Identification of multiple integration sites for Stx-phage $\Phi 24_B$ in the *Escherichia coli* genome, description of a novel integrase and evidence for a functional anti-repressor

Paul C. M. Fogg, Sharon M. Gossage, Darren L. Smith, Jon R. Saunders, Alan J. McCarthy and Heather E. Allison

Correspondence
Heather E. Allison
hallison@liv.ac.uk

Microbiology Research Group, Division of Integrative Biology, School of Biological Sciences, University of Liverpool, Liverpool L69 7ZB, UK

The key virulence factor in Shiga-toxigenic *Escherichia coli* is the expression of Shiga toxin (Stx), which is conferred by Stx-encoding temperate lambdoid phages (Stx-phages). It had been assumed that Stx-phages would behave similarly to λ phage. However, contrary to the λ superinfection immunity model, it has been demonstrated that double lysogens can be produced with the Stx-phage $\Phi 24_B$. Here, the $\Phi 24_B$ integrase gene is identified, and the preferred site of integration defined. Although an *E. coli* *int* gene was identified close to the $\Phi 24_B$ integration site, it was shown not to be involved in the phage integration event. An additional six potential integration sites were identified in the *E. coli* genome, and three of these were confirmed experimentally. Two of the other potential sites lie within genes predicted to be essential to *E. coli* and are therefore unlikely to support phage integration. A $\Phi 24_B$ gene, possessing similarity to the well-characterized P22 *ant* gene, was identified. RT-PCR was used to demonstrate that *ant* is transcribed in a $\Phi 24_B$ *E. coli* lysogen, and expression of an anti-repressor is the likely explanation for the absence of immunity to superinfection. Demonstration of the ability of $\Phi 24_B$ to form multiple lysogens has two potentially serious impacts. First, multiple integrated prophages will drive the evolution of bacterial pathogens as novel Stx-phages emerge following intracellular mutation/recombination events. Second, multiple copies of the *stx* gene may lead to an increase in toxin production and consequently increased virulence.

Received 18 July 2007
Revised 13 September 2007
Accepted 24 September 2007

INTRODUCTION

The integrases commonly associated with bacteria belong to a family of tyrosine recombinases involved in site-specific recombination (Rice, 2005). Some integrases are involved in the integration and excision of phage genomes, whilst others are essential for the maintenance of plasmid copy number or elimination of chromosomal dimers (Chen *et al.*, 2000; Sherratt *et al.*, 2004; Van Duyne, 2001). The first tyrosine recombinase extensively studied *in vitro* was that of bacteriophage λ (Craig, 2001; Ptashne, 2004). The minimal substrates for λ integration are a 21 bp segment of host DNA (*attB*) that surrounds the point of crossover (O_B) and ~240 bp of specific bacteriophage sequence (*attP*) that flanks the bacteriophage DNA crossover point (O_P). A combination of phage- and bacterial-encoded proteins regulates complete integration of the λ bacteriophage genome into the host chromosome (Biswas *et al.*, 2005).

Stx-phages are lambdoid phages that confer the ability to produce Shiga toxin (Stx) on their bacterial host. A partially successful attempt was made elsewhere to characterize the integrases and integration sites of seven Stx-phage isolates via a DNA-amplification strategy that employed oligonucleotide primer sets specific for the integrase genes from bacteriophages λ and 933W (Johansen *et al.*, 2001). Bacteriophage 933W was one of the earliest characterized Stx-phages for which there is a complete sequence (O'Brien *et al.*, 1984). Of the seven isolates induced and challenged with the primers, none was positive for λ -like integrase and only one was positive for 933W-like integrase, highlighting the need for a more universal integrase detection technique. The 933W-like integrases are the most common type associated with the Stx-phages sequenced to date (Johansen *et al.*, 2001). The integrase from the short-tailed Stx-phage $\Phi 24_B$, which has a relatively broad host range (James *et al.*, 2001) and unusually is capable of multiple lysogeny (Allison *et al.*, 2003), could not be identified using this published strategy, even though it shares a considerable amount of genomic DNA sequence with Stx-phage 933W (data not shown). In

Abbreviation: Stx, Shiga toxin.

view of the sequence diversity found amongst integrases and the mosaicism inherent in bacteriophage populations (Hatfull *et al.*, 2006; Johansen *et al.*, 2001), there is considerable scope for Stx-phage integration directed by as yet uncharacterized factors.

An initial comparison of 28 integrase protein sequences revealed the presence of four conserved amino acid residues (Arg, His-X-X-Arg and Tyr), known as the RHRY motif, in the more highly conserved C-terminal domain. There is good evidence that the RHRY motif forms the active site of the enzyme. Subsequently, a comparison of 105 recombinases identified two boxes of amino acid similarity that were present in all known integrases: box I (A²⁰²–G²²⁷ in λ), encompassing the His and both Arg residues; and box II (T³⁰⁶–D³⁴⁴ in λ), flanking the Tyr residue (Esposito & Scocca, 1997; Nunes-Duby *et al.*, 1998). These conserved regions were used to align the amino acid sequence of 32 inducible enteric phage integrases, resulting ultimately in their classification into eight distinct groups, for which a bank of 11 degenerate primer sets was designed to amplify all known integrases (Balding *et al.*, 2005).

Bacteriophages frequently integrate within, or in the vicinity of, tRNA genes; however, in the limited number of Stx-phages described thus far, tRNA insertion sites are rarely preferred (Campbell, 2003; Herold *et al.*, 2004). The model Stx-phage 933W and the non-inducible (remnant) phage VT2-Sakai, both encoding Stx2, are integrated into their respective hosts within the *wrbA* gene, which encodes a tryptophan repressor-binding protein (Plunkett *et al.*, 1999), whilst remnant Stx1-phages CP-933V and VT1-Sakai are integrated into *yehV* (a putative transcriptional regulator) (Makino *et al.*, 1999). The two Stx2-phages have completely identical integrases, as do the two Stx1-phages, with 62.6% amino acid identity shared between the two groups. Other known Stx-phage integration sites include *sbcB* (Stx2-phage; Makino *et al.*, 1999; Ohnishi & Hayashi, 2002), *ssrA* (cryptic CP-1639; Creuzburg *et al.*, 2005), *yecE* (Φ P27; Recktenwald & Schmidt, 2002), *varphi*297 (De Greve *et al.*, 2002), various phages from an O157:NM strain as well as various O26 strains (Bielaszewska *et al.*, 2006, 2007, respectively), and Z2755 (phage-6220; Koch *et al.*, 2003). The first two encode exonuclease I and 10Sa tmRNA, respectively, whilst the others encode hypothetical proteins of unknown function (Herold *et al.*, 2004). The temperate bacteriophages characterized to date have a distinct preference for a single specific integration site within the host chromosome, although it cannot be ruled out that secondary sites may be occupied under some circumstances. Indeed, the temperate bacteriophages λ and P2 utilize various insertion sites when the primary site is unavailable (Barreiro & Haggard-Ljungquist, 1992; Rutkai *et al.*, 2003).

