathogen genomes, with the 2 sequenced *E. coli* 0157 strains possessing an additional 1 Mbp of DNA, compared to the laboratory K-12 strain MG1655, which is mostly comprised of bacteriophage-related genes and insertion elements (IS) (Allison, 2007; Blattner *et al.*, 1997; Hayashi *et al.*, 2001; Perna *et al.*, 2001). The majority of this phage-related DNA is associated with lambda-like phages or phage remnants. Lambdoid phages undergo a temperate life cycle and share a distinct genome organisation with the well characterised bacteriophage lambda (Campbell, 1994). Despite these shared characteristics, there is a high degree of mosaicism amongst lambdoid phages whilst genome organisation remains similar to lambda (Casjens, 2005). Clearly, as lambdoid phage and carriers of the Stx-toxin genes, Stx-phages are major contributory factors to the manifestation of disease caused by Shiga toxin producing *Escherchia coli* strains and the evolution of shigatoxigenic bacteria.

Comprehensive methods for classifying and typing Stx-phage from outbreaks and reservoirs have yet to be documented. Characterisation has typically been achieved on the basis of morphological traits which do not take into account the genomic variation and gene heterogeneity apparent from the available sequence data (Pringle, 1999). Consequently, morphological classification of phages that carry *stx* genes is not informative. Typing of bacterial isolates has been carried out increasingly via the use of Multi-Loci Sequence Typing (MLST). A similar multi-loci approach has been

developed by Smith *et al.* (2007b) to characterise 113 wild-type, inducible Stx-phage isolates using an array of primer sets that target key modules involved in infection and propagation. The work described in this chapter contributes to that typing scheme. The STEC samples were obtained from two sources: Faculty of Veterinary Science, University of Liverpool and the Veterinary Laboratories Agency in Surrey. Both organisations have conducted surveys of STEC in farm animals and the farm environment. Here, the integrase genes present in induced phage lysates were identified by the application of an existing integrase bank of primers, designed to amplify all known tyrosine recombinase genes (Table 8 and Figure 21), supplemented with a primer pair specific to the conserved region of the new $\Phi 24_B$ integrase. Tyrosine recombinase genes are extremely diverse, and thus difficult to align. The existing universal bank of primers were designed by limiting the alignments to the well conserved RHRY box regions, allowing short stretches of sequence identity to be elucidated and primers designed. Degeneracy was introduced into the primers to allow maximal integrase detection by relatively few primer sets. Each primer pair is specific to the integrase group for which it is designed to amplify and does not cross react with the integrases of other groups. The primer degeneracy does, however, allow the detection of similar but as yet unknown integrases.

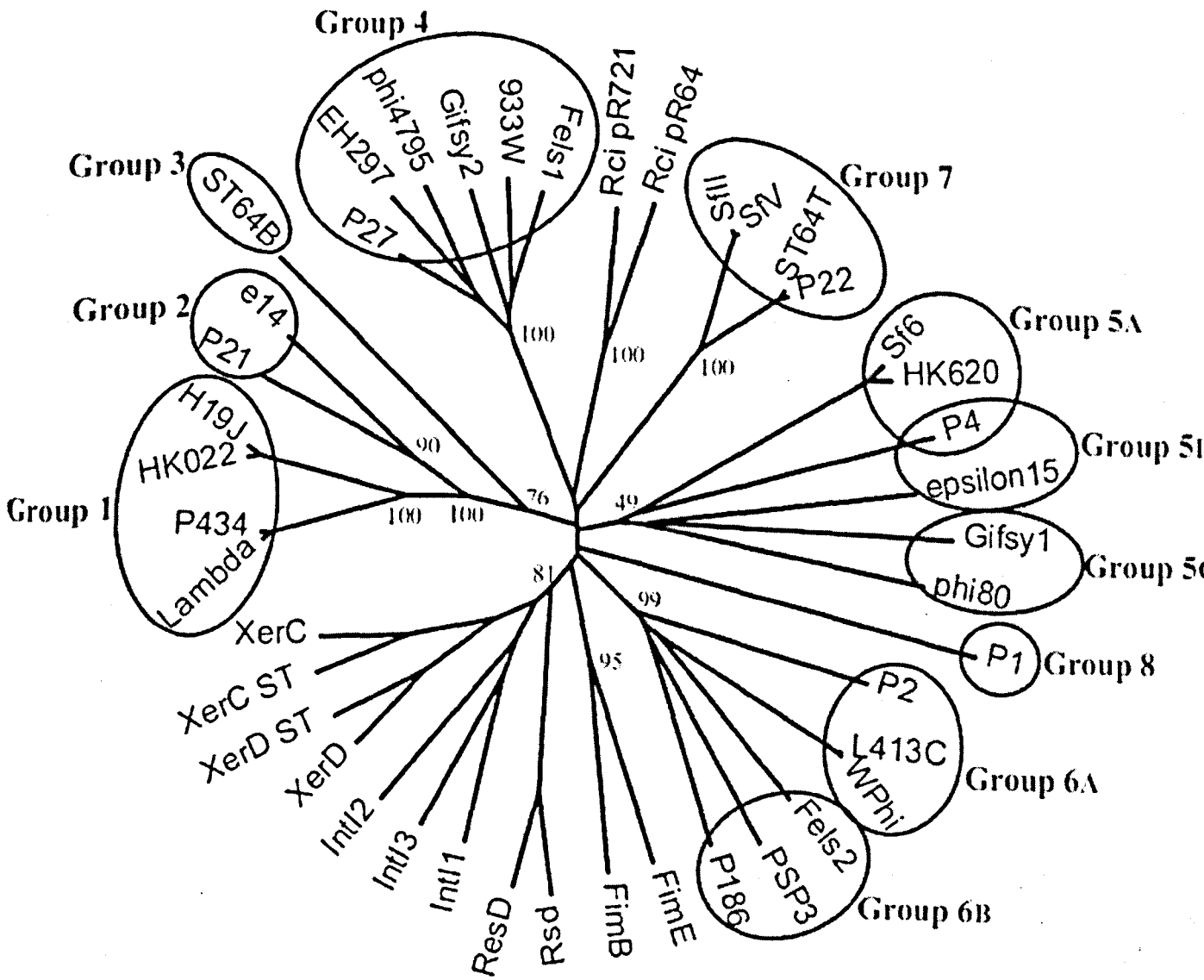


Figure 21. Taxonomic relationships of the conserved domains of 31 enteric phage tyrosine recombinases and 13 tyrosine recombinases from *Enterobacteriaceae* and their mobile genetic elements Phage lambda also represents HK97 and phage 933W also represents VT2-Sa. Circles indicate the groups into which the recombinases were divided for the purposes of PCR primer design (Balding *et al.*, 2005).

4.2 Specific Methods

4.2.1 Wild-Type Stx-Phage Induction and Propagation

All viable STEC samples and inducible bacteriophage were handled and propagated under containment level 3 (CAT 3) conditions. Overnight cultures of STEC isolates were inoculated into 10 ml of LB broth and incubated at 37 °C for 3 h. Norfloxacin (1 μg ml^{-1}) was added to each 10 ml approximately mid-exponential culture and incubated for 1 h at 37 °C. Cells from each culture (1 ml) were allowed to recover via subculture into 10 ml of LB broth which was further incubated at 37 °C for 2 h. The crude phage lysates produced were subjected to plaque assay using 10^{-1}-10^{-3} dilutions to produce semi-confluent lysis. The *E. coli* C strain WG5-rif^R was used as a host in the presence of 150 μg ml^{-1} rifampicin, which is sufficient to limit the growth of any STEC cells. After overnight incubation at 37 °C, the plates were flooded with 3 ml SM buffer and incubated overnight at 4 °C. The SM buffer (1 ml) was transferred to a fresh 1.5 ml tube and heated to 90 °C for 10 min. These heat-treated phage preparations were removed from the category level three containment laboratory and subsequently subjected to direct PCR amplification with a bank of integrase primers (Table 8) (2 μl of each phage preparation was used as an amplification template).

Table 8. Integrase primer names, sequences, degeneracy, length and known phage amplifiable Group 1-8 from Balding *et al.* (2005)

Grp	Phages	Primer Name	Primer Sequence	Length
1	Lambda, HK97, HK022,	1F	5'-GTTACMGGGCARMGAGTHGG-3'	20nt
	P434, H19J	1R	5'-ATGCCCGAGAAGAYGTTGAGC-3'	21nt
2	P21, e14	2F	5'-GTTACTGGWCARCGKTTAGG-3'	20nt
		2R	5'-GATCATCATKRTAWCGRTCGGT-3'	22nt
3	ST64B	3F	5'-AATGGARATWKCYTATYTVTGTGC-3'	24nt
		3R	5'-TCRTARTCTGARATYCCYTTBGC-3'	23nt
4	P27, 933W, Gifsy-2, EH297,	4F	5'-CTBGCMTGGGARGATATHGA-3'	20nt
	VT2-Sa, Fels-1, phi4795	4R	5'-GMCCAGCABGCATARGTRTG-3'	20nt
5A	P4, HK620, Sf6	5AF	5'-TGGRAKRAMKTCGAYTTYGA-3'	20nt
		5AR	5'-CAGTTGCMYYTCWATMGCGTCA-3'	22nt
5B	P4, epsilon15	5BF	5'-TWGTKCGTWMMAGTGAATT-3'	19nt
		5BR	5'-TKGWTRTATRCCGCWCGYAC-3'	20nt
5C	phi80, Gifsy-1	5CF	5'-GGRMARTYATAAAACKSG-3'	18nt
		5CR	5'-TGCCCGAGCAKCWTYTCA-3'	18nt
6	P2, P186, WPhi, Fels-2, L-	6AF	5'-CTGAGYACWGGAGSAMGWTGG-3'	21nt
	413C	6BF	5'-TVGCWACYGGCGCMMGRTGG -3'	20nt
		6R	5'-CCBCCRTTMATCATRAARTG-3'	20nt
7	P22, ST64T, SfII, phageV,	7F	5'-AACATYATMAAYCTKGARTGGCA-3'	23nt
	DLP12, qsr'	7R	5'-CGAACCATTTCKATRGACTCCCA-3'	23nt
8	P1	8F	5'-TGCTTATAACACCCTGTTACGTAT-3'	24nt
		8R	5'-CAGCCACCAGCTTGCATGATC-3'	21nt
$\Phi 24_B$	$\Phi 24_B$, phage-86	24_B *int* F	5'-GCCAGGCTTTCTGAGCTACG-3'	20nt
		24_B *int* R	5'- GCCTAAAATCATGCGTTCTCC-3'	21nt

4.3 Results

113 wild-type Stx-phage inductions were used as templates for PCR amplification with a bank of integrase primers (Table 8) with the aim of discovering the extent of *int* gene variation within the populations. In addition, seven further STEC strains were obtained from a previous study in which only one possessed an inducible phage with an identifiable integrase (Johansen *et al.*, 2001). The presence of *stx* genes in each lysate was confirmed by amplification with general *stx* primers (Smith et al., 2007b) capable of recognising all known variants of *stx1* and *stx2*. The *E. coli* C propagation strain, WG5, was used as a negative PCR control to ensure that no false positives of amplified bacterial chromosome DNA were obtained. The majority of phage amplifications revealed the presence of several integrase bands in each sample resulting either from multiple phages present in the template or alternatively multiple partial or functional integrase gene copies within a single phage. Only 37% of the samples contained a single integrase PCR amplification band; 41% gave two amplified bands; 20% comprised three or more (Fig. 22 and 23).

Of the twelve groups of integrases detectable with the bank of integrase primer sets, two groups were present in the vast majority of samples examined. The integrase of the well studied Stx-phage 933W (group 4) was present in 67% of inductions and the integrase of the Stx-phages $\Phi 24_B$ was detected in 58% of samples (Fig. 24). Furthermore, the percentage of samples containing one or other of these integrases was 88%. Otherwise, the highest frequency of detection was for groups 1 (19%) and 6B (14%). Only 3% of the isolates from which Stx phages could be induced contained an integrase that could not be identified with this bank of primers.

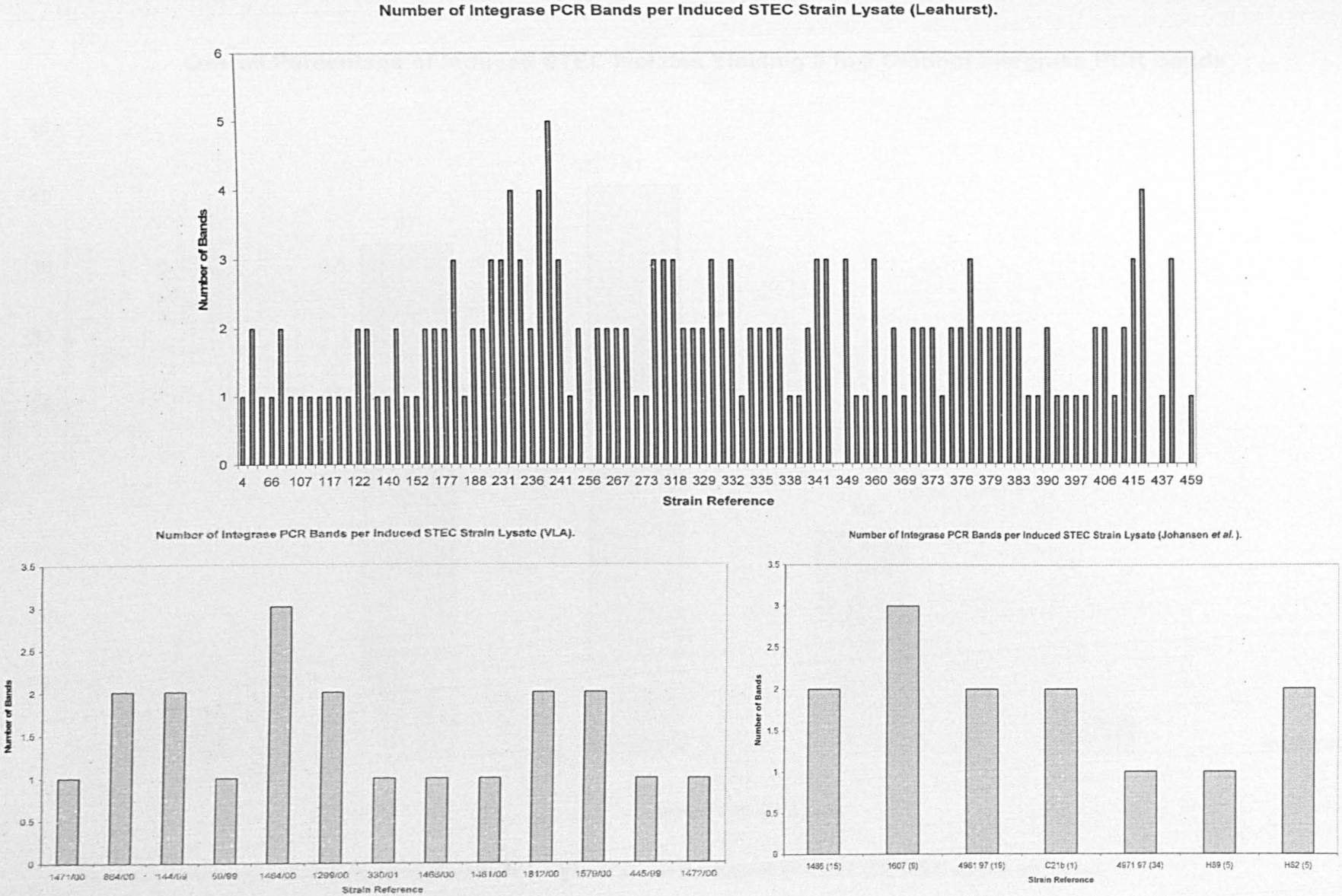


Figure 22. Histograms representing the number of integrase bands present in induced STEC cultures The number of integrase bands detected, by application of the integrase bank of primers, from three separate groups of isolates- Leahurst samples, VLA and seven strains detailed in Johansen *et al.* (2001). For full list of strain reference numbers and the number of integrases associated with each see Appendix I.

