Characterization of the *Salmonella* bacteriophage vB_SenS-Ent1

Dann Turner, Mohammed Hezwani, Shona Nelson, Vyv Salisbury and Darren Reynolds

Correspondence
Darren Reynolds
Darren.Reynolds@uwe.ac.uk

Centre for Research in Biosciences, Department of Applied Sciences, University of the West of England, Coldharbour Lane, Bristol BS16 1QY, UK

Received 31 March 2012
Accepted 7 June 2012

The bacteriophage vB_SenS-Ent1 (Ent1) is a member of the family *Siphoviridae* of tailed bacteriophages and infects a broad range of serovars of the enteric pathogen *Salmonella enterica*. The virion particle is composed of an icosahedral head 64 nm in diameter and a flexible, non-contractile tail of 116 × 8.5 nm possessing terminal fibres. The adsorption rate constant at 37 °C is 6.73×10^{-9} ml min^{-1}. Latent and eclipse periods are 25 and 20 min, respectively, and the burst size is 35 progeny particles per cell after 35 min at 37 °C. Sequencing revealed a circularly permuted, 42 391 bp dsDNA genome containing 58 ORFs organized into four major transcriptional units. Comparisons with the genome sequences of other bacteriophages revealed a high level of nucleotide sequence identity and shared orthologous proteins with the *Salmonella* phages SETP3, SE2 and KS7 (SS3e) and the *Escherichia* phages K1G, K1H, K1ind1 and K1ind3.

INTRODUCTION

Salmonellae are Gram-negative, facultatively anaerobic, non-sporulating and generally motile bacilli that are causative agents of typhoid fever, gastroenteritis and enteric fever in both humans and animals. The species *Salmonella enterica* is classified into five subspecies by differential biochemistry and further subdivided into serotypes based upon serology of the lipopolysaccharide (LPS) (O) and flagellar (H) antigens (Grimont & Weill, 2007). Epidemiological surveillance data present serovars of *Salmonella enterica* subspecies *enterica* as prominent aetiological agents of bacterial food-borne disease worldwide (ECDC, 2010; Scallan *et al.*, 2011). Non-typhoidal infections by *Salmonella* serovars other than Typhi and Paratyphi are generally self-limiting, with clinical manifestations ranging from mild to severe gastroenteritis. In a small proportion of cases, further complications arise, including bacteraemia, gastrointestinal bleeding and focal infections (Acheson & Hohmann, 2001). Transmission to humans is primarily associated with the ingestion of a wide variety of contaminated food products, but may also arise by contact with animals, contaminated water and infected individuals (Hanning *et al.*, 2009). In the USA, the economic burden associated with medical care and lost productivity due to salmonellosis is estimated at several billion dollars annually (Voetsch *et al.*, 2004).

The GenBank/EMBL/DDBJ accession number for the whole-genome sequence of *Salmonella* bacteriophage vB_SenS-Ent1 is HE775250.

Five supplementary tables and one figure are available with the online version of this paper.

All *Salmonella* phages reported thus far belong to the order *Caudovirales* (tailed phages) and represent three families: the *Siphoviridae*, *Podoviridae* and *Myoviridae*. Of these, only a small fraction are lytic, whilst the majority are capable of a temperate life cycle (Kropinski *et al.*, 2007). Lysogeny is widespread among strains of *Salmonella* and most, if not all, harbour prophages. To date, the complete genome sequences of 36 *Salmonella* phages are publicly available, representing just 3.97 % (35/880) of all available bacteriophage genome sequences. With an estimated 10^{31} bacteriophages existing in the environment (Suttle, 2005; Whitman *et al.*, 1998), current knowledge is limited regarding the identification and characterization of those phages able to utilize *Salmonella* as hosts. Here, we report the microbiological properties and bioinformatics analysis of a new *Salmonella* bacteriophage, vB_SenS-Ent1.

RESULTS

Isolation of vB_SenS-Ent1

The bacteriophage vB_SenS-Ent1, isolated from swine lagoon effluent, was named according to the recommendations outlined by Kropinski *et al.* (2009a) with the additional mnemonic Ent to denote the host *Salmonella* Enteritidis serotype. Ent1 propagates efficiently on *Salmonella* Enteritidis phage type 4 (PT4) and, upon plating in 0.6 % (w/v) agarose, produces relatively large plaques 5.69 ± 1.72 mm in diameter ($n=10$). The morphology of plaques may be termed as 'bull's-eye', but are best described as consisting of at least three discrete zones.