It has been demonstrated unequivocally, through the use of differentially labelled, isogenic recombinant phages (Φ 24_B::Cat and Φ 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 λ via immunity mutant experiments or homologous recombination events (Calef, 1967; Freifelder & Kirschner, 1971). 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 *Escherichia coli* hosts. The underlying mechanisms responsible for multiple lysogeny in *E. coli* infected with Stx-phage Φ 24_B are unknown, and it was the objective of this study to specifically identify the Φ 24_B integrase and the locations of the *attB* and the *attP* sites that are required for site-specific recombination. In addition, the factors that overcome the λ immunity system to permit multiple infections of isogenic temperate phages are investigated.

METHODS

Bacterial strains, bacteriophages and media. The bacterial strains, bacteriophages and plasmids used in this study are listed in Table 1. *E. coli* K-12 strain MC1061-Rif was the host of choice for bacteriophage infection and lysogen production. JM109- λ pir was the propagation strain for suicide plasmids, while Invitrogen One Shot TOP10 cells were used for all other recombinant work. Bacteria were routinely cultured in phage buffer [2.5%, w/v, Luria–Bertani (LB) broth (Merck) with 0.01M CaCl₂] with or without 1.5% (w/v) agar (Merck). Bacteriophage suspensions were routinely stored in the phage buffer at 4 °C. Where appropriate, the following antibiotics were added: rifampicin (3500 µg ml⁻¹), ampicillin (100 µg ml⁻¹), kanamycin (50 µg ml⁻¹), chloramphenicol (50 µg ml⁻¹), spectinomycin (100 µg ml⁻¹), tetracycline (10 µg ml⁻¹).

Lysogen production. Host bacteria were cultured to OD₆₀₀ 0.5 in phage buffer at 37 °C with shaking. Aliquots of these cultures were mixed with phage suspensions at an m.o.i. of 0.1 in a final volume of 100 µl. After 25 min, the mixture was plated onto selective media (LB agar), containing the appropriate antibiotic(s), and incubated overnight at 37 °C prior to identification and enumeration of resultant colonies (presumptive lysogens). Lysogens were confirmed as described previously (James *et al.*, 2001). Double and multiple lysogens were created as described previously (Allison *et al.*, 2003); briefly, a colony of a confirmed lysogen (obtained after either a primary or a secondary infection) was cultured to OD₆₀₀ 0.5 in phage buffer at 37 °C with shaking, and then infected with a differentially labelled phage to produce a double lysogen, resistant to two different antibiotics and capable of producing both phages. Multiple lysogens were produced as above, except that a double lysogen was infected with a differentially labelled phage.

PCR amplification parameters. DNA was amplified using 1.25 U Expand High Fidelity DNA polymerase (Roche Diagnostics) in a reaction mix containing 1 mM dNTP mix, 1.5 mM MgCl₂, proprietary buffer, 200 nM oligonucleotide primers (Table 2) and 100 ng template DNA. Cycling conditions consisted of an initial denaturation at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing for 1 min at various temperatures (Table 2) and extension at 68 °C for 1 min kb⁻¹, followed by a final extension at 68 °C for 7 min.

Southern blotting and chemiluminescent detection. Chromosomal DNA from naïve MC1061-Rif and Φ 24_B-infected

Table 1. Bacterial strains, bacteriophages and plasmids

Strain or plasmid	Relevant characteristics	Reference
MC1061-Rif	K-12 derivative, Rif ^R	James <i>et al.</i> (2001)
MC1061 Δ intS	MC1061 intS ⁻	This study
JM109- λ pir	K-12 derivative, λ pir ⁺	Penfold & Pemberton (1992)
One Shot TOP10	Chemically competent cells	Invitrogen
Φ 24 _B ::Cat	Stx-phage, stxA ₂ Δ cat	Allison <i>et al.</i> (2003)
Φ 24 _B ::Kan	Stx-phage, stxA ₂ Δ aph3	Allison <i>et al.</i> (2003)
Φ 24 _B ::Spec	Stx-phage, stxA ₂ Δ aadA	This study
NTP707	Φ 933W stx operon	Willshaw <i>et al.</i> (1987)
pKT230	Kan ^R , low copy number	Bagdasarian <i>et al.</i> (1981)
pPCMFintScomp	pKT230 carrying intS	This study
pJP5603	Kan ^R , R6K origin	Penfold & Pemberton (1992)
pPCMFintS	pJP5603 carrying intS interrupted by aadA	This study
pKNG101	Sm ^R , sacB, R6K origin	Kaniga <i>et al.</i> (1991)
pCR-Blunt	Blunt-ended cloning vector, lacZ α , Kan ^R , ccdB	Invitrogen
pCR2.1-TOPO	TA overhang cloning vector, lacZ α , Kan ^R , Amp ^R	Invitrogen

MC1061-Rif was digested with *Eco*RI, *Bam*HI or *Hinc*II (New England Biolabs) restriction endonucleases according to the manufacturer's recommendations. Each digest (1 μ g) along with 5 μ l DIG-labelled DNA marker IV (Roche Diagnostics) was separated on a 0.75% TAE agarose gel (40 mM Tris base, 20 mM glacial acetic acid, 1 mM EDTA, 0.75%, w/v, agarose) for 4 h at 4 V cm⁻¹. The DNA was transferred to a nylon membrane according to manufacturer's protocols (Roche Diagnostics). The DNA was UV cross-linked to the membrane using a UV Stratalinker 2400 (Stratagene) at 120 mJ. The membranes were placed in proprietary pre-hybridization solution for 1 h at 42 °C. The proprietary hybridization solution was then spiked with DIG-labelled probes (20 ng ml⁻¹), which were allowed to hybridize overnight at 42 °C. Washing and blocking of the membrane was carried out to the manufacturer's guidelines. Bands were resolved by chemiluminescent detection using CDP-Star reagent (Roche Diagnostics).