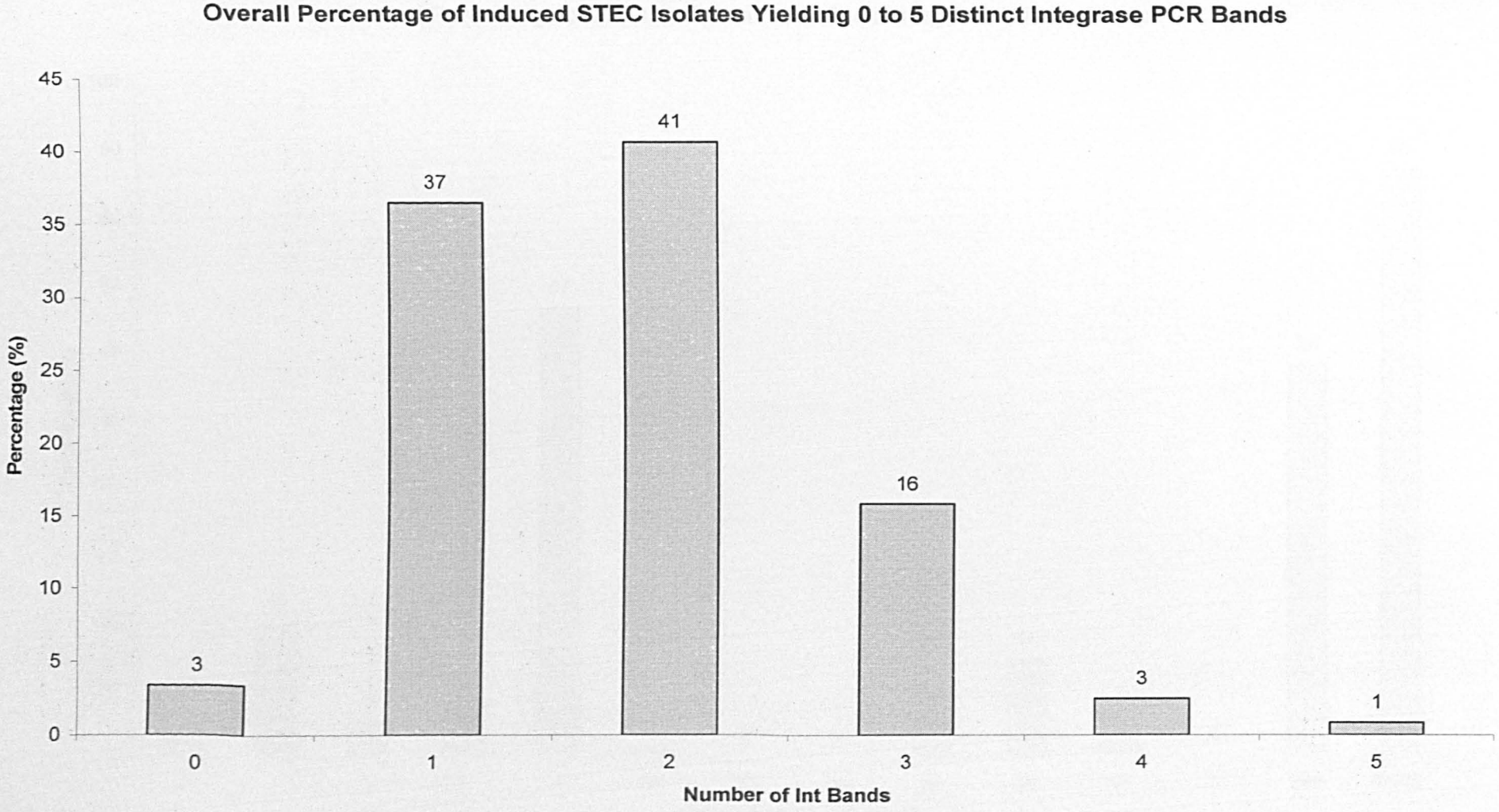


Figure 23. Summary of the distribution of multiple integrases across the full collection of 120 STEC strains examined

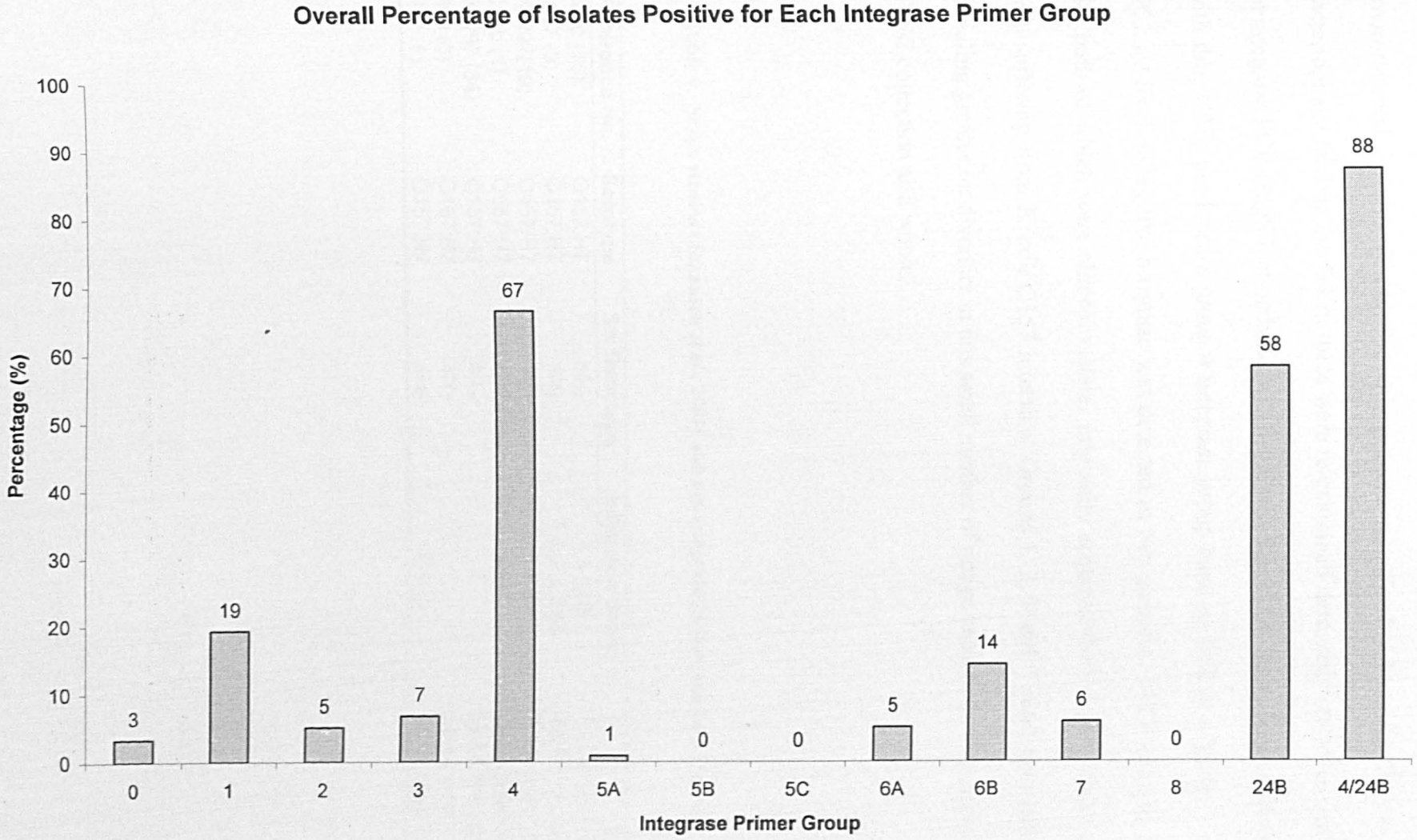


Figure 24. Distribution of different integrase groups across the STEC collection (Detected with the bank of primers listed in Table 8) 0 corresponds to no identifiable integrase gene present. For full list of strains plus integrase groups present in each, see Appendix I.

Johansen *et al.* (2001) attempted to amplify integrase genes from seven clinical and environmental isolates (Table 9) with primers specific for both 933W *int* and bacteriophage lambda *int*. When these were re-examined here, only one strain yielded an integrase PCR amplification band, C21b (1), with the 933W *int* primers. Consistent with this, C21b produced a group 4 integrase group band as well as a group 7 band (P22 *int*-like). $\Phi 24_B$-like integrase was detected in two samples, 1480 (15) and 1607 (9), both of which were clinical isolates originally acquired from a parent and child both suffering from *E. coli* O157 infection. Groups 1, 2, 3 and 7 were also detected revealing greater *int* diversity in this small number of phage inductions than across the STEC collection as a whole.

Table 9. Seven strains (Johansen *et al.*, 2001) and the integrase groups amplified by PCR

Reference No.	Serotype	*Stx* Genotype	Integrase Group	Source
1480 (15)	O157:H7	stx_2	2 & 24B	Sick Child (Heir *et al.*, 2000)
1607 (9)	O157:H7	stx_2	1 & 2 & 24B	Healthy Mother (Heir *et al.*, 2000)
498/97(19)	O157:H7	stx_2	1 & 2	Cattle Herd 2 (Vold *et al.*, 2001)
C21b (1)	O157:H7	stx_1 *and* stx_2	4 & 7	Cattle Herd 3 (Vold *et al.*, 2001)
497/97 (34)	O157:H7	stx_2	2	Cattle Herd 2 (Vold *et al.*, 2001)
H89 (5)	O157:H7	stx_2	7	Cattle Herd 1 (Vold *et al.*, 2001)
H82 (5)	O157:H7	stx_2	1 & 3	Cattle Herd 1 (Vold *et al.*, 2001)

4.4 Discussion

The reported mosaicism amongst the lambdoid phages, including Stx-phages, suggested that there would be a variety of integrase genes present within populations. However, this is not the case in the STEC collection examined here. 88% of samples contained either a 933W-like or $\Phi24_B$-like integrase gene. Prevalence of each of the other ten primer groups was < 20% with most (8 of the ten) < 10%. The lack of $\Phi24_B$ *int*-like genes in the GenBank database meant that broad specificity primers were not able to be designed. However, the target sequences chosen encompassed 3 of the conserved RHRY residues, including the catalytic tyrosine, and are 100 % identical to phage-86 *int* sequence, the only know comparable integrase gene. Therefore, the high incidence of the $\Phi24_B$–like integrases is all the more surprising because the $\Phi24_B$ primer pair contained no degenerate bases and as a consequence was far more stringent than the rest of the bank of primers. Accordingly, Smith *et al.* (2007b) found that, despite the diversity of bacteriophage tail genes amongst coliphages in general, 99% of the phages screened contained a gene homologous to the short tail spike gene previously identified in the sequenced Stx-phages 933W and Vt2-Sa (Makino *et al.*, 1999; Plunkett *et al.*, 1999) and also present in $\Phi24_B$ (Smith et al. 2007b). Although the lack of heterogeneity in tail genes correlates with the integrase data presented here, the reason for this low integrase diversity is less apparent. Tail spikes are involved in host cell recognition and therefore may be under the influence of powerful selection if the host receptor is well conserved *e.g.* YaeT (Smith *et al.*, 2007a); however integrases are not subject to the same pressures. Integrases are site specific recombinases and as such require complimentary DNA sequence between the *attP* and *attB* sites, thus the diversity of integrases utilised by Stx-phage could be limited by the presence or absence of the required *attB* sequence in the *E. coli* host.

Nevertheless, several of the rarely detected integrase groups comprise members able to recognise sequences in the *E. coli* genome; for example the group 5A integrases of Sf6 and HK620 recognise a site less than 100 bp upstream of the $\Phi24_B$ integration site (Casjens *et al.*, 2004; Clark *et al.*, 2001) and yet were detected in only one sample. It may be that the STEC hosts prevalent in the environments screened do not possess the required insertion sites for the various non-$\Phi24_B$ integrases or that, due to the increased remnant prophage supplementary DNA usually found in STEC and EHEC strains, the integration sites could already be occupied.

The data presented here suggests that the integrases possessed by bacteriophages 933W and $\Phi24_B$ are positively selected for *in vivo*. The underlying rationale for this conservation is, nevertheless, unresolved. It may be that these integrases confer a competitive advantage to the phage, for example, through increased efficiency of integration or a preferential site of insertion, less disruptive to host cell biology. Indeed the localisation of lambdoid *attB* sites, and the loci of defective bacteriophages, within a section of the *E. coli* genome (6-44 minutes) (Bachmann, 1990) alludes to the latter option or a combination of the two. It's also feasible that a competitive advantage may be conferred upon a phage, by the carriage of these integrase genes, if it is the integrase gene, or its associated regulatory mechanisms, that facilitate the production of multiple lysogens. If a resident prophage integrase was not under tight negative regulation, Int protein could build up in the cytoplasm and superinfection immunity to a second homoimmune bacteriophage could be negated. This mechanism for multiple lysogeny would allow for a diverse intracellular gene pool for recombination, whilst the integrase gene itself would be conserved. Although bacteriophage 933W has not been reported to form double lysogens, this has not been

conclusively proven either way and therefore the same pressures may apply to it as for $\Phi 24_B$ *int*, potentially explaining its pervasiveness in the isolates screened. In addition, $\Phi 24_B$ integrase has been shown here to be tolerant of a high degree of base pair mismatches in the complementary *attB* overlap region, i.e. up to 50 %. This characteristic could prospectively significantly increase the host range of a phage carrying the gene, and increase tolerance to the loss or occupation of the preferred integration site.

Alternatively the integrase preference may be in part due to a localised sampling bias. Only two of seven isolates obtained from Norway (Johansen *et al.*, 2001) contained a $\Phi 24_B$ integrase and only one possessed a 933W-like, group 4, integrase. Conversely, both the VLA and Leahurst samples (from different regions of the UK) do seem to have correlating integrase preferences, implying that location does not bias integrase prevalence, within the UK at least. The larger sample sizes of the Leahurst and VLA samples compared to the Norwegian isolates would suggest a more reliable nature of the data obtained from them. Also at least two of the seven Norwegian strains were likely to be identical (infected mother and child), compared to more random environmental samples and veterinary isolates from various sources. Clearly, investigation of further STEC samples, of both clinical and environmental origin, from a wider range of geographical areas would help to determine conclusively any preference for particular integrases.

4.5 Conclusions

- Bacteriophage integrase diversity in the STEC populations studied displayed surprisingly limited diversity, with the majority of integrase types belonging to one of two groups (933W-like and $\Phi 24_B$-like).
- The rationale underlying the lack of diversity is unclear but may be due to either the restricted number of sampling sites or an unknown selective advantage offered by those integrase types e.g. efficiency of integration, site of integration or ability to direct multiple lysogeny.
- Norwegian clinical and environmental isolates contained a greater variety of integrase types, however the small sample size (7 strains) makes conclusive commentary difficult.

Chapter 5: Discovery and Initial Characterisation of a P22-like Antirepressor Gene in $\Phi 24_B$

5.1 Background

Several temperate bacteriophages encode antirepressor (Ant) proteins that, when expressed, are responsible for the deactivation of repressor molecules. Under normal circumstances, these antirepressors are tightly regulated during both lysis and lysogeny. The best described Ant-encoding phage is the *Salmonella* phage P22 that possesses a bipartite immunity system comprising the disparate *ImmC* and *ImmI* regions (Susskind & Botstein, 1978). The former comprises an operon analogous to the lambda immunity region, whilst the later encodes an antagonistic system including the Ant protein. The 300 residue antirepressor protein does not function by interaction with the DNA operators but rather exerts its effect directly onto the P22 repressor, cII (Echols, 1972). The site of Ant binding to the repressor is the same as that recognised during RecA-mediated proteolytic cleavage. However, in this case the repressor protein is non-covalently bound and as such is recoverable *in vitro* in contrast to RecA alleviation of repression (Prell & Harvey, 1983; Roberts & Roberts, 1975). The activity of the antirepressor is not limited to specific deactivation of the P22 repressor. It is not only effective against the closely related *Salmonella* phage L and Px1 immunity systems but also super-infection of a lambda lysogen with bacteriophage P22 results in the elimination of lambda repressor activity (Susskind & Botstein, 1975).

Primarily, 2 repressors, Mnt and Arc, regulate transcription of the *ant* gene, both of which display a high degree of homology to lambda Cro (Sauer *et al.*, 1983). Mnt is

constitutively expressed during lysogeny and as such subdues production of antirepressor. Mnt is an 82 amino acid protein that exists in a tetrameric conformation in solution. The tetramer specifically binds to a 17 bp operator site, which is symmetrical about a single central base pair (Silbaq *et al.*, 2002). Each half of the operator is recognised by the symmetrically related Mnt dimers (see figure 25A). P22 Mnt deficient lysogens infected with P22 *mnt*$^{-}$ *ant*$^{+}$ phage display a loss of immunity to super-infection coupled with the induction of the resident prophage (Vershon *et al.*, 1985). The *arc* gene product also negatively regulates antirepressor expression during lytic infection via the co-operative binding of two Arc repressor dimers to a 21 bp operator site (see figure 25B). Each individual Arc dimer consists of matching, entwined monomers that form a ribbon-helix-ribbon DNA binding motif. In the co-crystal structure, the two Arc dimers each recognise and bind adjacent subsites of the operator site (Dostal *et al.*, 2005; Raumann *et al.*, 1994). A third negative regulatory system also contributes to the control of Ant expression during the lytic life cycle via anti-sense RNA inhibition, sar RNA. Sar RNA acts *in trans* by binding to the ant mRNA ribosome-binding thus site preventing translation (Liao *et al.*, 1987; Wu *et al.*, 1987). The presence of the Ant protein and the entire *ImmI* region are non-essential for the normal P22 life-cycle (Botstein *et al.*, 1975; Levine *et al.*, 1975). Furthermore constitutive over-expression of Ant results in a failure to produce phage progeny (Susskind, 1980).