Plaque centres exhibited clearing but contained a number of small colonies. Moving outwards, the second zone consisted of a halo of reduced turbidity, extending to the plaque boundary with the bacterial lawn. A third zone, consisting of a ring of increased turbidity relative to the surrounding bacterial lawn, was observed occasionally, just beyond the plaque boundary.

Virion morphology

Examination by transmission electron microscopy (TEM) revealed Ent1 as a typical member of the family *Siphoviridiae* of dsDNA bacteriophages (Fig. 1), similar to the Jersey morphotype (Ackermann, 2007). The phage particle consists of an icosahedral head of 64 nm mean diameter (n=20) and flexible, non-contractile tail (approx. 116 × 8.5 nm). The tail appears to have transverse striations and short fibres can be distinguished at the terminal end.

Adsorption and one-step growth

The infection process of Ent1 upon co-incubation with *S.* Enteritidis PT4 was investigated by standard adsorption and one-step growth assays. At 37 °C, Ent1 exhibited rapid adsorption, at 6.73×10^{-9} ml min^{-1} (R^2=0.992), to cells of *S.* Enteritidis PT4. Once adsorbed, vB_SenS-ENT1 exhibited a 20 min eclipse phase with host-cell lysis occurring after 35 min, releasing a mean of 35 new progeny particles per cell.

Host range and efficiency of plating (EOP)

Host range was assessed by the ability of Ent1 to form plaques upon a library of *Salmonella* consisting of 31 different isolates of *S. enterica* and panels of nine and eight for serovars Typhimurium and Enteritidis, respectively (Table S1, available in JGV Online). Ent1 exhibited plaque formation against 25 of 48 (52 %) *Salmonella* isolates when spot plates were prepared using log-fold dilutions of bacteriophage. Sensitive isolates were taken forward to ascertain the relative EOP; this is a more rigorous analysis of the ability of Ent1 to form plaques, as phage particles are distributed throughout the overlay agar rather than presented as a concentrated focal 'spot'. Hence, the formation of plaques results from productive infection of a bacterium by at least a single phage particle rather than by non-productive infection or lysis from without. The ability of Ent1 to form plaques under EOP assay conditions differed markedly from spot plates and the number of sensitive isolates was reduced to 14 (20.8 %), with values of relative EOP ranging from 5.67×10^{2} to 3.94×10^{-9}.

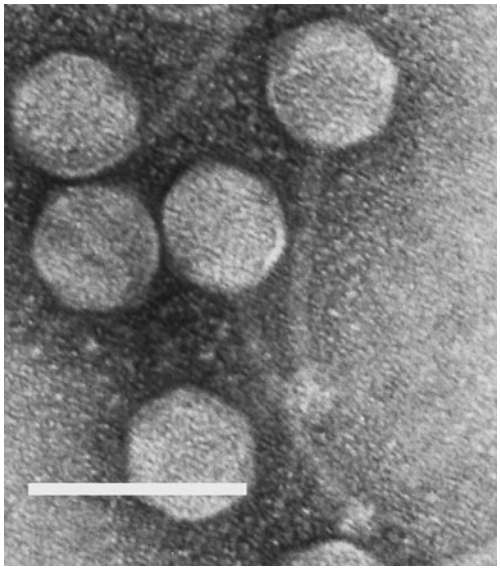

Fig. 1. Transmission electron micrograph of vB_SenS-Ent1. Bar, 100 nm.

With the exceptions of *S.* Enteritidis CBRI 1944, which was resistant to infection, and a reduced EOP of 4.04×10^{-7} for *S.* Enteritidis PT13a, Ent1 infected all *S.* Enteritidis isolates at similar or greater efficiencies to the propagating strain. For *S.* Typhimurium, Ent1 was able to form plaques on five of the nine isolates tested. Efficient infection relative to the host strain was observed for two Typhimurium isolates, CBRI 1960 and CBRI 1962.