Colony blots. Agar plates supporting the overnight growth of putative transformants were incubated at 4 °C for 1 h and the colonies were then transferred to nylon membrane disks (Hybond-N, Amersham Pharmacia Biotech). The membranes were placed sequentially on Whatman 3MM paper soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 15 min, neutralization solution (0.5 M Tris/HCl, pH 7.5, 3 M NaCl) for 15 min, and finally 2 $\times$ SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0) for 10 min. The DNA was UV cross-linked to the nylon membrane. Cell debris was removed mechanically between two sheets of Whatman 3MM paper soaked in distilled H₂O. Hybridization and chemiluminescent detection were carried out as described above for Southern blotting.

Creation of an isogenic MC1061 intS mutant. The CPS-53 (KpLE1) putative prophage integrase intS was amplified using intS Fwd *Age*I and intS Rev *Sal*I primers (Table 2) using Expand High Fidelity polymerase. The DNA was purified using the Eppendorf Perfectprep gel extraction kit (Eppendorf) and cloned into Zero Blunt cloning vector (Invitrogen) according to the manufacturers' protocols. Plasmid from the resulting transformants was purified using the Qiagen Miniprep kit and subjected to PCR using intS Inverse Fwd *Acc*I and intS Inverse Rev *Acc*I oligonucleotides (Table 2). The resulting amplification product was purified and 1 μ g digested with the restriction endonuclease *Acc*I (New England Biolabs) 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 using Expand High Fidelity polymerase (Table 2). After subsequent purification, 1 μ g 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 (268 and 224 bp) and compatible *Acc*I cohesive ends. The resulting plasmid was purified using the Qiagen Miniprep kit, and the mutated intS was excised from the plasmid backbone by sequential digestion 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 pJP560 suicide vector3. This plasmid is dependent upon λ pir for replication and was routinely propagated in the λ pir⁺ *E. coli* strain JM109. The intS-containing pJP5603 construct pPCMFintS was introduced into MC1061 via electroporation (2.5 V, 25 μ F, 200 Ω) by means of the Bio-Rad Gene Pulser (Bio-Rad Laboratories). Selection for kanamycin and spectinomycin resistance permitted the identification of clones produced by a single recombination event, which were confirmed by PCR and Southern hybridization. Passage of a single recombinant followed by screening for the loss of kanamycin resistance facilitated the identification of an intS knockout in MC1061 Δ intS, and ablation of the gene was confirmed by PCR.

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 (Table 2) and cloned into the low-copy-number vector pKT230 (Table 1) to produce pPCMFintScomp according to standard cloning protocols (Sambrook *et al.*, 1989). MC1061 Δ intS was used as the *E. coli* K-12 host.

Recombinant mutant phage construction. The plasmid NTP707 was digested with the restriction endonuclease *Pst*II, whilst the spectinomycin antibiotic-resistance cassette aadA was amplified with aadA Fwd *Pst*II and aadA Rev *Pst*II (Table 2) and digested with *Pst*II. The two fragments were ligated together and the resultant plasmid, NTP707-aadA, was transformed into *E. coli* strain MC1061, which was subsequently infected with Φ 24_B::Kan. Bacteriophages resulting from this infection were recovered by membrane filtration (0.2 μ m pore-size, Millipore), 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 plus tetracycline. Colonies displaying spectinomycin resistance but kanamycin and tetracycline sensitivity were deemed to be putative Φ 24_B::Spec recombinant phage lysogens, and were further examined for the ability to produce plaques, as described previously (Allison *et al.*, 2003; James *et al.*, 2001).