Another well-described antirepressor system is that encoded on the closely related temperate bacteriophages P1 and P7 (Heinrich *et al.*, 1995). Both phages possess a cI repressor protein analogous to lambda cI and an antirepressor protein similar to P22 Ant. The repressor protein of each phage, obtained from crude cellular extracts, co-

precipitates with the Ant1/2 antirepressor complex indicating a direct interaction during derepression. Ant1/2 also requires an additional 'loading-site' specific to each phage species, known as the site of Ant specificity or *sas*, in order to function (Riedel *et al.*, 1993a; Riedel *et al.*, 1993b). Regulation of antirepressor synthesis in P1/P7 is controlled by C4 anti-sense RNA repression (Riedel *et al.*, 1993b). Unusually, C4 is transcribed from the same promoter as the antirepressor protein and is produced as a precursor molecule, which must be post-transcriptionally activated to yield the functional product (Heinrich *et al.*, 1994). The active molecule represses synthesis of Icd, a small peptide translationally coupled to Ant. Repression is based upon the interaction of 2 short sequences, a'/a2 and b'/b2, encompassing the *icd-ant* ribosome binding site. This blockage of mRNA binding, and consequent translational repression, leads to the termination of Ant transcription via a Rho-dependent terminator (Biere *et al.*, 1992).

Examination of unannotated $\Phi24_B$ genome fragments revealed the presence of a putative antirepressor gene, similar to P22 Ant. Previous work has described the production of $\Phi24_B$ multiple lysogens (Allison *et al.*, 2003), however the mechanism by which this occurs is not yet clear. The presence of an active antirepressor in $\Phi24_B$ would raise the possibility that $\Phi24_B$ Ant could be interfering with efficient repression of super-infecting isogenic phage.

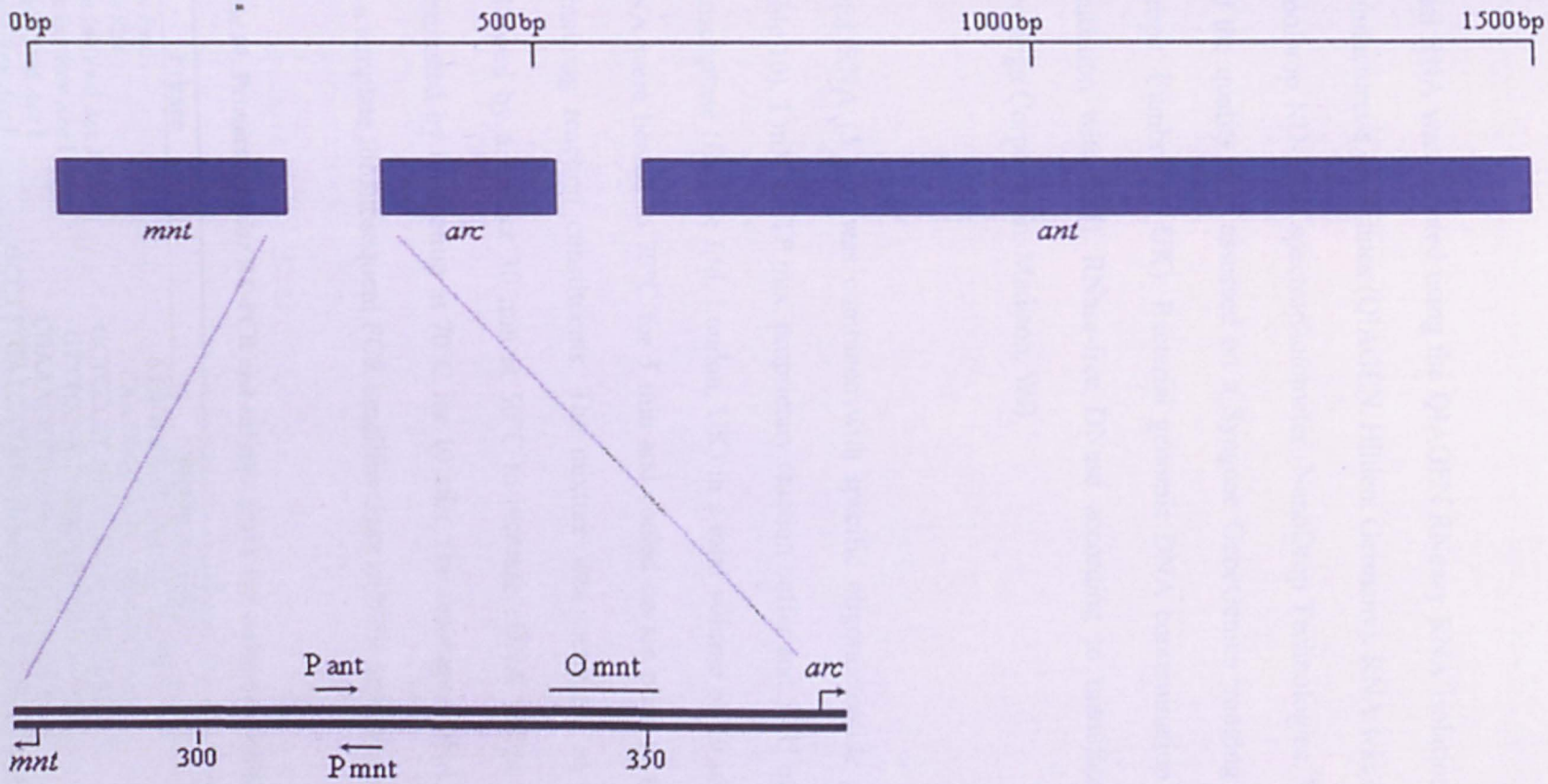


Figure 25. Illustration of P22 *ant* gene and regulatory components The *ant* gene is transcribed from P_{ant} and is regulated by two repressors, Mnt and Arc, which tightly control expression of the antirepressor during lysogeny and lysis respectively.

5.2 Specific Methods

5.2.1 Reverse-Transcriptase PCR

Total RNA was isolated using the QIAGEN RNeasy RNA isolation kit according to manufacturer's guidelines (QIAGEN, Hilden, Germany). RNA was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE), and the quality was assessed on a Syngene GeneGenius imaging system (Syngene Europe, Cambridge, UK). Bacterial genomic DNA contamination was removed by incubation with RQ1 RNase-free DNase according to manufacturer's guidelines (Promega Corporation, Madison, WI)

Total RNA (5 µg) was combined with specific oligonucleotide primers (200 nM, Table 10), 1 mM dNTP mix, proprietary reaction buffer and 50 U of BioScript reverse transcriptase (Bioline Ltd, London, UK) in a total volume of 30 µl. The primers and RNA were heated to 70°C for 5 min and cooled on ice prior to the addition of the remaining reaction constituents. The mixture was incubated at 42°C for 30 min followed by a further 30 min at 50°C to increase cDNA yields. The reaction was terminated by incubation at 70°C for 10 min. The resultant cDNA was used directly as a template for subsequent PCR amplifications with the specific primers.

Table 10: Primers used for RT-PCR and *ant* gene knock-out vector production

Primer	Sequence (5'-3')	Annealing
ant Fwd	ATGGCTGTTCAAATTTCTGTCG	60 °C
ant Rev	CAATGCCTGTGAGCGGTTGG	60 °C
ant Inv Fwd *Acc* I	GCTCCCTCATCCTGTAGACAGAACG	60 °C
ant Inv Rev *Acc* I	GTTTCGGGTCTACTCTTAAACTCGC	60 °C
aadA Fwd *Acc* I	CGAAACCTTGCGGTCGACCGCCAGCC	60 °C
aadA Rev *Acc* I	GCCTTTCATGATATGTCGACCAATTTGTGTAGGGC	60 °C

5.2.2 Strategy to Abrogate $\Phi 24_B$ *ant* Gene

The $\Phi 24_B$ putative P22-like antirepressor, *ant*, was amplified with *ant* Fwd and *ant* Rev primers and the relevant parameters (Table 10) using EXPAND high fidelity polymerase, purified using the Eppendorf Perfect Prep Gel Extraction Kit and cloned into Zero Blunt cloning vector according to manufacturers' protocols. Plasmid from transformants was purified using QIAGEN's mini-prep kit and subjected to PCR amplification using *ant* Inv Fwd *Acc* I and *ant* Inv Rev *Acc* I oligonucleotides with the relevant parameters (Table 10). The resulting amplification product was purified and 1 µg digested with the restriction endonuclease *Acc* I for 1 h at 37 °C. The antibiotic resistance cassette, *aadA*, encoding resistance to spectinomycin, was amplified with *aadA* Fwd *Acc* I and *aadA* Rev *Acc* I oligonucleotides and EXPAND High Fidelity Polymerase with the relevant parameters (Table 10). After subsequent purification 1 µg of the amplified resistance cassette was digested with the restriction enzyme *Acc* I for 1 h at 37 °C. The spectinomycin resistance cassette possessing *Acc* I 3' overhangs was then cloned into the Zero Blunt backbone containing the *ant* flanking regions and compatible *Acc*I cohesive ends. The resulting plasmid (Fig. 26) was purified using QIAGEN's mini prep kit and the interrupted *ant* gene was confirmed by PCR using combinations of *ant* Fwd, *ant* Rev, *aadA* Fwd *Acc* I and *aadA* Rev *Acc* I primers with Taq recombinant DNA polymerase and relevant parameters (Table 10).

The insert was then excised from the plasmid backbone by sequential digest using the restriction endonucleases *Age* I and *Sal* I. Following heat inactivation of the restriction endonucleases the construct was ligated into *Age I*/*Sal* I digested suicide vector, pKNG101. This suicide plasmid possesses an R6K origin of replication (λ pir

dependent) and the *sacB* gene, which when expressed is lethal to *E. coli* when grown on > 5% sucrose media. The constructed pKNG101 + *ant*Δ*aadA* was introduced into chemically-competent $\Phi 24_B$::Kan lysogens. Selection for spectinomycin resistance allowed the identification of clones produced by a single recombination event, which were confirmed by PCR. Passage of an established single recombinant followed by positive selection for the loss of *sacB*, *i.e.* any cells able to grow on 7.5% sucrose, enabled the identification of an *ant* knock-out in $\Phi 24_B$::Kan. Ablation of the gene was confirmed by PCR with combinations of *ant* Fwd, *ant* Rev, *aadA* Fwd *Acc* I and *aadA* Rev *Acc* I primers with Taq recombinant DNA polymerase under relevant conditions (Table 10). The process was repeated for $\Phi 24_B$::Cat lysogens.

The presence of *sacB* results in cell death in the presence of sucrose as cytotoxic metabolic intermediates build up within the cell, allowing positive selection for the loss of the plasmid during a double recombination event. Use of this plasmid was problematic for several reasons. Firstly, propagation and preparation of plasmid DNA was difficult due to a low plasmid copy number and volatility of the plasmid. The low copy number problem was overcome by simply increasing the volume of culture subjected to plasmid purification from 1 ml to 10 ml. The volatility issue was more difficult as storage of plasmid containing cultures frequently resulted in loss of plasmid during subsequent propagation. To overcome this, plasmid DNA was frozen and electroporated into competent cells to produce fresh plasmid cultures as needed. A further difficulty encountered was the process of selection for plasmid loss following a double recombination event. The *sacB* system requires fine tuning of the growth media. All NaCl had to be removed and the precise percentage of sucrose, lethal to *E. coli* K-12 strain MC1061, had to be determined. A range of sucrose

concentrations, 5 % - 12.5 % (w/v), were tested for lethal effects on cells carrying the pKNG101 plasmid. 7.5 % sucrose efficiently killed all of the pKNG101 test cells whilst allowing normal confluent growth of the naïve control strain. Following growth media optimisation, there still remained spurious growth on occasion, which made selection of true double recombinant colonies difficult. As a result, secondary PCR screening was required to identify the true positive colonies.

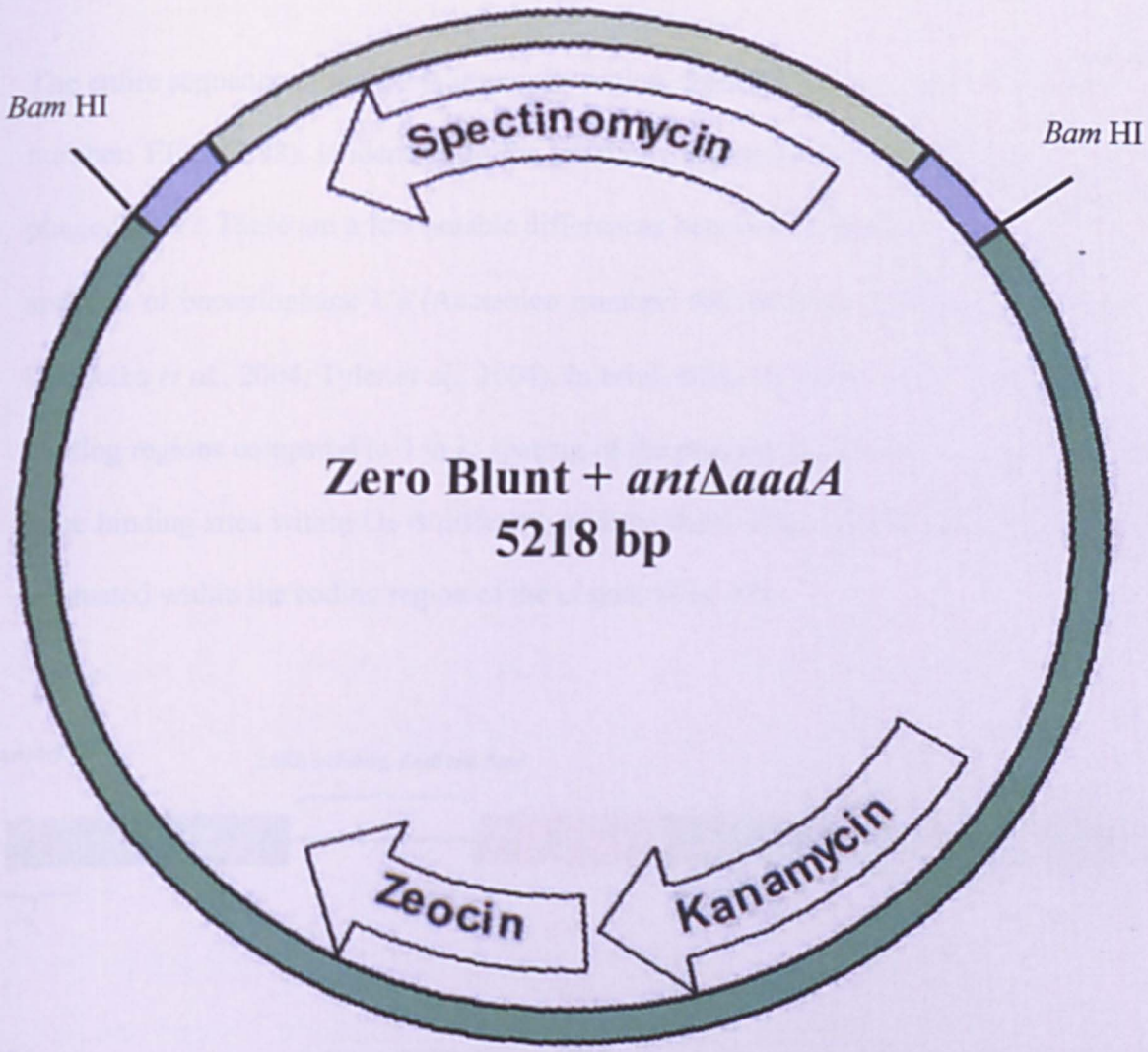


Figure 26. Plasmid map of the intermediate antirepressor knock-out construct The contruct consists of *ant* interrupted by a spectinomycin resistance cassette in a Zero Blunt cloning vector. Locations of *Bam* HI restriction endonuclease cut sites required for insert excision are included.