Table 2. DNA amplification primers

Primer	Sequence (5'–3') or source	Annealing temperature
<i>int</i> Probe Fwd	CTGGAAGTAATCCGCAGG	50 °C
<i>int</i> Probe Rev	AGCTCTTTCGTTCTTAGG	50 °C
pGEM Fwd	ACTCACTATAGGGCGAATTCCG	55 °C
pGEM Rev	CTGCAGGTCGACTCTAGAGG	55 °C
Integ Fwd	GTGTCACTCAGTGCCAATGC	55 °C
<i>intS</i> Fwd <i>Age</i> I	TCGCTTCAGACCGGTGACAGCCGCACTCC	55 °C
<i>intS</i> Rev <i>Sal</i> I	GATGAGTCGACACTTATACACACAAAGC	55 °C
<i>ant</i> Fwd	ATGGCTGTTCAAATTTCTGTCTG	60 °C
<i>ant</i> Rev	CAATGCCTGTGAGCGGTTGG	60 °C
<i>aadA</i> Fwd <i>Acc</i> I	CGAAACCTTGCGGTGACCGCCAGCC	60 °C
<i>aadA</i> Rev <i>Acc</i> I	GCCTTTCATGATATGTCGACCAATTTGTGTAGGGC	60 °C
<i>aadA</i> Fwd <i>Pst</i> I	CGAAACCTGACAGTCGTTCCGACGCC	60 °C
<i>aadA</i> Rev <i>Pst</i> I	GCCTTTCATGATATATCTGCAGATTTGTGTAGGGC	60 °C
<i>intS</i> Inverse Fwd <i>Acc</i> I	CAGCACGATTGTCGACGAGCAC	65 °C
<i>intS</i> Inverse Rev <i>Acc</i> I	CCTCGCAGTCGACGCTTTATCC	65 °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
Naïve 1 Fwd	CGTGGTCGGTTCTGTTGAAA	50 °C
Naïve 1 Rev	GTGTATAAGGGTGTTGAGC	50 °C
Naïve 2 Fwd	CTCTTACATTCGCTTATATATTGACC	50 °C
Naïve 2 Rev	CCAATCCAGAACGCGATGG	50 °C
Naïve 3 Fwd	CAAATTCCATCAATGCCAGG	50 °C
Naïve 3 Rev	AGTTATCACCCTGCTTCACC	50 °C
Naïve 4 Fwd	GATGATCCCATTAAACGTAGTAGC	50 °C
Naïve 4 Rev	GGAGAAAGTGCATCATATACG	50 °C
24 _B <i>int</i> Probe Fwd	CACATCAAATCCGATATTAAGC	55 °C
24 _B <i>int</i> Probe Rev	CTCAACAAGAAAAAGGCTCG	55 °C
24 _B <i>int</i> Fwd	GCCAGGCTTTCTGAGCTACG	60 °C
24 _B <i>int</i> Rev	GCCTAAAATCATGCGTTCTCC	60 °C
24 _B <i>int</i> Seq F	CCCGATTACAGCAGATGC	50 °C
24 _B <i>int</i> Seq R	ACTGTAATCTACGGAATGC	50 °C
24 _B <i>int</i> Seq F2	GAAAGCCGCAAACATCAGC	60 °C
24 _B <i>int</i> Seq R2	CTCAACAAGAAAAAGGCTCG	55 °C
24 _B <i>int</i> Seq F RC	CTGCATCTGCTGTAATCG	55 °C
24 _B <i>int</i> Seq F2 RC	CTGATGTTTGCGGCTTTTCG	60 °C
24 _B <i>int</i> Seq R2 RC	AGCCTTTTCTTGTTGAGGC	55 °C
24 _B <i>int</i> Seq F3	AACCGAGCTGTTTAGCGTGC	60 °C
24 _B <i>int</i> Seq F3 RC	CGTGTATAATAAACACGACTGG	55 °C
Integrase Bank	Balding <i>et al.</i> (2005)	50 °C

RT-PCR. Total RNA was isolated using the Qiagen RNeasy RNA isolation kit according to the manufacturer's guidelines. RNA was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies), and the quality was assessed on a Syngene GeneGenius imaging system (Syngene Europe). Bacterial genomic DNA carryover was removed by incubation with RQ1 RNase-free DNase according to the manufacturer's guidelines (Promega).

Total RNA (5 µg) was combined with specific oligonucleotide primers (200 nM; Table 2), 1 mM dNTP mix, proprietary reaction buffer and 50 U BioScript reverse transcriptase (Bioline) 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 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 (according to the manufacturer's recommendations). 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.

Sequence analysis and alignments. 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>; Nicholas *et al.*, 1997), BioEdit Sequence Alignment Editor (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>; Hall, 1999) and CLUSTALW Multiple Sequence Alignment Program (<http://www.ebi.ac.uk/clustalw/>; Thompson *et al.*, 1994).

RESULTS

Identification of the phage integrase

In order to identify the integration sites for $\Phi 24_B$ in the *E. coli* genome, the fact that the phage integrase would be located near to the site of integration in the bacterial lysogen was exploited (Campbell, 1994). To specifically identify the integrase of Stx-phage $\Phi 24_B$, we used a bank of 11 degenerate PCR primers that had been designed to amplify all known integrase genes (Balding *et al.*, 2005). When applied to a suspension of $\Phi 24_B$ phage particles, a DNA fragment of the expected size was generated with only one of the primer pairs, and this product was purified and sequenced (data not shown). BLASTN analysis of this sequence revealed 100 % identity to the *intS* integrase gene of the remnant prophage element CPS-53 (KpLE1), located within the genome of *E. coli* K-12 strain MG1655 (Blattner *et al.*, 1997). A specific, DIG-labelled *intS* probe was amplified using the *int* Probe Fwd/Rev primers (Table 2) and subsequently used to probe restriction endonuclease-digested (*Eco*RI, *Bam*HI or *Hinc*II) chromosomal DNA from naïve MC1061 or MC1061 $\Phi 24_B$ lysogen cells by Southern hybridization. This enabled detection of a single DNA fragment in both the naïve MC1061 and the $\Phi 24_B$ lysogen chromosomes; however, the latter was ~1 kb larger in size in both the *Eco*RI and *Bam*HI digests (Fig. 1a). This size discrepancy suggested that the integration of

Fig. 1. Identification of the $\Phi 24_B$ integrase and integration site. (a) Southern Blot of naïve MC1061 and $\Phi 24_B$ lysogen DNA (lanes N and $\Phi 24_B$, respectively) digested with *Eco*RI, *Bam*HI or *Hinc*II (lanes E, B or H, labelled below, respectively) and hybridized with the DIG-labelled integrase probe. Arrows to the right of the blot indicate the deduced size of the DNA bands (in bp). (b) Map of the primary integration site of the $\Phi 24_B$ lysogen, depicting the location of the restriction endonuclease sites, the $\Phi 24_B$ integrase, the *intS* gene and the *attB* sequence. The numbers indicate the distance of a given location from the upstream *Eco*RI site. The *Bam*HI sites are located beyond the boundaries of the map.

bacteriophage $\Phi 24_B$ had occurred in the vicinity of the host *intS* gene (Fig. 1b). Thus, one or more of the multiple integration events previously reported for $\Phi 24_B$ in *E. coli* (Allison *et al.*, 2003) might in fact have been directed by the host-cell *intS* gene. If the creation of double lysogens by this phage was in fact directed, at least in part, by *IntS*, then double lysogens of $\Phi 24_B$ are not a consequence of the phage, but are due to host functions, and the λ phage model of immunity to superinfection is not confounded. In order to experimentally address the role that *IntS* plays in the production of double $\Phi 24_B$ lysogens, an isogenic MC1061 mutant was created in which the endogenous *intS* gene was ablated. This mutant was constructed by replacing *intS* with a spectinomycin-resistance cassette, *aadA*. Although the yield of lysogens from infection of MC1061 $\Delta intS$ with $\Phi 24_B$ was significantly lower than that obtained when wild-type MC1061 was the host ($P < 0.05$ in a parallel infection assay; data not shown), both single and double lysogens could nevertheless be produced in MC1061 $\Delta intS$. Furthermore, lysogen yields could not be returned to the original values through complementation of the *intS* deletion by reintroduction of *intS* on the low-copy-number plasmid pKT230 (Bagdasarian *et al.*, 1981). Lysogens are routinely detected in wild-type MC1061 from a single infection event of a naïve population at a rate of 1 in every 6724 cells. Double lysogens are routinely detected from a single infection of a single lysogen at a rate of 1 in every 4931 cells, and triple lysogens are routinely detected from a single infection of a double lysogen at a rate of 1 in every 339 cells (Table 3). Consequently, it can be concluded that the presence of an additional copy of *intS* on the *E. coli* genome does not explain the phenomenon of multiple lysogeny with $\Phi 24_B$.

In order to determine the complete $\Phi 24_B$ integrase gene sequence, and to define the exact site of phage insertion, *EcoRI*-digested *E. coli* MC1061 $\Delta intS$ $\Phi 24_B$ lysogen genomic DNA fragments of ~ 4.7 kb, corresponding to the increased band size (Fig. 1a), were excised and cloned into *EcoRI* pre-cut pGEM cloning vector. The resulting transformants were screened for ampicillin and spectinomycin resistance, and the presence of *aadA* in resistant colonies was confirmed by PCR amplification with *aadA* Fwd/Rev *Accl* (Table 2). In this way, 4.7 kb of DNA sequence containing the entire $\Phi 24_B$ *int* gene, the *attB* site and most of the *intS* gene interrupted by the spectinomycin cassette was obtained (Fig. 1b). Alignment of the DNA sequences from the genome sequence of *E. coli* K-12 strain MG1655 (Blattner *et al.*, 1997) and the doubly resistant transformant enabled the identification of the integration point. As expected from the original Southern blot analysis, the insertion occurs 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). The extent of complementary sequence overlap between the $\Phi 24_B$ *attP* and the MC1061 *attB* sites was determined to be 24 bp (Fig. 2a). The $\Phi 24_B$ integrase identified here (accession no. EF397940) has only

Table 3. Numbers of lysogens produced following $\Phi 24_B$ infection

Host cells	Number of lysogens*	SEM	Sample size
Naïve MC1061	2245.6	80.81	9
MC1061 $\Phi 24_B::Kan$ lysogen	3122.5	70.2	8
MC1061 $\Phi 24_B::Kan$ $\Phi 24_B::Cat$ lysogen	45 722.2	844.1	9

*The number of lysogens was determined as described previously (Allison *et al.*, 2003; James *et al.*, 2001) using the number of cells indicated: naïve MC1061, 1.51×10^8 c.f.u. ml⁻¹; MC1061 $\Phi 24_B::Kan$ lysogen, 1.54×10^8 c.f.u. ml⁻¹; MC1061 $\Phi 24_B::Kan$ $\Phi 24_B::Cat$ lysogen, 1.55×10^8 c.f.u. ml⁻¹ infected at an m.o.i. of 0.1 with $\Phi 24_B::Kan$, $\Phi 24_B::Cat$ or $\Phi 24_B::Spec$, respectively.

been found once before, on Stx-phage-86 (96 % identity; accession no. AB255463), which has yet to be described in the scientific literature. The amounts of overall amino acid sequence identity and similarity shared between the $\Phi 24_B$ *int* gene and *intS* were only 14 and 24 %, respectively. BLASTX analysis revealed a match to the conserved domain database for bacteriophage P4 integrase, whilst the closest matches in GenBank, other than phage-86, were to prophages of *Photobacterium profundum* and *Shewanella frigidimarina* (46 and 38 % identity, respectively; Fig. 3).

Location of the phage integration sites

Identification of the $\Phi 24_B$ DNA integrase sequence enabled the production of a DIG-labelled $\Phi 24_B$ -*int*-specific probe, which was allowed to hybridize to the DNA from two single lysogens (MC1061 carrying $\Phi 24_B::Cat$ or $\Phi 24_B::Kan$) and from two double lysogens (MC1061 infected first with $\Phi 24_B::Cat$ then $\Phi 24_B::Kan$ and vice versa) digested with either *EcoRI* or *AvaI*. 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. 4). The DNA fragments corresponding to the three novel insertion sites were excised and cloned to produce three subgenomic libraries. The presence of the *int*-containing sequence in each of these libraries was confirmed by PCR using $\Phi 24_B$ *int* Fwd/Rev primers (Table 2). 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 2). A 350 bp product was amplified from the library prepared from *AvaI* fragments of approximately 3 kb (Fig. 4, 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* (Blattner *et al.*, 1997). The degree of homology between the core *attP* sequence and

Fig. 2. Analysis of the Φ24_B *attB* sites. (a) Alignment of Φ24_B *attP*, naïve MC1061 *attB* and integrated lysogen sequences containing the 24 bp overlap region highlighted in black. (b) Alignment of the previously defined 24 bp overlap region of all four MC1061 integration sites. The sequences are labelled at the right with Roman numerals to indicate the identity of each *attB* site.

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 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, identified as putative integration sites A–F (PIA–F, lettered in order of their appearance in the genome), including the two already identified above. The first integration site, near the MC1061 *intS* gene, was identified *in silico* as site PID, and will be referred to as site I. 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 2). Of these, only PCR amplification of DNA prepared from a double lysogen (MC1061 infected first with Φ24_B::Cat then Φ24_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 produced a 400 bp band. 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. 4).

Box I:

Box II:

Fig. 3. Comparison of Φ24_B integrase sequence with that of the most closely related integrases. CLUSTALW amino acid sequence alignments are shown of the conserved box I and box II regions of the integrases of Φ24_B, P4 and the three closest BLASTX matches.

Fig. 4. 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 DNA was obtained from lysogens, infected with the phages indicated above the lanes. In all cases, single lysogens were obtained following a single infection event and the double lysogens (found on the right-hand side of the blot; the top phage name is the second infecting phage) were obtained following a second phage infection of the single lysogens (left-hand side of the blot). DNA was digested with either *Eco*RI (E) or *Ava*I (A). Bands corresponding to the identified integration sites are identified with Roman numerals at either the left (E lanes) or right (A lanes). *Bands used to produce the subgenomic libraries. Arrows to the right of the figure indicate approximate molecular mass. The bands corresponding in size to *Ava*I-digested 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=PIC, III=PIB. PIA and PIF lie within essential genes (Gerdes *et al.*, 2003).