5.3 Results

The entire sequence of the $\Phi24_B$ immunity region, from O_L through to *cro* (Accession number: EF 517298), is identical to the immunity region of the fully sequenced Stx-phage 933W. There are a few notable differences between the $\Phi24_B$ immunity region and that of bacteriophage λ's (Accession number: NC_001416) (Fattah *et al.*, 2000; Koudelka *et al.*, 2004; Tyler *et al.*, 2004). In brief, $\Phi24_B$ O_L possesses only 2 operator binding regions compared to 3 in λ; spacing of the promoters $_pR$ and $_pRM$ between the three binding sites within O_R is different; and the third $\Phi24_B$ operator binding region is situated within the coding region of the *cI* gene (Fig. 27).

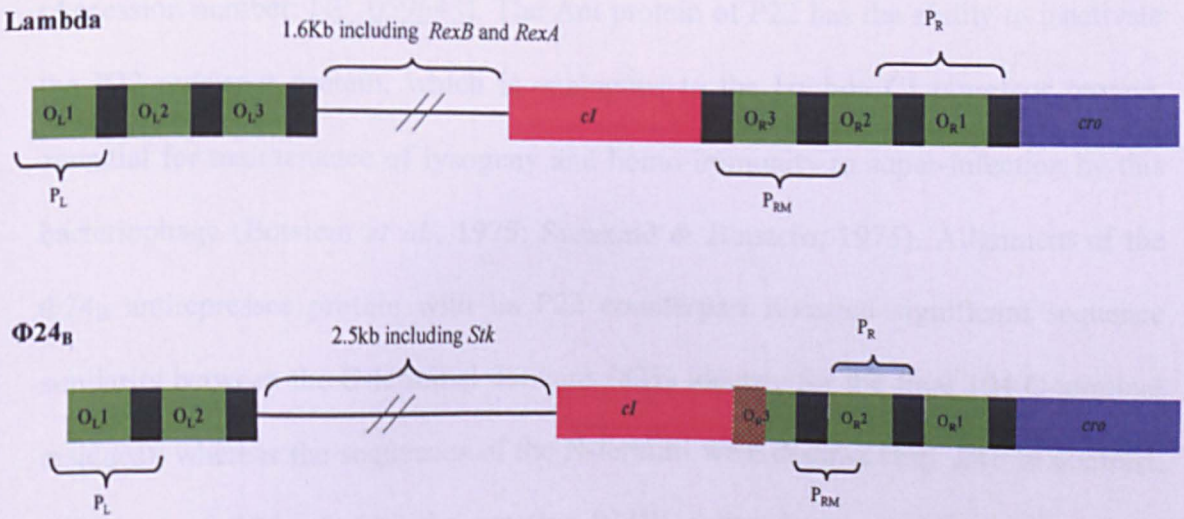


Figure 27. Comparative immunity regions of $\Phi24_B$ and lambda

The differences between the respective immunity regions of bacteriophage lambda and $\Phi24_B$ offer the possibility that there may be sufficient disparity in function to allow the production of multiple $\Phi24_B$ lysogens in a single *E. coli* host via reduced repressor function or efficiency. Previous work (Fattah *et al.*, 2000; Koudelka *et al.*, 2004; Tyler *et al.*, 2004) suggests that this hypothesis is unlikely as the identical

2004; Tyler *et al.*, 2004) suggests that this hypothesis is unlikely as the identical 933W immunity region conforms to the lambda immunity model. Furthermore, multiple lysogens of 933W have never been reported.

During the course of this research 10 µg of $\Phi24_B$;;Cat phage DNA was purified and submitted to the Wellcome Trust Sanger Institute, Cambridge, for whole genome sequencing. The full genome sequence did not become available until close to the conclusion of this research project, however, unprocessed contigs of various lengths were provided in advance. Examination of these unannotated fragments of the $\Phi24_B$ genome sequence revealed the presence of a putative antirepressor gene possessing 37% amino acid homology to the well characterised P22 antirepressor gene, ant (Accession number: NP_059643). The Ant protein of P22 has the ability to inactivate the P22 repressor protein, which is analogous to the lambda CI repressor protein, essential for maintenance of lysogeny and homo-immunity to super-infection by this bacteriophage (Botstein *et al.*, 1975; Susskind & Botstein, 1975). Alignment of the $\Phi24_B$ antirepressor protein with its P22 counterpart revealed significant sequence similarity between the C-terminal domains (82% identity for the final 104 C-terminal residues), whereas the sequences of the N-termini were distinct (Fig. 28). In contrast, alignment of $\Phi24_B$ Ant to the putative 933W antirepressor, which appears to be truncated and incomplete, produced an apparent low level of shared amino acid sequence (12% sequence identity).

In order to establish if the *ant* gene is transcribed in $\Phi24_B$ lysogens, total RNA from mid-log naïve and lysogen cultures was harvested. In addition, total RNA of mid-log *E. coli* MC1061 cultures undergoing $\Phi24_B$ infection at two multiplicities of infection

(0.1 and 1.0) was also harvested. The RNA samples were subjected to reverse-transcription PCR with internal, *ant* specific primers (Table 10), expected to produce a 456 bp fragment. The use of $\Phi24_B$ *int* Fwd/Rev primers, which should produce a fragment of 414 bp in length, served as an internal positive control for the RT-PCR reaction. In this way, the presence of *ant* mRNA transcripts was identified in the lysogen and infection samples, and were found to be absent in the control RNA harvested from the naïve *E. coli* cultures (Fig. 29). Thus, this $\Phi24_B$ gene is actively expressed during bacteriophage lysogeny/infection and, as demonstrated for phage P22 (Susskind & Botstein, 1978), could explain the absence of homo-immunity leading to the creation of multiple lysogens by the lambdoid Stx-phage described here.

To definitively determine whether Ant influences the formation of multiple $\Phi24_B$ lysogens, a strategy was embarked upon to disrupt the *ant* gene by replacement of the majority of the gene with a spectinomycin resistance cassette (*aadA*). Initial sequence data did not include the entire *ant* gene sequence, thus limiting the ability to produce an effective ablation approach. In addition, construction of the interrupted *ant* gene proved difficult due to problems associated with inverse PCR and restriction endonuclease cleavage site incorporation into oligonucleotide primers. Subsequent receipt of complete sequence data allowed a more comprehensive gene knock-out strategy to be developed, including larger flanking regions to increase recombination frequency/efficiency and a more significant degree of gene replacement. The PCR complications encountered previously were overcome and an intermediate plasmid constructed using Zero Blunt cloning vector as a backbone (Fig. 26). The *ant*Δ*aadA* insert was transferred to the suicide plasmid pKNG101, which contains the

conditional lethal gene *sacB*. Replacement of the $\Phi 24_B$::Kan putative *ant* gene was achieved, however, due to time constraints, deletion of the $\Phi 24_B$::Cat *ant* gene was not completed. Assessment of biological function of $\Phi 24_B$::Kan *ant*$^-$ was also not investigated.

Figure 28

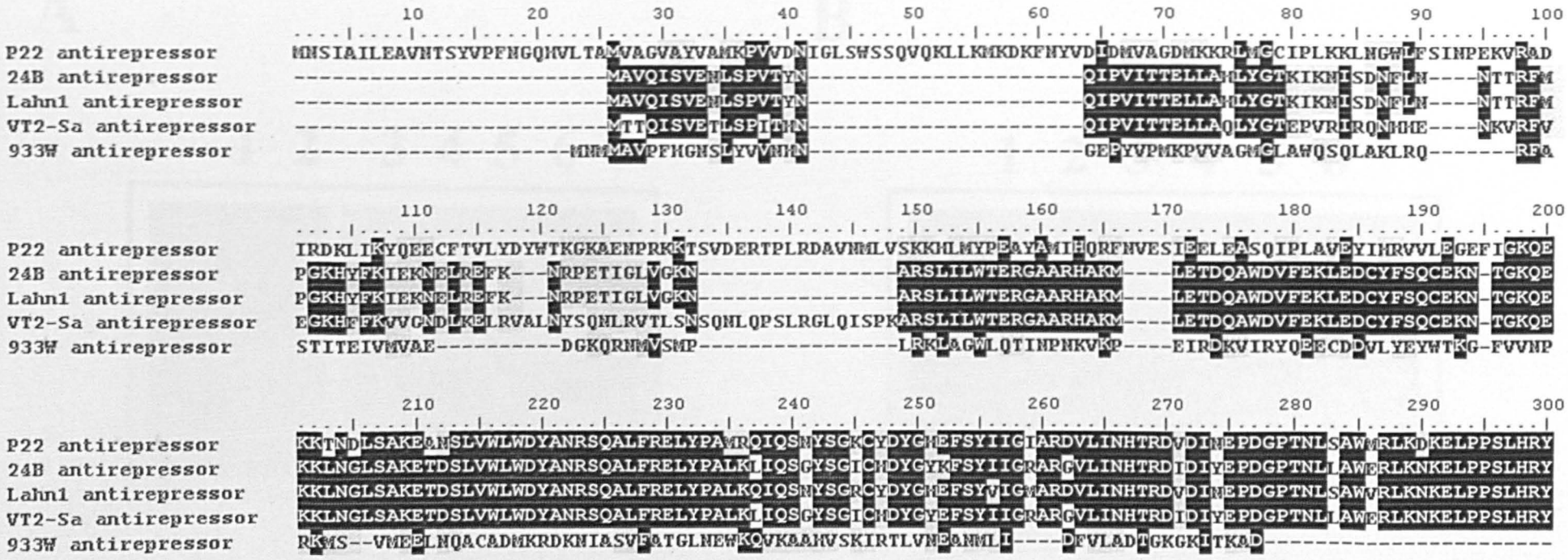


Figure 28. Comparison of $\Phi 24_B$ antirepressor with published homologues ClustalW multiple sequence alignment of $\Phi 24_B$ Ant to the well defined P22 Ant as well as the two closest Blastx matches (Lahn2 and VT2-Sa antirepressors) and the Ant protein of the well characterised Stx-phage, 933W.

Figure 29

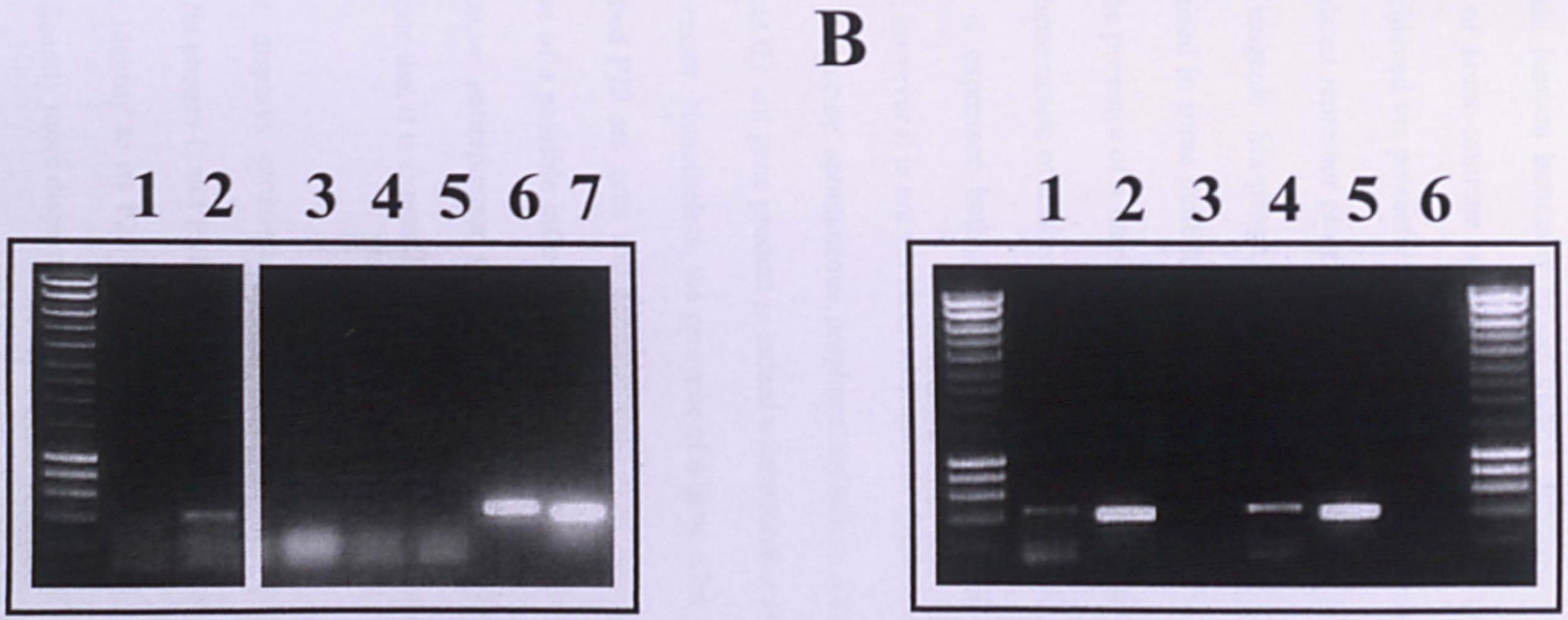


Figure 29. Analysis of $\Phi 24_B$ *ant* transcription RT-PCR amplifications of mid-log naïve MC1061 and $\Phi 24_B$ lysogen RNA demonstrating transcription of $\Phi 24_B$ *ant* in the lysogen only. Panel A. Lane 1: MC1061 cDNA + ant F/R; Lane 2: Lysogen cDNA + ant F/R; Lane 3: MC1061 RNA + ant F/R; Lane 4: Lysogen RNA + ant F/R; Lane 5: MC1061 DNA + ant F/R; Lane 6: Lysogen DNA + ant F/R; Lane 7: Lysogen cDNA + $\Phi 24_B$ *int* F/R. RT-PCR amplifications of mid-log MC1061 undergoing infection at multiplicities of infection of 0.1 and 1.0. Panel B. Lane 1: MoI 0.1 cDNA + ant F/R; Lane 2: MoI 0.1 cDNA + $\Phi 24_B$ *int* F/R; Lane 3: MoI 0.1 RNA + ant F/R; Lane 4: MoI 1.0 cDNA + ant F/R; Lane 5: MoI 1.0 cDNA + $\Phi 24_B$ *int* F/R; Lane 6: MoI 1.0 RNA + ant F/R.

5.4 Discussion

The classic lambda immunity model clearly precludes the production of multiple lysogens of homo-immune lambdoid bacteriophage by consecutive super-infection. This is achieved via powerful repression of transcription in an incoming phage by an auto-regulated repressor protein. It has been unequivocally proven with differentially labelled isogenic Stx-phage, $\Phi24_B$::Cat and $\Phi24_B$::Kan, that immunity can be compromised in some lambdoid phage by an unknown mechanism(s) (Allison *et al.*, 2003). The presence of a functional antirepressor in $\Phi24_B$ is one plausible explanation for the phenomenon of multiple lysogen production. The putative antirepressor found in $\Phi24_B$ is expressed both in a lysogen culture and a culture undergoing active infection; however it is not yet clear whether this expression is occurring within stable lysogens or during spontaneous prophage induction. Furthermore, it has not been proven that the ant gene product is indeed a functional antirepressor active against the $\Phi24_B$ repressor. Nonetheless, the presence of a gene with significant homology to the well studied P22 ant gene and demonstration of its transcription in some capacity is suggestive of a possible influence on $\Phi24_B$ immunity. Having identified the presence of a putative antirepressor gene in the $\Phi24_B$ phage genome, the first step is to demonstrate that it is expressed, and that has been achieved here.

P22 Ant displays cross-specificity with the repressors of the closely related *Salmonella* phages L and Px1 as well as the lambda CI protein, which possesses 31% sequence identity to its P22 homologue. $\Phi24_B$ repressor and P22 repressor proteins are significantly more disparate at the amino acid sequence level, 15.4% identity (Fig. 30), and thus would be expected to possess different binding specificities for any antirepressor acting upon them. This may, therefore, explain the observed divergence

in N-terminal sequence. Less than 10% of the N-termini of the P22 and $\Phi 24_B$ antirepressor proteins are similar, in addition, $\Phi 24_B$ Ant N-terminus is significantly shorter than in P22 Ant (122 and 197 amino acids respectively). In contrast the final 103 C-termini residues possess 81.5% identity suggesting a conserved functional domain. Despite the well documented function and regulation of P22 Ant, the CI protein binding domain has not been elucidated and therefore the hypothesised CI binding activity of the amino-terminus can not be corroborated by experimental data.