In order to identify the final integration site (site IV) indicated by the data in Fig. 4, transformants produced from the library created with 10 kb *Eco*RI fragments from a double lysogen (MC1061 infected first with $\Phi 24_B::\text{Kan}$ then $\Phi 24_B::\text{Cat}$; Fig. 4, fragment IV*) were subjected to colony blot hybridization with a $\Phi 24_B$ *int*-specific DIG-labelled probe. 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 2). The plasmids from three transformant clones were purified and sequenced. BLASTN analysis revealed that the final integration site indicated in Fig. 4 fell within the *sgcA* gene, which encodes a putative transport protein encoded on the KpLE2 phage-like element in *E. coli* K-12 MG1655 (Blattner *et al.*, 1997). Alignment of the sequence to the other three insertion sites reduced the consensus from a motif containing 11 conserved residues to one containing 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, to site I (Fig. 3b). Site IV was not identified *in silico*. The sequences of all insertion sites in the naïve *E. coli* MC1061 genome were compared to those found in the sequenced *E. coli* strain MG1655 (Blattner *et al.*, 1997) using primer pairs Naïve 1–4 Fwd/Rev (Table 2) 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* (Gerdes *et al.*, 2003), and are therefore unlikely to be utilized for phage integration to produce lysogens.

A third recombinant phage, $\Phi 24_B::\text{Spec}$, was constructed and used to infect an existing double lysogen culture. Putative triple lysogens were confirmed by screening for resistance to kanamycin, chloramphenicol and spectinomycin, as well as the ability to release all three marked phages upon induction. The $\Phi 24_B$ 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 in a non-essential gene for which actual $\Phi 24_B$ 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 $\Phi 24_B$ has therefore been demonstrably extended to the ability to form triple lysogens, and four different sites of bacteriophage integration, which are scattered across the chromosome, have been identified within the *E. coli* host genome (Fig. 2b). However, the question of how the λ superinfection immunity model (Ptashne, 2004; Allison *et al.*, 2003) is overcome remained.

Identification of an expressed anti-repressor gene

The entire sequence of the $\Phi 24_B$ immunity region, from O_L through to *cro* (accession no. EF517298), was determined and found to be identical to the immunity region of the fully sequenced Stx-phage 933W. There are a few notable differences between the $\Phi 24_B$ immunity region and that of the bacteriophage λ immunity region, which have been

described elsewhere (Fattah *et al.*, 2000; Koudelka *et al.*, 2004; Tyler *et al.*, 2004). Briefly, O_L possesses only two operator binding regions, the spacing of the promoters p_R and p_{RM} between the three binding sites within O_R is different, and the third operator binding region lies within the coding region of the *cl* gene. There are two alternatives to explain the formation of double lysogens. Either the immunity region is sufficiently different from λ such that it has reduced function or there is an anti-repressor. Earlier work (Fattah *et al.*, 2000; Koudelka *et al.*, 2004; Tyler *et al.*, 2004) suggests that the first option is unlikely. By examining sequences of other anti-repressor proteins (encoded by *ant* genes), a gene possessing 37% homology to the P22 anti-repressor gene *ant* (accession no. NP_059643), was identified in unannotated large fragments of the $\Phi 24_B$ genome sequence. The Ant protein of P22 has the ability to inactivate the P22 repressor protein, which is analogous to the λ C1 repressor protein, essential for maintenance of lysogeny and homoimmunity to superinfection by this bacteriophage (Botstein *et al.*, 1975; Susskind & Botstein, 1975). The P22 anti-repressor is also known to be capable of neutralizing the λ repressor (Schaefer & McClure, 1997). Alignment of the $\Phi 24_B$ anti-repressor protein with its P22 counterpart revealed significant sequence similarity between the C-terminal domains (82% identity for the final 104 C-terminal residues), while the sequences of the N termini were distinct (Fig. 5). In contrast, alignment of $\Phi 24_B$ Ant to the 933W anti-repressor, which appears to be truncated and incomplete, produced an apparent low level of shared amino acid sequence (12% sequence identity). In order to establish whether the *ant* gene is transcribed in $\Phi 24_B$ lysogens, total RNA from mid-exponential naïve and lysogen cultures was harvested. In addition, total RNA of mid-exponential *E. coli* MC1061 cultures undergoing $\Phi 24_B$

infection at two m.o.i. values (0.1 and 1.0) was also harvested. The RNA samples were subjected to RT-PCR with internal *ant*-specific primers (Table 2), expected to produce a 456 bp fragment. The use of 24_B *int* Fwd/Rev primers, which should produce a fragment 414 bp in length, served as an internal positive control for the RT-PCR. In this way, the presence of *ant* mRNA transcripts was identified in the lysogen/infection samples, and these were absent in the control RNA harvested from the naïve *E. coli* cultures (Fig. 6). Thus, this $\Phi 24_B$ gene is actively expressed during bacteriophage infection and, as demonstrated for phage P22 (Susskind & Botstein, 1978), could explain the avoidance of homoimmunity leading to the creation of multiple lysogens by the lambdoid Stx-phage described here.

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

Fig. 5. Relatedness of $\Phi 24_B$ putative anti-repressor sequence to other known phage antiterminators. A CLUSTALW multiple sequence alignment is shown of $\Phi 24_B$ Ant to the well-defined P22 Ant, Ant from the two closest BLASTX matches (Lahn1 and VT2-Sa) and the Ant protein from the truncated gene of the well-characterized Stx-phage 933W.