In P22, Ant production is tightly regulated during both the lytic and lysogenic life cycles. Analysis of the gene organisation flanking the putative $\Phi 24_B$ antirepressor revealed the absence of any P22-like regulatory operon (Fig. 31). Instead, $\Phi 24_B$ ant and surrounding genes are homologous to those found in the remnant lambdoid prophage VT2-Sa. Function has not been assigned to these genes and therefore no speculation can be attempted on their role in the regulation of any functional antirepressor. The replacement of P22-like regulatory genes with an alternative genetic arrangement or complete absence of direct regulation may explain the altered immunity profile observed for $\Phi 24_B$ compared to P22, which does not form multiple lysogens unless *ant* regulation is compromised. The RT-PCR bands achieved, although not quantitative, are relatively faint in comparison to control bands suggesting a low level of *ant* transcription. A basal level of antirepressor production may be sufficient to allow at least low level expression of integrase from an incoming homo-immune bacteriophage, sufficient to facilitate integration into the host genome whilst allowing an existing prophage to be maintained. This could also account for the apparent increased rate of spontaneous induction observed in $\Phi 24_B$ lysogens compared to bacteriophage lambda *in vitro*. Observation that there is a stronger *ant*

RT-PCR band for a lysogen undergoing infection at an MOI of 1 compared to MOI 0.1 and an uninfected lysogen culture is, although not quantitative, suggestive that expression either occurs from an incoming bacteriophage or is up-regulated in a resident prophage during infection. Alternatively, there is the possibility that *ant* expression is a result of rightward expression during spontaneous induction into the lytic life cycle, rather than from a resident lysogenic phage. In order for this eventuality to occur, any regulatory system present must be divergent from the established P22 system.

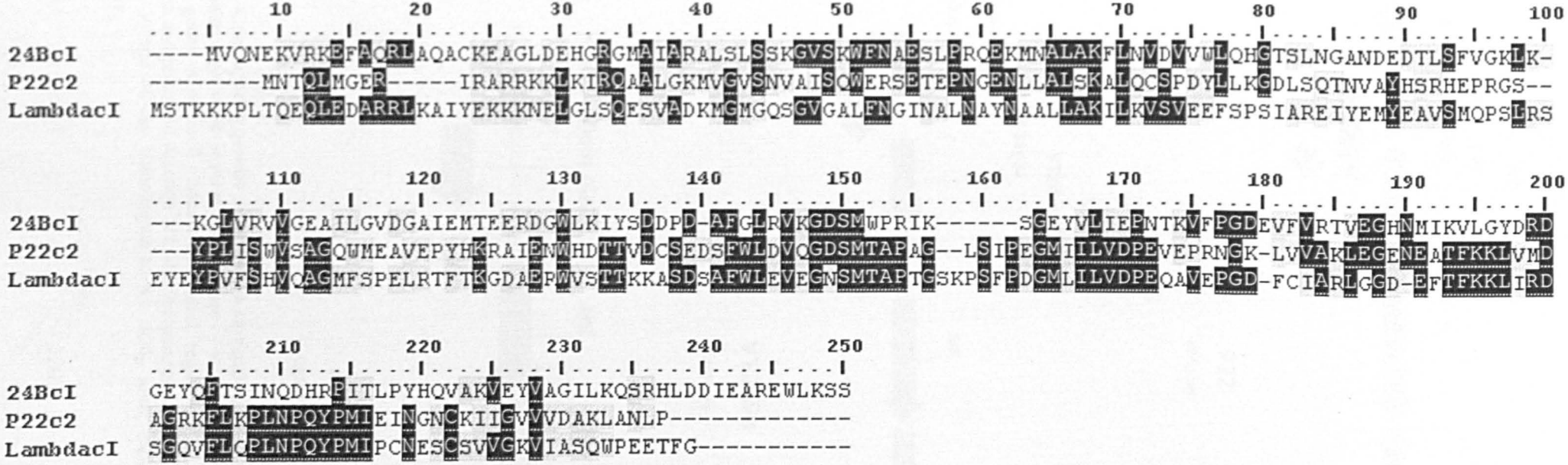


Figure 30. ClustalW multiple sequence alignment of the repressor proteins from P22, lambda and $\Phi 24_B$

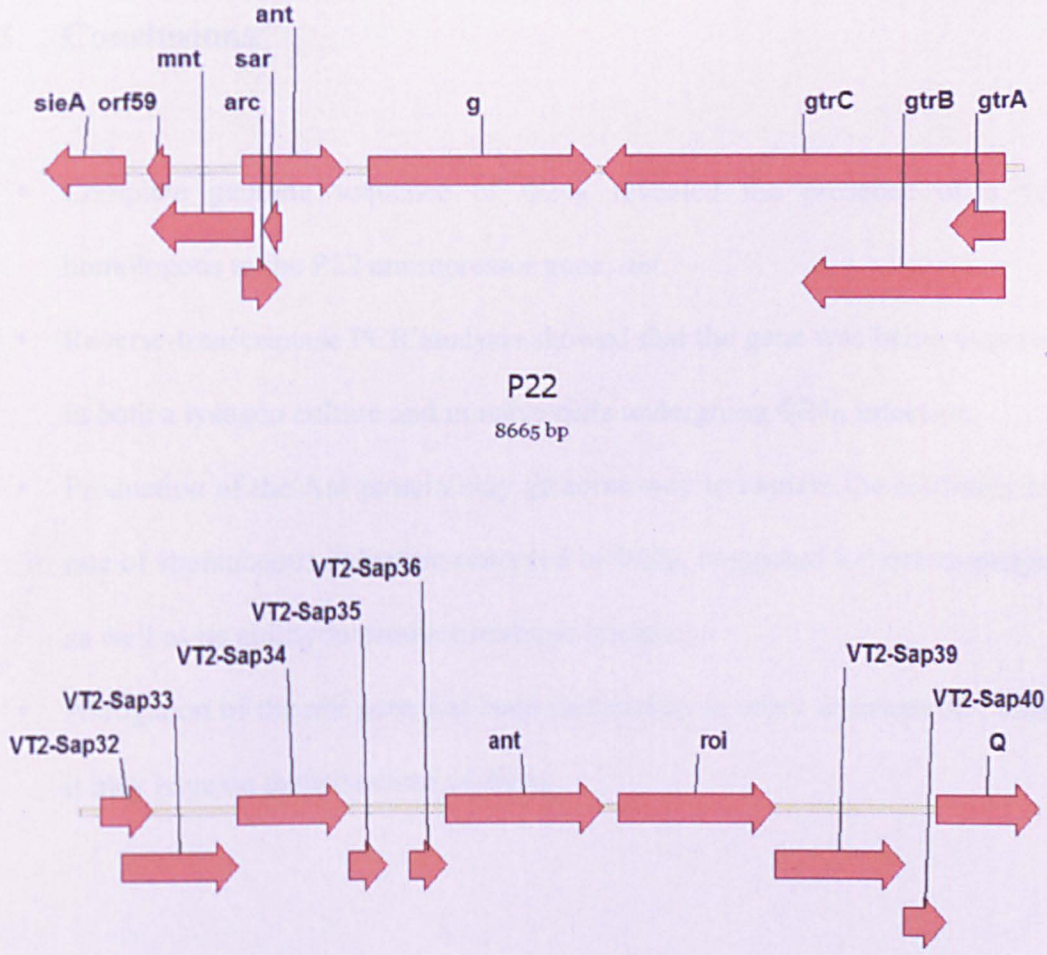


Figure 31. Comparative genetic organisation surrounding the respective antirepressor or putative antirepressor regions of several key phages $\Phi 24_B$ *ant* is flanked by genes dissimilar to the regulatory genes of the well characterised *ImmC* region of P22. The closest known match in gene organisation and presumed function is to the uncharacterised putative *ant* region of Stx-phage VT2-Sa. Sap33-35, 39 and 40 are homologous to the $\Phi 24_B$ *ninB*, *dam*, *ninE*, *ninG* and *ninH*-like genes, respectively.

5.5 Conclusions

- Complete genome sequence of $\Phi 24_B$ revealed the presence of a gene homologous to the P22 antirepressor gene, *ant*.
- Reverse-transcriptase PCR analysis showed that the gene was being expressed in both a lysogen culture and in naïve cells undergoing $\Phi 24_B$ infection.
- Production of the Ant protein may go some way to explain the relatively high rate of spontaneous induction observed in $\Phi 24_B$, compared to bacteriophage λ, as well as its ability to produce multiple lysogens.
- Abrogation of the *ant* gene has been undertaken in order to assess any impact it may have on the behaviour of $\Phi 24_B$.

Chapter 6: Final Discussion

Bacteriophages play an important role in the evolution of both Gram negative and Gram positive bacterial pathogens (Boyd & Brussow, 2002) via the horizontal transfer of virulence determinants (Blaisdell *et al.*, 1996). Within bacteriophage populations, the modular theory of evolution describes the exchange of regions of the genome, usually of comparable or similar function, between integrated phages, whether they are inducible or remnant (Botstein, 1980). This process of modular substitution provides phages with substantial genomic diversity and in turn can lead to the emergence of novel pathogens (Saunders *et al.*, 2001; Sherratt *et al.*, 2004). These recombination events between phages will occur while they are harboured within a bacterial host cell or when bacteriophages within a bacterial host are exposed to new phages during a fresh infection event (Brussow *et al.*, 2004). The presence of multiple bacteriophage genomes within a host is limited by two major mechanisms: prophage loss or decay; immunity to super-infection by a homo-immune phage. The latter is the focus of this thesis. The ability of a temperate Stx-encoding bacteriophage ($\Phi 24_B$) to form double lysogens has been described (Allison *et al.*, 2003), and this provides an intracellular environment for recombination to further enhance the continued expansion of diversity amongst Stx-phages (Allison, 2007). Here, we begin to characterise this phenomenon by describing four distinct integration sites utilised by $\Phi 24_B$ in an *E. coli* K-12 genome, and extend the observation of double lysogeny to achieve the production of a triple lysogen. Multiple insertion sites for bacteriophages in bacterial genomes have previously been described only when the primary site is either absent or occupied *i.e.* they are not utilised concurrently (Barreiro & Haggard-Ljungquist, 1992; Rutkai *et al.*, 2003).

Though $\Phi 24_B$ does possess an integrase gene with a similar active site to the *intS* gene of *E. coli* strain MG1655, and $\Phi 24_B$ does integrate near the integration site utilised by *intS*, the integration site for $\Phi 24_B$ is distinct from the integration site recognised by the *E. coli intS* gene product. Creation of the *intS* knockout in the genome of *E. coli* strain MC1016 did reduce the efficiency of lysogen formation, but double lysogens could still be formed. Furthermore, complementation of the *intS* gene did not restore the original phenotype, so it must be concluded that additional polar effects of the *intS* knockout are responsible for the measured drop in lysogen formation efficiency. The genes located immediately downstream of *intS* are *yfdG, yfdH* and *yfdI*, and they are predicted to be membrane proteins that play a role in outer membrane biogenesis (Riley *et al.*, 2006). For example, the expression of these genes may have been affected in the *int*S knockout mutant, with concomitant effects on cell fitness and susceptibility to Stx-phage infection. The phage $\Phi 24_B$ does appear to prefer a single site (I) for integration near the *E. coli* chromosomal copy of *intS*, and site I possesses 24 bp of 100% sequence identity to the *attP* site and without fail is occupied in all of the lysogen DNA preparations examined by Southern hybridisation. The supplementary sites do share a degree of conserved consensus sequence within the 24 bp overlap region of the primary site, and in all cases this consensus sequence has overall identity in excess of 50%. Although a further three potential integration sites were identified in the Colibri database (PIA, E and F), PIA and PIF interrupt the expression of two genes, *insE5* and *yfiF*, respectively, predicted to encode essential function for *E. coli* (Gerdes *et al.*, 2003). A defined, minimal consensus sequence cannot yet be reported as production of at least 5 knockout mutants (I, II, III, IV and PIE) would be required before site-directed mutation experiments could be attempted in order to determine the minimal sequences necessary for integration. The DNA

sequence of $\Phi 24_B$ *int* identified here shares 100% identity within the conserved boxes (Fig. 17) associated with catalytic activity at the protein level and 96% nucleotide identity overall with the *int* gene of the recently submitted Stx2 phage-86 (accession number: NC 008464) isolated from an enterohaemorrhagic *Escherichia coli* serotype O86:H- strain, which also possesses a 24 bp *attP* overlap region identical to the one carried by $\Phi 24_B$. It is also interesting to note that the only homology these two phages appear to share is the integrase gene and its cognate *att* site. The sequence differences between the two integrase genes are external to the conserved box regions vital for catalytic activity. The location of phage-86 integration within the O86:H- strain's genome has not been described, however given the integrase and *attP* overlap similarities with $\Phi 24_B$, it can be suggested that both phages utilise the same site.

The precise rationale underlying the non-functional $\Phi 24_B$ super-infection immunity system is as yet unclear. Alignment of the immunity regions of $\Phi 24_B$ (Accession number: EF517298) and the model Stx-phage, 933W (Plunkett *et al.*, 1999), revealed 100% identity in sequence conservation between both phages. Nevertheless, in contrast to bacteriophage lambda, in both sequences the O_{R3} DNA-binding site is located within the *cI* ORF, possibly influencing the availability of the binding site to the repressor or Cro proteins *in situ*, and the O_{L3} DNA-binding site is completely absent (Koudelka *et al.*, 2004; Tyler *et al.*, 2004). Consequently, the combination of these factors may result in reduced efficiency of repression, however, the production of double 933W lysogens has not been reported. Alternatively, sequence data from $\Phi 24_B$ reveals the presence of a gene homologous to the *ant* gene of bacteriophage Lahn1 (Allison, 2007). Another homologue of this gene forms part of the *immI* antirepressor operon in P22 and, when expressed, is responsible for the deactivation

of the repressor protein along with the ensuing derepression of phage gene transcription (Botstein *et al.*, 1975; Susskind & Botstein, 1975; Susskind & Botstein, 1978). The P22 *ant* gene product has also been shown to be effective against lambda repressor cI (Susskind & Botstein, 1975). The presence of the *ant* gene product could go some way to explaining the absence of immunity or complete immunity to a superinfecting, homoimmune phage. In P22, *ant* expression is tightly regulated by two different proteins (Mnt and Arc) as well as a complementary transcript, *sar* (Susskind & Botstein, 1975). The sequence surrounding the $\Phi24_B$ *ant* gene does not encode comparable control factors. In fact, the $\Phi24_B$ *ant* gene is located 6.5 kb downstream of the *cro* gene and 1.2 kb upstream of the *Q* gene (data not shown). The demonstration here that the $\Phi24_B$ antirepressor gene is transcribed in a lysogenised host supports the hypothesis that the cI repressor protein is inactivated to an extent whereby an additional infection event is permitted, even if the infecting phage possesses an identical immunity region to an existent prophage. Although there is a $\Phi24_B$ *ant* homologue in the 933W genome, it does not appear to be intact, has very little sequence identity with $\Phi24_B$ *ant* and it is located just upstream of the lysis genes.

In addition to insertion site identification, Southern hybridisation analyses revealed the presence of a band that corresponds to the size of digested circular genomic bacteriophage DNA (Fig 19), as extrapolated from $\Phi24_B$ genome restriction analysis. The method of lysogen genomic DNA extraction excluded the isolation of extracellular free phage. Therefore, the extra-chromosomal phage DNA in a lysogen must either result from a surprising number of spontaneously induced, unreleased viral particles or, alternatively, from circularised plasmid-like pseudo-lysogens (Williamson *et al.*, 2001; Wommack & Colwell, 2000). Both circumstances are events

that might have a deleterious outcome during a Shiga toxin producing *E. coli* (STEC) infection. In the first instance, there would be a very high level of spontaneous induction (which has been reported for Stx-phages 933W and H-19B (Livny & Friedman, 2004)), and since *stx* gene expression is linked to the expression of the late genes, this will result in higher than expected levels of toxin. A pseudolysogenic state will amplify the number of *stx* genes within a bacterial cell that could be used as a template for Stx expression, a situation that would also result in higher than expected levels of Stx production. Increased Shiga toxin production would be expected to result in more severe STEC mediated disease and higher rates of complication with the sequelae such as haemolytic uraemic syndrome (HUS) (Siegler *et al.*, 2001).

Clearly, the ability of a single bacterium to harbour multiple phage, whether closely related or not, could have a profound impact on pathogen evolution, providing an increased gene pool for genetic exchange *in situ* (Abremski & Hoess, 1992; Allison, 2007) and possibly even support enhanced spontaneous mutation rates of duplicated/redundant phage genes that may ultimately result in the formation of new phage components with altered host range and other properties. Furthermore, the presence of several *stx* gene copies in a single host cell will in turn lead to amplification of toxin load and, upon lysis, manifestation of more severe disease pathology. The mechanisms underpinning the immunity patterns observed need to be further investigated, but evidence for an actively transcribed antirepressor is an important development. The potential of the antirepressor to influence immunity and the effect of divergence of the immunity region structure from the established lambda model are key factors that impact upon the evolution of Stx-phages and their role as drivers of the emergence of new Stx-producing bacterial pathogens.

References

Abremski, K. E. & Hoess, R. H. (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng* **5**, 87-91.

Acheson, D. W., Moore, R., De Breucker, S., Lincicome, L., Jacewicz, M., Skutelsky, E. & Keusch, G. T. (1996). Translocation of Shiga toxin across polarized intestinal cells in tissue culture. *Infect Immun* **64**, 3294-3300.

Allison, H. E., Sergeant, M. J., James, C. E., Saunders, J. R., Smith, D. L., Sharp, R. J., Marks, T. S. & McCarthy, A. J. (2003). Immunity Profiles of Wild-Type and Recombinant Shiga-Like Toxin-Encoding Bacteriophages and Characterization of Novel Double Lysogens. *Infect Immun* **71**, 3409-3418.

Allison, H. E. (2007). Stx-phages: Drivers and Mediators of the Evolution of STEC and STEC-like Pathogens. *Future Microbiol* **2**, 165-174.

Andreoli, S. P., Trachtman, H., Acheson, D. W. K., Siegler, R. L. & Obrig, T. G. (2002). Hemolytic uremic syndrome: epidemiology, pathophysiology, and therapy. *Pediatr Nephrol* **17**, 293-298.

Aoki, Y., Yosida, Y. & Takeda, T. (1999). Shiga toxin 2 promotes the stem cell differentiation into granulocytes in the bone marrow causing marked granulocytosis in the peripheral blood [1]. *J Infect* **39**, 97.

Argos, P., Landy, A., Abremski, K. & other authors (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* **5**, 433-440.

Ashkenazi, S., Cleary, K. R., Pickering, L. K., Murray, B. E. & Cleary, T. G. (1990). The association of Shiga toxin and other cytotoxins with the neurologic manifestations of shigellosis. *J Infect Dis* **161**, 961-965.

Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol Mol Biol Rev* **54**, 130-197.

Bagdasarian, M., Lurz, R., Ruckert, B., Franklin, F. C., Bagdasarian, M. M., Frey, J. & Timmis, K. N. (1981). Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. *Gene* **16**, 237-247.

Balding, C., Bromley, S. A., Pickup, R. W. & Saunders, J. R. (2005). Diversity of phage integrases in *Enterobacteriaceae*: development of markers for environmental analysis of temperate phages. *Environ Microbiol* **7**, 1558-1567.

Banatvala, N., Griffin, P. M., Greene, K. D., Barrett, T. J., Bibb, W. F., Green, J. H. & Wells, J. G. (2001). The United States National Prospective Hemolytic Uremic Syndrome Study: microbiologic, serologic, clinical, and epidemiologic findings. *J Infect Dis* **183**, 1063-1070.

Barkocy-Gallagher, G. A., Arthur, T. M., Rivera-Betancourt, M., Nou, X., Shackelford, S. D., Wheeler, T. L. & Koohmaraie, M. (2003). Seasonal prevalence of Shiga toxin-producing *Escherichia coli*, including O157:H7 and non-O157 serotypes, and *Salmonella* in commercial beef processing plants. *J Food Prot* **66**, 1978-1986.

Barreiro, V. & Haggard-Ljungquist, E. (1992a). Attachment sites for bacteriophage P2 on the *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and detection of a P2-like remnant in *E. coli* K-12 derivatives. *J Bacteriol* **174**, 4086-4093.

Bauer, M. E. & Welch, R. A. (1996). Characterization of an RTX toxin from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **64**, 167-175.

Bertani, G. & Six, E. (1958). Inheritance of prophage P2 in bacterial crosses. *Virology* **6**, 357-381.

Beutin, L. (2006). Emerging Enterohaemorrhagic *Escherichia coli*, Causes and Effects of the Rise of a Human Pathogen. *J Vet Med Series B* **53**, 299-305.

Biere, A. L., Citron, M. & Schuster, H. (1992). Transcriptional control via translational repression by c4 antisense RNA of bacteriophages P1 and P7. *Genes Dev* **6**, 2409-2416.

Bilge, S. S., Clausen, C. R., Lau, W. & Moseley, S. L. (1989). Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated *Escherichia coli* to HEp-2 cells. *J Bacteriol* **171**, 4281-4289.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. & Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* **435**, 1059-1066.

Blaisdell, B. E., Campbell, A. M. & Karlin, S. (1996). Similarities and dissimilarities of phage genomes. *PNAS* **93**, 5854-5859.

Blattner, F. R., Plunkett, G., III, Bloch, C. A. & other authors (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462.

Boerlin, P., McEwen, S. A., Boerlin-Petzold, F., Wilson, J. B., Johnson, R. P. & Gyles, C. L. (1999). Associations between Virulence Factors of Shiga Toxin-Producing *Escherichia coli* and Disease in Humans. *J Clin Microbiol* **37**, 497-503.

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

Botstein, K., Lew, K. K., Jarvik, V. & Swanson, C. A. (1975). Role of antirepressor in the bipartite control of repression and immunity by bacteriophage P22. *J Mol Biol* **91**, 439-462.

Boyd, B. & Lingwood, C. (1989). Verotoxin receptor glycolipid in human renal tissue. *Nephron* **51**, 207-210.

Boyd, B., Tyrrell, G., Maloney, M., Gyles, C., Brunton, J. & Lingwood, C. (1993). Alteration of the glycolipid binding specificity of the pig oedema toxin from globotetraosyl to globotriaosyl ceramide alters *in vivo* tissue targeting and results in a verotoxin 1-like disease in pigs. *J Exp Med* **177**, 1745-1753.

Boyd, E. F. & Brussow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* **10**, 521-529.

Brabban, A. D., Hite, E. & Callaway, T. R. (2005). Evolution of foodborne pathogens via temperate bacteriophage-mediated gene transfer. *Foodborne Pathog Dis* **2**, 287-303.

Brunder, W., Schmidt, H. & Karch, H. (1996). KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* **142 (Pt 11)**, 3305-3315.

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

Brunder, W., Karch, H. & Schmidt, H. (2006). Complete sequence of the large virulence plasmid pSFO157 of the sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H- strain 3072/96. *Int J Med Microbiol* **296**, 467-474.

Brussow, H., Canchaya, C. & Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiol Mol Biol Rev* **68**, 560-602.

Buchrieser, C. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* **38**, 760-771.

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

Calef, E. (1967). Mapping of integration and excision crossovers in superinfection double lysogens for phage lambda in *Escherichia coli*. *Genetics* **55**, 547-556.

Caliezi, C., Wuillemin, W. A., Zeerleder, S., Redondo, M., Eisele, B. & Hack, C. E. (2000). C1-Esterase Inhibitor: An Anti-Inflammatory Agent and Its Potential Use in the Treatment of Diseases Other Than Hereditary Angioedema. *Pharmacol Rev* **52**, 91-112.

Campbell, A. (1994). Comparative Molecular Biology of Lambdoid Phages. *Annu Rev Microbiol* **48**, 193-222.

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

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

Campellone, K. G. & Leong, J. M. (2003). Tails of two Tirs: actin pedestal formation by enteropathogenic *E. coli* and enterohemorrhagic *E. coli* O157:H7. *Curr Opin Microbiol* **6**, 82-90.

Casjens, S., Winn-Stapley, D. A., Gilcrease, E. B., Morona, R., Kuhlewein, C., Chua, J. E. H., Manning, P. A., Inwood, W. & Clark, A. J. (2004). The Chromosome of *Shigella flexneri* Bacteriophage Sf6: Complete Nucleotide Sequence, Genetic Mosaicism, and DNA Packaging. *J Mol Biol* **339**, 379-394.

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

Cerqueira, A. M. F., Guth, B. E. C., Joaquim, R. M. & Andrade, J. R. C. (1999). High occurrence of Shiga toxin-producing *Escherichia coli* (STEC) in healthy cattle in Rio de Janeiro State, Brazil. *Vet Microbiol* **70**, 111-121.

Chart, H., Jenkins, C., Smith, H. R., Hedges, D. & Rowe, B. (1998). Haemolysin production by strains of Verocytotoxin-producing *Escherichia coli*. *Microbiology* **144 (Pt 1)**, 103-107.

Chen, Y. & Rice, P. A. (2003). New insight into site-specific recombination from Flp recombinase-DNA structures. *Annu Rev Biophys Biomol Struct* **32**, 135-159.

Clark, A. J. (1973). Recombination deficient mutants of *E. coli* and other bacteria. *Annu Rev Genet* **7**, 67-86.

Clark, A. J., Inwood, W., Cloutier, T. & Dhillon, T. S. (2001). Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J Mol Biol* **311**, 657-679.

Cobbold, R. & Desmarchelier, P. (2000). A longitudinal study of Shiga-toxigenic *Escherichia coli* (STEC) prevalence in three Australian diary herds. *Vet Microbiol* **71**, 125-137.

Cohen, M. B. & Giannella, R. A. (1992). Hemorrhagic colitis associated with *Escherichia coli* O157:H7. *Adv Intern Med* **37**, 173-195.

Cordovez, A., Prado, V., Maggi, L. & other authors (1992). Enterohemorrhagic *Escherichia coli* associated with hemolytic-uremic syndrome in Chilean children. *J Clin Microbiol* **30**, 2153-2157.

Court, D. L., Oppenheim, A. B. & Adhya, S. L. (2007). A New Look at Bacteriophage lambda Genetic Networks. *J Bacteriol* **189**, 298-304.

Craig, N. L. & Nash, H. A. (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell* **39**, 707-716.

Dewell, G. A., Ransom, J. R., Dewell, R. D. & other authors (2005). Prevalence of and risk factors for *Escherichia coli* O157 in market-ready beef cattle from 12 U.S. feedlots. *Foodborne Pathog Dis* **2**, 70-76.

Dobrindt, U. (2005). (Patho-)Genomics of *Escherichia coli*. *Int J Med Microbiol* **295**, 357-371.

Dodd, I. B., Shearwin, K. E., Perkins, A. J., Burr, T., Hochschild, A. & Egan, J. B. (2004). Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev* **18**, 344-354.

Dodd, I. B., Shearwin, K. E. & Egan, J. B. (2005). Revisited gene regulation in bacteriophage lambda. *Curr Opin Genet Dev*, 145-152.

Dostal, L., Misselwitz, R. & Welfle, H. (2005). Arc Repressor-Operator DNA Interactions and Contribution of Phe10 to Binding Specificity. *Biochemistry* **44**, 8387-8396.

Echols, H. (1972). Developmental Pathways for the Temperate Phage: Lysis Vs Lysogeny. *Annu Rev Genet*, 157-190.

Eklund, M., Nuorti, J. P., Ruutu, P. & Siitonen, A. (2005). Shigatoxigenic *Escherichia coli* (STEC) infections in Finland during 1998-2002: a population-based surveillance study. *Epidemiol Infect* **133**, 845-852.

Elliott, E. J., Robins-Browne, R. M., O'Loughlin, E. V. & other authors (2001). Nationwide study of haemolytic uraemic syndrome: clinical, microbiological, and epidemiological features. *Arch Dis Child* **85**, 125-131.

Esposito, D. & Scocca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucl Acids Res* **25**, 3605-3614.

Faguy, D. M. & Doolittle, W. F. (2000). Horizontal transfer of catalase-peroxidase genes between Archaea and pathogenic bacteria. *Trends Genet* **16**, 196-197.

Faruque, S. M., Kamruzzaman, M., Asadulghani, Sack, D. A., Mekalanos, J. J. & Nair, G. B. (2003a). CTXPhi -independent production of the RS1 satellite phage by *Vibriocholerae*. *Proc Natl Acad Sci U S A* **100**, 1280-1285.

Faruque, S. M., Sack, D. A., Sack, R. B., Colwell, R. R., Takeda, Y. & Nair, G. B. (2003b). Inaugural Article: Emergence and evolution of *Vibrio cholerae* O139. *Proc Natl Acad Sci U S A* **100**, 1304-1309.

Fattah, K. R., Mizutani, S., Fattah, F. J., Matsushiro, A. & Sugino, Y. (2000). A comparative study of the immunity region of lambdoid phages including Shiga-toxin-converting phages: molecular basis for cross immunity. *Genes Genet Syst* **75**, 223-232.

Fischer, H., Konig, P., Dierich, M. P. & Allerberger, F. (2001). Hemolytic-uremic syndrome surveillance to monitor trends in infection with *Escherichia coli* O157 and non-O157 enterohemorrhagic *E. coli* in Austria. *Pediatr Infect Dis J* **20**, 316-318.

Fleischmann, R. D., Adams, M. D., White, O. & other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Freeman, V. J. (1951). Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol* **61**, 675-688.

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

Garcia-Aljaro, C., Muniesa, M., Jofre, J. & Blanch, A. R. (2006). Newly identified bacteriophages carrying the *stx2g* Shiga toxin gene isolated from *Escherichia coli* strains in polluted waters. *FEMS Microbiol Lett* **258**, 127-135.

Garcia, E., Nedialkov, Y. A., Elliott, J., Motin, V. L. & Brubaker, R. R. (1999). Molecular Characterization of KatY (Antigen 5), a Thermoregulated Chromosomally Encoded Catalase-Peroxidase of *Yersinia pestis*. *J Bacteriol* **181**, 3114-3122.

Garred, O., Deurs, B. V. & Sandvig, K. (1995a). Furin-induced Cleavage and Activation of Shiga Toxin. *J Biol Chem* **270**, 10817-10821.

Garred, O., Dubinina, E., Holm, P. K., Olsnes, S., Van Deurs, B., Kozlov, J. V. & Sandvig, K. (1995b). Role of Processing and Intracellular Transport for Optimal Toxicity of Shiga Toxin and Toxin Mutants. *Exp Cell Res* **218**, 39-49.

Gerdes, S. Y., Scholle, M. D., Campbell, J. W. & other authors (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185**, 5673-5684.

Gopaul, D. N., Guo, F. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J* **17**, 4175-4187.

Grabow, W. O., and P. Coubrough. (1986). Practical direct plaque assay for coliphages in 100-302 ml samples of drinking water. *Appl. Environ. Microbiol.* **52**, 430–433.

Greenblatt, J., Mah, T. F., Legault, P., Mogridge, J., Li, J. & Kay, L. E. (1998). Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. *Cold Spring Harb Symp Quant Biol* **63**, 327-336.

Griffin, P. M., Olmstead, L. C. & Petras, R. E. (1990). *Escherichia coli* O157:H7-associated colitis. A clinical and histological study of 11 cases. *Gastroenterology* **99**, 142-149.

Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**, 567-605.

Groth, A. C. & Calos, M. P. (2004). Phage Integrases: Biology and Applications. *J Mol Biol* **335**, 667-678.

Grys, T. E., Siegel, M. B., Lathem, W. W. & Welch, R. A. (2005). The StcE Protease Contributes to Intimate Adherence of Enterohemorrhagic *Escherichia coli* O157:H7 to Host Cells. *Infect Immun* **73**, 1295-1303.

Gyles, C. L. (2007). Shiga toxin-producing *Escherichia coli*: An overview. *J Anim Sci* **85**, 45-62.

Harrison, L. M., Cherla, R. P., van den Hoogen, C., van Haaften, W. C. E., Lee, S.-Y. & Tesh, V. L. (2005). Comparative evaluation of apoptosis induced by Shiga toxin 1 and/or lipopolysaccharides in human monocytic and macrophage-like cells. *Microb Pathog* **38**, 63-76.

Hayashi, T., Makino, K., Ohnishi, M. & other authors (2001). Complete Genome Sequence of Enterohemorrhagic *Eschelichia coli* O157:H7 and Genomic Comparison with a Laboratory Strain K-12. *DNA Res* **8**, 11-22.

Heinrich, J., Citron, M., Gunther, A. & Schuster, H. (1994). Second-site suppressors of the bacteriophage P1 virs mutant reveal the interdependence of the *c4*, *icd*, and *ant* genes in the P1 *imml* operon. *J Bacteriol* **176**, 4931-4936.

Heinrich, J., Velleman, M. & Schuster, H. (1995). The tripartite immunity system of phages P1 and P7. *FEMS Microbiol Rev* **17**, 121-126.