Fig. 6. Analysis of $\Phi 24_B$ putative *ant* transcription. (a) RT-PCR was performed on the total RNA from naïve MC1061 and $\Phi 24_B$ lysogen to demonstrate the transcription of $\Phi 24_B$ *ant* in the lysogen, indicated by the presence of a 456 bp fragment. Lanes: 1, MC1061 cDNA+*ant* Fwd/Rev; 2, lysogen cDNA+*ant* Fwd/Rev; 3, MC1061 RNA+*ant* Fwd/Rev; 4, lysogen RNA+*ant* Fwd/Rev; 5, MC1061 DNA+*ant* Fwd/Rev; 6, lysogen DNA+*ant* Fwd/Rev; 7, lysogen cDNA+ $\Phi 24_B$ *int* Fwd/Rev. (b) RT-PCR was performed on the total RNA from MC1061 cells that were infected with $\Phi 24_B$ at different m.o.i. Lanes: 1, m.o.i. 0.1 with cDNA+*ant* Fwd/Rev; 2, m.o.i. 0.1 with cDNA+ $\Phi 24_B$ *int* Fwd/Rev; 3, m.o.i. 0.1 with RNA+*ant* Fwd/Rev; 4, m.o.i. 1.0 with cDNA+*ant* Fwd/Rev; 5, m.o.i. 1.0 with cDNA+ $\Phi 24_B$ *int* Fwd/Rev; 6, m.o.i. 1.0 with RNA+*ant* Fwd/Rev.

within a host is limited by two major mechanisms: prophage loss or decay, and immunity to superinfection by a homoimmune phage (Brussow *et al.*, 2004), the latter of which is the focus of this study. The ability of a temperate Stx-encoding bacteriophage ($\Phi 24_B$) to form double lysogens has been described previously (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 characterize this phenomenon by describing four distinct integration sites utilized by $\Phi 24_B$ in an *E. coli* K-12 genome, and extend our observation of double lysogeny to achieve the production of a triple lysogen. Multiple insertion sites for bacteriophages in bacterial genomes have been described previously only when the primary site is either absent or occupied, i.e. they are not utilized concurrently (Barreiro & Haggard-Ljungquist, 1992; Rutkai *et al.*, 2003).

Although $\Phi 24_B$ does possess an integrase gene with an active site similar to that of the *intS* gene of *E. coli* strain MG1655, and $\Phi 24_B$ does integrate near the integration site utilized by *intS*, the integration site for $\Phi 24_B$ is distinct from the integration site recognized by the *E. coli intS* gene product. Creation of an *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). The expression of these genes may have been affected in the *intS* knockout mutant, with

concomitant effects on cell fitness and susceptibility to Stx-phage infection. Phage $\Phi 24_B$ does appear to prefer a single site (I) for integration, and this is near the *E. coli* chromosomal copy of *intS*. Site I possesses 24 bp of 100 % sequence identity to the *attP* site of $\Phi 24_B$, and without fail it was occupied in all of the lysogen DNA preparations examined by Southern hybridization. The supplementary sites, which are scattered across the *E. coli* genome, 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 an 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 functions for *E. coli* (Gerdes *et al.*, 2003). A defined, minimal consensus sequence cannot yet be reported, as the production of at least five 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. It is also clear from our data that the integrated $\Phi 24_B$ prophages in double and multiple lysogens are not tandemly linked to one another but are scattered across the genome, each having integrated into the chromosomal molecule independently. It was also found that a single infection event can result in a double integration event, as shown in Fig. 4, lanes 1 and 2. This chloramphenicol-resistant lysogen has, at some point, undergone two separate integration events with the same phage. The DNA sequence of $\Phi 24_B$ *int* identified here shares 100 % identity within the conserved boxes (Fig. 3) 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 no. NC 008464) isolated from an enterohaemorrhagic *E.*

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 area that these two phages appear to share is the integrase gene with 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 genome has not been described; however, given the integrase and *attP* overlap similarities with $\Phi 24_B$, it can be suggested that both phages utilize the same site.

The precise mechanism underlying the non-functional $\Phi 24_B$ superinfection immunity system is as yet unclear. Alignment of the immunity regions of $\Phi 24_B$ and the model Stx-phage 933W (Plunkett *et al.*, 1999) revealed 100% identity in sequence conservation between the two phages (accession nos EF517298 and AF125520, respectively). Nevertheless, in contrast to bacteriophage λ , for 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* anti-repressor 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, 1978). The P22 *ant* gene product has also been shown to be effective against λ 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. The data presented here do not provide definitive evidence that *ant* suppresses immunity to superinfection, but do provide supporting evidence that the $\Phi 24_B$ *ant* could be involved. 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 sequences surrounding the $\Phi 24_B$ *ant* gene do not encode comparable control factors. In fact, the $\Phi 24_B$ *ant* gene is located 6.5 kb downstream of the *cro* gene and 1.2 kb upstream of the *Q* gene (data not shown). Our demonstration that the $\Phi 24_B$ anti-repressor 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 existing prophage. Although there is a $\Phi 24_B$ *ant* homologue in the 933W genome, it does not appear to be intact, as very little sequence identity is shared with $\Phi 24_B$ *ant*, and in 933W this homologue is located just downstream of the lysis

gene. Further work will be required to confirm that the Ant protein interacts with the $\Phi 24_B$ repressor and to understand how and when *ant* is expressed.

In addition to insertion site identification, Southern hybridization analyses revealed the presence of a band that could only have been produced by *AvaI* restriction endonuclease digestion if the phage genome was circularized (Fig. 4), with the recognized *AvaI* sites lying at each distal end of the integrated prophage sequence, as extrapolated from $\Phi 24_B$ 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 circularized plasmid-like pseudolysogens. 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 might 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 might also result in higher than expected levels of Stx production. Increased Stx production might be expected to result in more severe STEC-mediated disease and higher rates of complication with sequelae such as haemolytic uraemic syndrome (HUS) (Siegler *et al.*, 2001).

The ability of a single bacterium to harbour multiple phages, whether closely related or not, could have a profound impact on pathogen evolution, providing an increased gene pool for genetic exchange *in situ* (Allison, 2007) and possibly even supporting enhanced spontaneous mutation rates of duplicated/redundant phage genes that may ultimately result in the formation of new phage components that alter host range, etc. Furthermore, the presence of several *stx* gene copies within a host might in turn lead to amplification of toxin load and, upon lysis, manifestation of more severe disease pathology. The mechanisms that underpin the immunity patterns observed need to be further investigated, but evidence for an actively transcribed anti-repressor is an important first step. The potential of the anti-repressor to influence immunity and the effect of divergence of the immunity region structure from the established λ 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.