Heir, E., Lindstedt, B. A., Vardund, T., Wasteson, Y. & Kapperud, G. (2000). Genomic fingerprinting of shigatoxin-producing *Escherichia coli* (STEC) strains: comparison of pulsed-field gel electrophoresis (PFGE) and fluorescent amplified-fragment-length polymorphism (FAFLP). *Epidemiol Infect* **125**, 537-548.

Hicks, S., Frankel, G., Kaper, J. B., Dougan, G. & Phillips, A. D. (1998). Role of Intimin and Bundle-Forming Pili in Enteropathogenic *Escherichia coli* Adhesion to Pediatric Intestinal Tissue *In Vitro*. *Infect Immun* **66**, 1570-1578.

Jacob, F. & Wollman, E. L. (1961). Viruses and genes. *Sci Am* **204**, 93-107.

James, C. E., Stanley, K. N., Allison, H. E., Flint, H. J., Stewart, C. S., Sharp, R. J., Saunders, J. R. & McCarthy, A. J. (2001). Lytic and Lysogenic Infection of Diverse *Escherichia coli* and *Shigella* Strains with a Verocytotoxigenic Bacteriophage. *Appl Environ Microbiol* **67**, 4335-4337.

Johansen, B. K., Wasteson, Y., Granum, P. E. & Brynestad, S. (2001). Mosaic structure of Shiga-toxin-2-encoding phages isolated from *Escherichia coli* O157:H7 indicates frequent gene exchange between lambdoid phage genomes. *Microbiology* **147**, 1929-1936.

Kaniga, K., Delor, I. & Cornelis, G. R. (1991). A wide-host-range suicide vector for improving reverse genetics in gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*. *Gene* **109**, 137-141.

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

Keusch, G. T., Grady, G. F., Mata, L. J. & McIver, J. (1972). The pathogenesis of *Shigella* diarrhea. I. Enterotoxin production by *Shigella dysenteriae* I. *J Clin Invest* **51**, 1212-1218.

Khan, W. A., Dhar, U., Salam, M. A., Griffiths, J. K., Rand, W. & Bennish, M. L. (1999). Central nervous system manifestations of childhood shigellosis: prevalence, risk factors, and outcome. *Pediatrics* **103**, E18.

King, A. J., Sundaram, S., Cendoroglo, M., Acheson, D. W. K. & Keusch, G. T. (1999). Shiga toxin induces superoxide production in polymorphonuclear cells with subsequent impairment of phagocytosis and responsiveness to phorbol esters. *J Ifect Dis* **179**, 503-507.

Kistemann, T., Zimmer, S., Vagsholm, I. & Andersson, Y. (2004). GIS-supported investigation of human EHEC and cattle VTEC O157 infections in Sweden: geographical distribution, spatial variation and possible risk factors. *Epidemiol Infect* **132**, 495-505.

Klapproth, J.-M. A., Scaletsky, I. C. A., McNamara, B. P., Lai, L.-C., Malstrom, C., James, S. P. & Donnenberg, M. S. (2000). A Large Toxin from Pathogenic *Escherichia coli* Strains That Inhibits Lymphocyte Activation. *Infect Immun* **68**, 2148-2155.

Klotz, M. G. & Loewen, P. C. (2003). The Molecular Evolution of Catalatic Hydroperoxidases: Evidence for Multiple Lateral Transfer of Genes Between Prokaryota and from Bacteria into Eukaryota. *Mol Biol Evol* **20**, 1098-1112.

Knight, K. L. & Sauer, R. T. (1992). Biochemical and genetic analysis of operator contacts made by residues within the beta-sheet DNA binding motif of Mnt repressor. *Embo J* **11**, 215-223.

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

Kobiler, O., Rokney, A. & Oppenheim, A. B. (2007). Phage Lambda CIII: A Protease Inhibitor Regulating the Lysis-Lysogeny Decision. *PLoS ONE* **2**, e363.

Koudelka, A. P., Hufnagel, L. A. & Koudelka, G. B. (2004). Purification and Characterization of the Repressor of the Shiga Toxin-Encoding Bacteriophage 933W: DNA Binding, Gene Regulation, and Autocleavage. *J Bacteriol* **186**, 7659-7669.

Krogh, B. O. & Shuman, S. (2000). Catalytic mechanism of DNA topoisomerase IB. *Mol Cell* **5**, 1035-1041.

Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **58**, 913-949.

Lathem, W. W., Grys, T. E., Witowski, S. E., Torres, A. G., Kaper, J. B., Tarr, P. I. & Welch, R. A. (2002). StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. *Mol Microbiol* **45**, 277-288.

Lathem, W. W., Bergsbaken, T. & Welch, R. A. (2004). Potentiation of C1 Esterase Inhibitor by StcE, a Metalloprotease Secreted by *Escherichia coli* O157:H7. *J Exp Med* **199**, 1077-1087.

Law, D. (2000). Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*. *J Appl Microbiol* **88**, 729-745.

Lawrence, J. G. & Ochman, H. (1998). Molecular archaeology of the *Escherichia coli* genome. *PNAS* **95**, 9413-9417.

Levine, M., Truesdell, S., Ramakrishnan, T. & Bronson, M. J. (1975). Dual control of lysogeny by bacteriophage P22: An antirepressor locus and its controlling elements. *J Mol Biol* **91**, 421-438.

Liao, S. M., Wu, T. H., Chiang, C. H., Susskind, M. M. & McClure, W. R. (1987). Control of gene expression in bacteriophage P22 by a small antisense RNA. I. Characterization *in vitro* of the Psar promoter and the sar RNA transcript. *Genes Dev* **1**, 197-203.

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

Little, J. W., Edmiston, S. H., Pacelli, L. Z. & Mount, D. W. (1980). Cleavage of the *Escherichia coli* lexA protein by the recA protease. *Proc Natl Acad Sci U S A* **77**, 3225-3229.

Little, J. W. (1991). Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease. *Biochimie* **73**, 411-421.

Little, J. W. (2005). Threshold effects in gene regulation: When some is not enough. *Proc Natl Acad Sci USA* **102**, 5310-5311.

Liu, J., Akahosh, T., Sasahana, T., Kitasato, H., Namai, R., Sasaki, T., Inoue, M. & Kondo, H. (1999). Inhibition of neutrophil apoptosis by verotoxin 2 derived from *Escherichia coli* O157:H7. *Infect Immun* **67**, 6203-6205.

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

Lommel, S., Benesch, S., Rohde, M., Wehland, J. & Rottner, K. (2004). Enterohaemorrhagic and enteropathogenic *Escherichia coli* use different mechanisms for actin pedestal formation that converge on N-WASP. *Cell Microbiol* **6**, 243-254.

Low, J. C., McKendrick, I. J., McKechnie, C., Fenlon, D., Naylor, S. W., Currie, C., Smith, D. G. E., Allison, L. & Gally, D. L. (2005). Rectal Carriage of Enterohemorrhagic *Escherichia coli* O157 in Slaughtered Cattle. *Appl Environ Microbiol* **71**, 93-97.

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

Mahon, B. E., Griffin, P. M., Mead, P. S. & Tauxe, R. V. (1997). Hemolytic uremic syndrome surveillance to monitor trends in infection with *Escherichia coli* O157:H7 and other shiga toxin-producing *E. coli*. *Emerg Infect Dis* **3**, 409-412.

Mainil, J. (1999). Shiga/verocytotoxins and Shiga/verotoxigenic *Escherichia coli* in animals. *Vet Res* **30**, 235-257.

Makino, K., Ishii, K., Yasunaga, T. & other authors (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., Yokoyama, K., Kubota, Y. & other authors (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.

Marches, O., Ledger, T. N., Boury, M. & other authors (2003). Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G2/M transition. *Mol Microbiol* **50**, 1553-1567.

Marr, M. T., Datwyler, S. A., Meares, C. F. & Roberts, J. W. (2001). Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. *Proc Natl Acad Sci USA* **98**, 8972-8978.

Matthews, L., McKendrick, I. J., Ternent, H., Gunn, G. J., Synge, B. & Woolhouse, M. E. (2006). Super-shedding cattle and the transmission dynamics of *Escherichia coli* O157. *Epidemiol Infect* **134**, 131-142.

Maurer, R., Meyer, B. J. & Ptashne, M. (1980). Gene regulation at the right operator (OR) of bacteriophage lambda: I. OR3 and autogenous negative control by repressor. *J Mol Biol* **139**, 147-161.

McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. & Kaper, J. B. (1995). A Genetic Locus of Enterocyte Effacement Conserved Among Diverse Enterobacterial Pathogens. *PNAS* **92**, 1664-1668.

McDaniel, T. K. & Kaper, J. B. (1997). A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol Microbiol* **23**, 399-407.

McDonough, M. A. & Butterton, J. R. (1999). Spontaneous tandem amplification and deletion of the Shiga toxin operon in *Shigella dysenteriae* 1. *Mol Microbiol* **34**, 1058-1069.

Michel, L. O., Sandkvist, M. & Bagdasarian, M. (1995). Specificity of the protein secretory apparatus: secretion of the heat-labile enterotoxin B subunit pentamers by different species of Gram- bacteria. *Gene* **152**, 41-45.

Michel, P., Wilson, J. B., Martin, S. W., Clarke, R. C., McEwen, S. A. & Gyles, C. L. (1999). Temporal and geographical distributions of reported cases of *Escherichia coli* O157:H7 infection in Ontario. *Epidemiol Infect* **122**, 193-200.

Mizutani, S., Nakazono, N. & Sugino, Y. (1999). The So-called Chromosomal Verotoxin Genes are Actually Carried by Defective Prophages. *DNA Res* **6**, 141-143.

Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., Hsu, P. L., Ross, W. & Landy, A. (1981). Structure and function of the phage lambda att site: size, int-binding sites, and location of the crossover point. *Cold Spring Harb Symp Quant Biol* **45 Pt 1**, 429-437.

Nagarajan, R., Kwon, K., Nawrot, B., Stec, W. J. & Stivers, J. T. (2005). Catalytic Phosphoryl Interactions of Topoisomerase IB. *Biochemistry* **44**, 11476-11485.

Nataro, J. P. & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* **11**, 142-201.

Nataro, J. P., Steiner, T. S. & Guerrant, R. L. (1998). Enteroaggregative *Escherichia coli*. *Emerg Infect Dis* **4**, 251-261.

Naylor, S. W., Low, J. C., Besser, T. E., Mahajan, A., Gunn, G. J., Pearce, M. C., McKendrick, I. J., Smith, D. G. E. & Gally, D. L. (2003). Lymphoid Follicle-Dense Mucosa at the Terminal Rectum Is the Principal Site of Colonization of Enterohemorrhagic *Escherichia coli* O157:H7 in the Bovine Host. *Infect Immun* **71**, 1505-1512.

Nicholls, L., Grant, T. H. & Robins-Browne, R. M. (2000). Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic *Escherichia coli* to epithelial cells. *Mol Microbiol* **35**, 275-288.

Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucl Acids Res* **26**, 391-406.

O'Loughlin, E. V. & Robins-Browne, R. M. (2001). Effect of Shiga toxin and Shiga-like toxins on eukaryotic cells. *Microbes Infect* **3**, 493-507.

Olsnes, S., Reisbig, R. & Eiklid, K. (1981). Subunit structure of *Shigella* cytotoxin. *J Biol Chem* **256**, 8732-8738.

Omisakin, F., MacRae, M., Ogden, I. D. & Strachan, N. J. C. (2003). Concentration and Prevalence of *Escherichia coli* O157 in Cattle Feces at Slaughter. *Appl Environ Microbiol* **69**, 2444-2447.

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. & Adhya, S. (2005). Switches in bacteriophage lambda development. *Annu Rev Genet* **39**, 409-429.

Paton, J. C. & Paton, A. W. (1998). Pathogenesis and Diagnosis of Shiga Toxin-Producing *Escherichia coli* Infections. *Clin Microbiol Rev* **11**, 450-479.

Peiffer, I., Servin, A. L. & Bernet-Camard, M. F. (1998). Piracy of decay-accelerating factor (CD55) signal transduction by the diffusely adhering strain *Escherichia coli* C1845 promotes cytoskeletal F-actin rearrangements in cultured human intestinal INT407 cells. *Infect Immun* **66**, 4036-4042.

Peiffer, I., Bernet-Camard, M. F., Rousset, M. & Servin, A. L. (2001). Impairments in enzyme activity and biosynthesis of brush border-associated hydrolases in human intestinal Caco-2/TC7 cells infected by members of the Afa/Dr family of diffusely adhering *Escherichia coli*. *Cell Microbiol* **3**, 341-357.

Penfold, R. J. & Pemberton, J. M. (1992). An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene* **118**, 145-146.

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

Perna, N. T., Plunkett, G., Burland, V. & other authors (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529-533.

Pitari, G. M. (2003). Bacterial enterotoxins are associated with resistance to colon cancer. *Proc Natl Acad Sci USA* **100**, 2695-2699.

Plunkett, G., III, Rose, D. J., Durfee, T. J. & Blattner, F. R. (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.

Prell, H. H. & Harvey, A. M. (1983). P22 antirepressor protein prevents in vivo recA-dependent proteolysis of P22 repressor. *Mol Gen Genet* **190**, 427-431.

Pringle, C. R. (1999). Virus Taxonomy - 1999. *Arch Virol* **144**, 421.

Ptashne, M. (2004). *A genetic switch: phage lambda revisited*, 3rd edn. Cold Spring Harbour: Cold Spring Harbour Laboratory Press.

Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. (1994). DNA recognition by beta-sheets in the Arc represser-operator crystal structure. *Nature* **367**, 754-757.

Raumann, B. E., Knight, K. L. & Sauer, R. T. (1995). Dramatic Changes In Dna-Binding Specificity Caused By Single Residue Substitutions In An Arc/Mnt Hybrid Repressor. *Nat Struct Biol* **2**, 1115-1122.

Reisbig, R., Olsnes, S. & Eiklid, K. (1981). The cytotoxic activity of *Shigella* toxin. Evidence for catalytic inactivation of the 60 S ribosomal subunit. *J Biol Chem* **256**, 8739-8744.

Richardson, S. E., Rotman, T. A., Jay, V., Smith, C. R., Becker, L. E., Petric, M., Olivieri, N. F. & Karmali, M. A. (1992). Experimental verocytotoxemia in rabbits. *Infect Immun* **60**, 4154-4167.

Riedel, H.-D., Heinrich, J., Heisig, A., Choli, T. & Schuster, H. (1993a). The antirepressor of phage P1 Isolation and interaction with the C1 repressor of P1 and P7. *FEBS Lett* **334**, 165-169.

Riedel, H. D., Heinrich, J. & Schuster, H. (1993b). Cloning, expression, and characterization of the *icd* gene in the *imm1* operon of bacteriophage P1. *J Bacteriol* **175**, 2833-2838.

Riley, L. W., Remis, R. S., Helgerson, S. D. & other authors (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* **308**, 681-685.

Riley, M., Abe, T., Arnaud, M. B. & other authors (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucl Acids Res* **34**, 1-9.

Roberts, J. W. & Roberts, C. W. (1975). Proteolytic cleavage of bacteriophage lambda repressor in induction. *Proc Natl Acad Sci U S A* **72**, 147-151.

Robinson, C. M., Sinclair, J. F., Smith, M. J. & O'Brien, A. D. (2006). Shiga toxin of enterohemorrhagic *Escherichia coli* type O157:H7 promotes intestinal colonization. *Proc Natl Acad Sci U S A* **103**, 9667-9672.

Rohatgi, R., Nollau, P., Ho, H.-Y. H., Kirschner, M. W. & Mayer, B. J. (2001). Nck and Phosphatidylinositol 4,5-Bisphosphate Synergistically Activate Actin Polymerization through the N-WASP-Arp2/3 Pathway. *J Biol Chem* **276**, 26448-26452.

Roldgaard, B. B., Scheutz, F., Boel, J., Aabo, S., Schultz, A. C., Cheasty, T., Nielsen, E. M., Olsen, K. E. & Christensen, B. B. (2004). VTEC O157 subtypes associated with the most severe clinical symptoms in humans constitute a minor part of VTEC O157 isolates from Danish cattle. *Int J Med Microbiol* **294**, 255-259.

Ross, W., Landy, A., Kikuchi, Y. & Nash, H. (1979). Interaction of Int protein with specific sites on lambda *att* DNA. *Cell* **18**, 297-307.