ACKNOWLEDGEMENTS

We would like to thank the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the UK Department for Environment, Food and Rural Affairs (DEFRA) for their financial support of this project through both a funded research studentship (P. C. M. F.) and

project grants. We also gratefully acknowledge Dr Nick R. Thomson at the Sanger Centre for the release of raw contig sequence data obtained while sequencing the genome of $\Phi 24_B$.

REFERENCES

- Allison, H. E. (2007). Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. *Future Microbiol* 2, 165–174.
- Allison, H. E., 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.
- 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.
- Barreiro, V. & Haggard-Ljungquist, E. (1992). 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.
- Bielaszewska, M., Prager, R., Zhang, W., Friedrich, A. W., Mellmann, A., Tschape, H. & Karch, H. (2006). Chromosomal dynamism in progeny of outbreak-related sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM. *Appl Environ Microbiol* 72, 1900–1909.
- Bielaszewska, M., Prager, R., Kock, R., Mellmann, A., Zhang, W., Tschape, H., Tarr, P. I. & Karch, H. (2007). Shiga toxin gene loss and transfer *in vitro* and *in vivo* during enterohemorrhagic *Escherichia coli* O26 infection in humans. *Appl Environ Microbiol* 73, 3144–3150.
- Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. & Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by λ integrase. *Nature* 435, 1059.
- Blaisdell, B. E., Campbell, A. M. & Karlin, S. (1996). Similarities and dissimilarities of phage genomes. *Proc Natl Acad Sci U S A* 93, 5854–5859.
- Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462.
- Botstein, D. (1980). A theory of modular evolution for bacteriophages. *Ann 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, E. F. & Brussow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol* 10, 521.
- 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.
- Calef, E. (1967). Mapping of integration and excision crossovers in superinfection double lysogens for phage lambda in *Escherichia coli*. *Genetics* 55, 547–556.
- 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.
- 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.
- Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. (2000). Crystal structure of a Flp recombinase–Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* 6, 885–897.
- 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.
- Craig, N. L. (2001). *Mobile DNA II*. Washington, DC: American Society for Microbiology.
- Creuzburg, K., Kohler, B., Hempel, H., Schreier, P., Jacobs, E. & Schmidt, H. (2005). Genetic structure and chromosomal integration site of the cryptic prophage CP-1639 encoding Shiga toxin 1. *Microbiology* 151, 941–950.
- De Greve, H., Qizhi, C., Deboeck, F. & Hernalsteens, J. P. (2002). The Shiga-toxin VT2-encoding bacteriophage varphi297 integrates at a distinct position in the *Escherichia coli* genome. *Biochim Biophys Acta* 1579, 196–202.
- Esposito, D. & Scocca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* 25, 3605–3614.
- 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.
- 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.
- Gerdess, S. Y., Scholle, M. D., Campbell, J. W., Balázs, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I. & other authors (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* 185, 5673–5684.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41, 95–98.
- Hatfull, G. F., Pedulla, M. L., Jacobs-Sera, D., Balázs, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I. & other authors (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet* 2, e92.
- Herold, S., Karch, H. & Schmidt, H. (2004). Shiga toxin-encoding bacteriophages – genomes in motion. *Int J Med Microbiol* 294, 115–121.
- 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.

- Koch, C., Hertwig, S. & Appel, B. (2003). Nucleotide sequence of the integration site of the temperate bacteriophage 6220, which carries the Shiga toxin gene *stx*(10x3). *J Bacteriol* **185**, 6463–6466.
- 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.
- Livny, J. & Friedman, D. I. (2004). Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. *Mol Microbiol* **51**, 1691–1704.
- Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I. & 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.
- Nicholas, K. B., Nicholas, H. B., Jr & Deerfield, D. W., II (1997). GeneDoc: analysis and visualization of genetic variation. *EMBnet News* **4** (2). http://www.embnet.org/files/shared/EMBnetNews/embnet_news_4_2.pdf.
- 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. *Nucleic Acids Res* **26**, 391–406.
- O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W. & Formal, S. B. (1984). Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226**, 694–696.
- Ohnishi, M. & Hayashi, T. (2002). Genetic diversity of enterohemorrhagic *Escherichia coli*. *Nippon Rinsho* **60**, 1077–1082 (in Japanese).
- Penfold, R. J. & Pemberton, J. M. (1992). An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene* **118**, 145–146.
- 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.
- Ptashne, M. (2004). *A Genetic Switch: Phage Lambda Revisited*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Recktenwald, J. & Schmidt, H. (2002). The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage ϕ P27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. *Infect Immun* **70**, 1896–1908.
- Rice, P. A. (2005). Resolving integral questions in site-specific recombination. *Nat Struct Mol Biol* **12**, 641.
- Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M. & other authors (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot – 2005. *Nucleic Acids Res* **34**, 1–9.
- Rutkai, E., Dorgai, L., Sirot, R., Yagil, E. & Weisberg, R. A. (2003). Analysis of insertion into secondary attachment sites by phage λ 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, NY: Cold Spring Harbor Laboratory.
- Saunders, J. R., Allison, H. E., James, C. E., McCarthy, A. J. & Sharp, R. (2001). Phage-mediated transfer of virulence genes. *J Chem Tech Biotech* **76**, 662–666.
- Schaefer, K. L. & McClure, W. R. (1997). Antisense RNA control of gene expression in bacteriophage P22. I. Structures of *sar* RNA and its target, *ant* mRNA. *RNA* **3**, 141–156.
- Sherratt, D. J., Soballe, B., Barre, F. X., Filipe, S., Lau, I., Massey, T. & Yates, J. (2004). Recombination and chromosome segregation. *Philos Trans R Soc Lond B Biol Sci* **359**, 61–69.
- 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.
- 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.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- 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.
- Van Duyne, G. D. (2001). A structural view of Cre-*loxP* site-specific recombination. *Annu Rev Biophys Biomol Struct* **30**, 87–104.
- 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.

Edited by: D. L. Gally