Rutkai, E., Dorgai, L., Sirot, R., Yagil, E. & Weisberg, R. A. (2003). Analysis of Insertion into Secondary Attachment Sites by Phage lambda and by *int* Mutants with Altered Recombination Specificity. *J Mol Biol* **329**, 983-996.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Sandhu, K. S., Clarke, R. C., McFadden, K., Brouwer, A., Louie, M., Wilson, J., Lior, H. & Gyles, C. L. (1996). Prevalence of the *eaeA* gene in verotoxigenic *Escherichia coli* strains from dairy cattle in Southwest Ontario. *Epidemiol Infect* **116**, 1-7.

Sandvig, K. (2001). Shiga toxins. *Toxicon* **39**, 1629-1635.

Sansonetti, P. (2002). Host-pathogen interactions: the seduction of molecular cross talk. *Gut* **50**, S2.

Sarkar, D., Radman-Livaja, M. & Landy, A. (2001). The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. *Embo J* **20**, 1203-1212.

Sauer, R. T., Pabo, C. O., Meyer, B. J., Ptashne, M. & Backman, K. C. (1979). Regulatory functions of the lambda repressor reside in the amino-terminal domain. *Nature* **279**, 396-400.

Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P. & Susskind, M. M. (1983). Primary structure of the *imm1* immunity region of bacteriophage P22. *J Mol Biol* **168**, 699-713.

Saunders, J. R., Allison, H. E., James, C. E., McCarthy, A. J. & Sharp, R. (2001). Phage-mediated transfer of virulence genes. *J Chem Technol Biotechnol* **76**, 662-666.

Schmidt, H. & Karch, H. (1996). Enterohemolytic phenotypes and genotypes of shiga toxin-producing *Escherichia coli* O111 strains from patients with diarrhea and hemolytic- uremic syndrome. *J Clin Microbiol* **34**, 2364-2367.

Schmidt, H., Kernbach, C. & Karch, H. (1996). Analysis of the EHEC *hly* operon and its location in the physical map of the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* **142 (Pt 4)**, 907-914.

Schmidt, H., Henkel, B. & Karch, H. (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.

Shapiro, J. A. (1979). Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc Natl Acad Sci U S A* **76**, 1933-1937.

Sherratt, D., SÃ¸balle, B., Barre, F., Filipe, S., Lau, I., Massey, T. & Yates, J. (2004). Recombination and chromosome segregation. *Philos Trans R Soc Lond B Biol Sci* **359**, 61-69.

Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1972). Prophage lambda at unusual chromosomal locations: I. Location of the secondary attachment sites and the properties of the lysogens. *J Mol Biol* **63**, 483-503.

Shotland, Y., Koby, S., Teff, D., Mansur, N. & Oren, D. A. (1997). Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of *Escherichia coli*. *Mol Microbiol* **24**, 1303-1310.

Siegler, R. L., Pysher, T. J., Tesh, V. L. & Taylor, F. B., Jr. (2001). Response to single and divided doses of Shiga toxin-1 in a primate model of hemolytic uremic syndrome. *J Am Soc Nephrol* **12**, 1458-1467.

Silbaq, F. S., Ruttenberg, S. E. & Stormo, G. D. (2002). Specificity of Mnt 'master residue' obtained from *in vivo* and *in vitro* selections. *Nucl Acids Res* **30**, 5539-5548.

Sinclair, J. F. & O'Brien, A. D. (2004). Intimin types alpha, beta, and gamma bind to nucleolin with equivalent affinity but lower avidity than to the translocated intimin receptor. *J Biol Chem* **279**, 33751-33758.

Sjogren, R., Neill, R., Rachmilewitz, D. & other authors (1994). Role of Shiga-like toxin I in bacterial enteritis: comparison between isogenic *Escherichia coli* strains induced in rabbits. *Gastroenterology* **106**, 306-317.

Smith, D. L., James, C. E., Sergeant, M. J., Yaxian, Y., Saunders, J. R., McCarthy, A. J. and Allison, H. E. (2007a). Short-tailed Stx-phages exploit the conserved YaeT protein to disseminate Shiga toxin genes amongst enterobacteria. *J Bacteriol* (Published online ahead of print on 10 August 2007)

Smith, D. L., Wareing, B. M., Fogg, P. C. M., Riley, L. M., Spencer, M., Cox, M. J., Saunders, J. R., McCarthy, A. J. and Allison, H. E. (2007b). A multi-loci characterization scheme for Shiga-toxin encoding bacteriophages. (Submitted)

Spangler, B. D. (1992). Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin. *Microbiol Mol Biol Rev* **56**, 622-647.

Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L. & Read, R. J. (1992). Crystal structure of the cell-binding B oligomer of verotoxin-1 from *E. coli*. *Nature* **355**, 748-750.

Steiner, T. S., Nataro, J. P., Poteet-Smith, C. E., Smith, J. A. & Guerrant, R. L. (2000). Enteroaggregative *Escherichia coli* expresses a novel flagellin that causes IL-8 release from intestinal epithelial cells. *J Clin Invest* **105**, 1769-1777.

Stevens, M. P., van Diemen, P. M., Frankel, G., Phillips, A. D. & Wallis, T. S. (2002). Efa1 Influences Colonization of the Bovine Intestine by Shiga Toxin-Producing *Escherichia coli* Serotypes O5 and O111. *Infect Immun* **70**, 5158-5166.

Stevens, M. P., Roe, A. J., Vlisidou, I., van Diemen, P. M., La Ragione, R. M., Best, A., Woodward, M. J., Gally, D. L. & Wallis, T. S. (2004). Mutation of *toxB* and a Truncated Version of the *efa-1* Gene in *Escherichia coli* O157:H7 Influences

the Expression and Secretion of Locus of Enterocyte Effacement-Encoded Proteins but not Intestinal Colonization in Calves or Sheep. *Infect Immun* **72**, 5402-5411.

Suh, J. K., Hovde, C. J. & Robertus, J. D. (1998). Shiga Toxin Attacks Bacterial Ribosomes as Effectively as Eukaryotic Ribosomes. *Biochemistry* **37**, 9394-9398.

Susskind, M. M. & Botstein, D. (1975). Mechanism of action of *Salmonella* phage P22 antirepressor. *J Mol Biol* **98**, 413-424.

Susskind, M. M. & Botstein, D. (1978). Molecular genetics of bacteriophage P22. *Microbiol Rev* **42**, 385-413.

Susskind, M. M. (1980). A new gene of bacteriophage P22 which regulates synthesis of antirepressor. *J Mol Biol* **138**, 685-713.

Takeda, Y., Folkmanis, A. & Echols, H. (1977). Cro regulatory protein specified by bacteriophage lambda. Structure, DNA-binding, and repression of RNA synthesis. *J Biol Chem* **252**, 6177-6183.

Tesh, V. L., Ramegowda, B. & Samuel, J. E. (1994). Purified Shiga-like toxins induce expression of proinflammatory cytokines from murine peritoneal macrophages. *Infect and Immun* **62**, 5085-5094.

Tran Van Nhieu, G., Bourdet-Sicard, R., Dumenil, G., Blocker, A. & Sansonetti, P. J. (2000). Bacterial signals and cell responses during *Shigella* entry into epithelial cells. *Cell Microbiol* **2**, 187-193.

Tyler, J. S., Mills, M. J. & Friedman, D. I. (2004). The Operator and Early Promoter Region of the Shiga Toxin Type 2-Encoding Bacteriophage 933W and Control of Toxin Expression. *J Bacteriol* **186**, 7670-7679.

Unkmeir, A. & Schmidt, H. (2000). Structural Analysis of Phage-Borne stx Genes and Their Flanking Sequences in Shiga Toxin-Producing *Escherichia coli* and *Shigella dysenteriae* Type 1 Strains. *Infect Immun* **68**, 4856-4864.

Van Duyne, G. D. (2001). A Structural View of Cre-loxP Site-Specific Recombination. *Annu Rev of Biophys Biomol Struct* **30**, 87-104.

Van Duyne, G. D. (2005). Lambda Integrase: Armed for Recombination. *Curr Biol* **15**, R658-R660.

van Setten, P. A., Monnens, L. A., Verstraten, R. G., van den Heuvel, L. P. & van Hinsbergh, V. W. (1996). Effects of verocytotoxin-1 on nonadherent human monocytes: binding characteristics, protein synthesis, and induction of cytokine release. *Blood* **88**, 174-183.

Varnado, C. L., Hertwig, K. M., Thomas, R., Roberts, J. K. & Goodwin, D. C. (2004). Properties of a novel periplasmic catalase-peroxidase from *Escherichia coli* O157:H7. *Arch Biochem Biophys* **421**, 166-174.

Velarde, J. J. & Nataro, J. P. (2004). Hydrophobic Residues of the Autotransporter EspP Linker Domain Are Important for Outer Membrane Translocation of Its Passenger. *J Biol Chem* **279**, 31495-31504.

Vershon, A. K., Youderian, P., Susskind, M. M. & Sauer, R. T. (1985). The bacteriophage P22 Arc and Mnt repressors. Overproduction, purification, and properties. *J Biol Chem* **260**, 12124-12129.

Vold, L., Sandberg, M., Jarp, J. & Wasteson, Y. (2001). Occurrence and characterization of *Escherichia coli* O157 isolated from cattle in Norway. *Vet Res Commun* **25**, 13-26.

Wagner, E. G. H. & Simons, R. W. (1994). Antisense RNA Control in Bacteria, Phages, and Plasmids. *Annu Rev Microbiol* **48**, 713-742.

Wagner, P. L., Livny, J., Neely, M. N., Acheson, D. W. K., Friedman, D. I. & Waldor, M. K. (2002). Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol Microbiol* **44**, 957-970.

Wang, H., Yang, C. H., Lee, G., Chang, F., Wilson, H., del Campillo-Campbell, A. & Campbell, A. (1997). Integration specificities of two lambdoid phages (21 and e14) that insert at the same *attB* site. *J Bacteriol* **179**, 5705-5711.

Weaver, A. M., Heuser, J. E., Karginov, A. V., Lee, W.-l., Parsons, J. T. & Cooper, J. A. (2002). Interaction of Cortactin and N-WASp with Arp2/3 Complex. *Curr Biol* **12**, 1270-1278.

Whiteson, K. L., Chen, Y., Chopra, N., Raymond, A. C. & Rice, P. A. (2007). Identification of a Potential General Acid/Base in the Reversible Phosphoryl Transfer Reactions Catalyzed by Tyrosine Recombinases: Flp H305. *Chem Biol* **14**, 121-129.

Widiasih, D. A., Ido, N., Omoe, K., Sugii, S. & Shinagawa, K. (2004). Duration and magnitude of faecal shedding of Shiga toxin-producing *Escherichia coli* from naturally infected cattle. *Epidemiol Infect* **132**, 67-75.

Williamson, S. J., McLaughlin, M. R. & Paul, J. H. (2001). Interaction of the PhiHSIC Virus with Its Host: Lysogeny or Pseudolysogeny? *Appl Environ Microbiol* **67**, 1682-1688.

Willshaw, G. A., Smith, H. R., Scotland, S. M., Field, A. M. & Rowe, B. (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.

Willshaw, G. A., Cheasty, T., Smith, H. R., O'Brien, S. J. & Adak, G. K. (2001). Verocytotoxin-producing *Escherichia coli* (VTEC) O157 and other VTEC from human infections in England and Wales: 1995-1998. *J Med Microbiol* **50**, 135-142.

Wolf, M. K. (1997). Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*. *Clin Microbiol Rev* **10**, 569-584.

Wommack, K. E. & Colwell, R. R. (2000). Virioplankton: Viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**, 69.

Wright, A. (1971). Mechanism of Conversion of the *Salmonella* O Antigen by Bacteriophage ε34. *J Bacteriol* **105**, 927-936.

Wu, T. H., Liao, S. M., McClure, W. R. & Susskind, M. M. (1987). Control of gene expression in bacteriophage P22 by a small antisense RNA. II. Characterization of mutants defective in repression. *Genes Dev* **1**, 204-212.

Yin, S., Bushman, W. & Landy, A. (1985). Interaction of the lambda Site-Specific Recombination Protein Xis with Attachment Site DNA. *Proc Natl Acad Sci U S A* **82**, 1040-1044.

Yu, A., Bertani, L. E. & Haggard-Ljungquist, E. (1989). Control of prophage integration and excision in bacteriophage P2: nucleotide sequences of the *int* gene and *att* sites. *Gene* **80**, 1-11.

Appendix I: Full List of Wild-Type STEC Strains and Integrase Screening Results

Leahurst DS Ref#	Int Group	VLA Ref#	Int Group	Johansen Ref#	Int Group
4	4	1471/00	24B	1486 (15)	2 + 24B
15	4	864/00	6B + 24B	1607 (9)	1, 2 + 24B
51	4	144/99	4 + 24B	4981 97(19)	1 + 2
66	4	59/99	24B	C21b (1)	4 + 7
94	4	1464/00	4, 6B + 24B	4971 97 (34)	2
97	4	1299/00	4 + 24B	1189 (5)	7
107	4	330/01	24B	1182 (5)	1 + 3
113	4	1466/00	24B		
114	4	1461/00	24B		
117	4	1812/00	4 + 24B		
119	4	1579/00	4 + 24B		
121	4	445/99	24B		
122	1+4	1472/00	24B		
124	4				
125	4				
140	-				
145	4				
151	4				
152	4				
153	4				
157	4				
177	4				
182	1+4				
187	4				
188	1+4				
189	1+4				
209	1+4				
231	3+4				
232	3,4+6B				
233	3+4				
236	4				
237	3,4+6B				
238	1,2,5A+7				
241	1+3				
254	4				
255	4				
256	-				
257	4				
260	4				
267	4+7				
268	4+7				
272	4				
273	4				
279	3,4+7				
280	3,4+7				

318	4, 6B + 24B
319	4 + 24B
328	4 + 6B
329	4 +24B
330	1, 4 + 24B
331	4 + 24B
332	4, 6B + 24B
333	4
334	4 + 24B
335	4 + 24B
336	4 + 24B
337	4 + 24B
338	24B
339	24B
340	1 + 24B
341	1, 4 + 24B
342	2, 4 + 24B
348	-
349	1 + 4 + 6B
350	1
352	1
360	1, 6B + 24B
366	24B
368	4 + 24B
369	4
371	4 + 24B
372	4 + 24B
373	4 + 24B
374	4
375	4 + 24B
376	6B + 24B
377	4, 6A + 6B
378	4 + 24B
379	4 + 6A
380	6A + 6B
382	1 + 6B
383	4 + 24B
384	6A
387	24B
390	1 + 24B
391	4
392	4
397	1
399	24B
400	1 + 24B
406	6B + 24B
408	6B
410	4 + 24B
415	1, 4 + 24B
424	1, 4, 6B + 24B
431	-
437	6A
439	4, 6A + 6B

456	-
459	4

Appendix I: Full list of integrase groups identified by PCR screening for each of the induced STEC strains from Leahurst, Veterinary Laboratories Agency and Johansen *et al.* (2001).

Appendix II: Statistical analysis of $\Phi 24_B$::Kan infections of MC1061 wild-type, MC1061 *intS*⁻ and MC1061 *intS*⁻ complemented with pKT230 + *intS*

Descriptive Statistics: Wild-Type, Knock-Out, Complemented

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
WT	6	0	320.7	31.4	77.0	223.0	223.8	355.0	381.3	388.0
KO	6	0	62.00	5.58	13.67	41.00	53.00	61.00	73.00	82.00
Comp	6	0	63.83	4.48	10.98	46.00	54.25	65.50	74.25	75.00

Two-Sample T-Test and CI: Wild-Type vs. KO

	N	Mean	StDev	SE Mean
WT	6	320.7	77.0	31
KO	6	62.0	13.7	5.6

```
Difference = mu (WT) - mu (KO)
Estimate for difference:  258.667
95% CI for difference:  (176.621, 340.713)
T-Test of difference = 0 (vs not =): T-Value = 8.10  P-Value = 0.000  DF = 5
```

Two-Sample T-Test and CI: Knock-Out vs. Complemented

	N	Mean	StDev	SE Mean
KO	6	62.0	13.7	5.6
Comp	6	63.8	11.0	4.5

```
Difference = mu (KO) - mu (Comp)
Estimate for difference:  -1.83333
95% CI for difference:  (-18.02440, 14.35774)
T-Test of difference = 0 (vs not =): T-Value = -0.26  P-Value = 0.804  DF = 9
```

Two-Sample T-Test and CI: Wild-Type vs. Complemented

	N	Mean	StDev	SE Mean
WT	6	320.7	77.0	31
Comp	6	63.8	11.0	4.5

```
Difference = mu (WT) - mu (Comp)
Estimate for difference:  256.833
95% CI for difference:  (175.233, 338.434)
T-Test of difference = 0 (vs not =): T-Value = 8.09  P-Value = 0.000  DF = 5
```