From the panel of serovars tested, no pattern of serotype somatic and flagella antigen sensitivity to plaque formation by Ent1 could be identified. It was noted that Ent1 formed plaques at low efficiency upon the rough strain *S. enterica* e,h: e,n,x,z15 [Veterinary Laboratories Agency (VLA) S07538-07] but did not plaque upon another rough isolate, *S. enterica* z10 (VLA S07520-07). No lysis was observed when Ent1 was co-incubated with representatives of four other Gram-negative bacterial species (Table S1).

Genome properties

A single consensus sequence at 30-fold coverage was obtained by 454 pyrosequencing. The vB_SenS-Ent1 particle encapsulates a linear dsDNA molecule 42 391 bp in length with a mean G+C content of 49.79 mol%, lower than the mean of 52 mol% reported for serovars of *S. enterica* (McClelland *et al.*, 2001; Reen *et al.*, 2005; Thomson *et al.*, 2008). The genome was opened upstream of the putative small terminase subunit. A total of 58 ORFs, accounting for 92.8 % of the sequence, were predicted (Fig. 2), and the genome may be structured into four clusters on the basis of transcriptional direction. The genes are tightly spaced at a density of 1.37 kb^{-1} and a mean length of 678 bp. Short overlaps between the stop codon of one gene and the start codon of the adjacent downstream gene are common. No tRNA genes were discovered with either tRNAscan-SE or ARAGORN. As for many bacteriophages, only a limited number of protein functions could be predicted by sequence similarity and the presence of conserved domains (Tables S2 and S3). As such, 56 % of the ORFs were annotated as

hypothetical and, of these, two did not have a related bacterial, phage or prophage match in the current databases.

Genome clustering

BLASTN and BLASTX database searches using the entire Ent1 consensus sequence returned the highest nucleotide identity and protein similarity for members of the family *Siphoviridae* infecting *Salmonella*, phages SE2, SETP3 and KS7, and *Escherichia coli*, phages K1H, K1G, K1ind1 and K1ind3. Pairwise nucleotide sequence alignment of Ent1 with these members of the family *Siphoviridae* using MAUVE revealed that these phages all share a similar genomic organization. Notably, all possess division of the genes involved in head and tail structure by a regulatory gene cluster encoded in the opposite orientation.

Dot-plot analysis of 33 tailed phages reported to infect *Salmonella*, including the prophages Gifsy-1, Gifsy-2, ST64T and ST64B, revealed a high degree of genetic diversity at the nucleotide sequence level, but that sufficient similarity exists between members to allow tentative relational clustering (Fig. 3) (Hatfull *et al.*, 2008).

Whilst synteny of gene order is often preserved within arrays of genes, particularly those encoding structural and assembly proteins in the family *Siphoviridae*, bacteriophage genomes are strikingly mosaic (Hatfull, 2008). Individual genes and gene segments are exchangeable by horizontal transfer with members of the population by a number of potential mechanisms (Hatfull & Hendrix, 2011). Where similarities in nucleotide sequence are no longer apparent, the comparison of protein sequences provides a useful tool to delineate common ancestries. Protein sequence similarity has been used to produce the phage proteomic tree (Rohwer & Edwards, 2002), aid taxonomic classification of the families *Podoviridae* and *Myoviridae* (Lavigne *et al.*, 2008, 2009) and designate *Mycobacteria* phage proteins into 'phamilies' (Cresawn *et al.*, 2011). The number of orthologous proteins encoded by the Ent1, SETP3, KS7 and SE2 genomes were identified using CoreGenes 3.0, employing the default BLASTP threshold score of 75 (Zafar *et al.*, 2002). Ent1 shares 49, 48 and 50 protein homologues with SETP3, KS7 and SE2, respectively, with a total of 40 homologues shared between the four genomes.

Genome architecture

The genome structure of Ent1 exhibits a modular organization, with four gene clusters representing two early and two late transcriptional regions for which different functions may be assigned. Late clusters encompassed genes encoding proteins involved in packaging (gp01–gp03), morphogenesis (gp05–gp19) and lysis (gp51 and gp52). With the exception of gp42, all early ORFs with inferred DNA replication (gp32–gp45) or regulatory (gp20–gp24) functions are encoded on the complementary

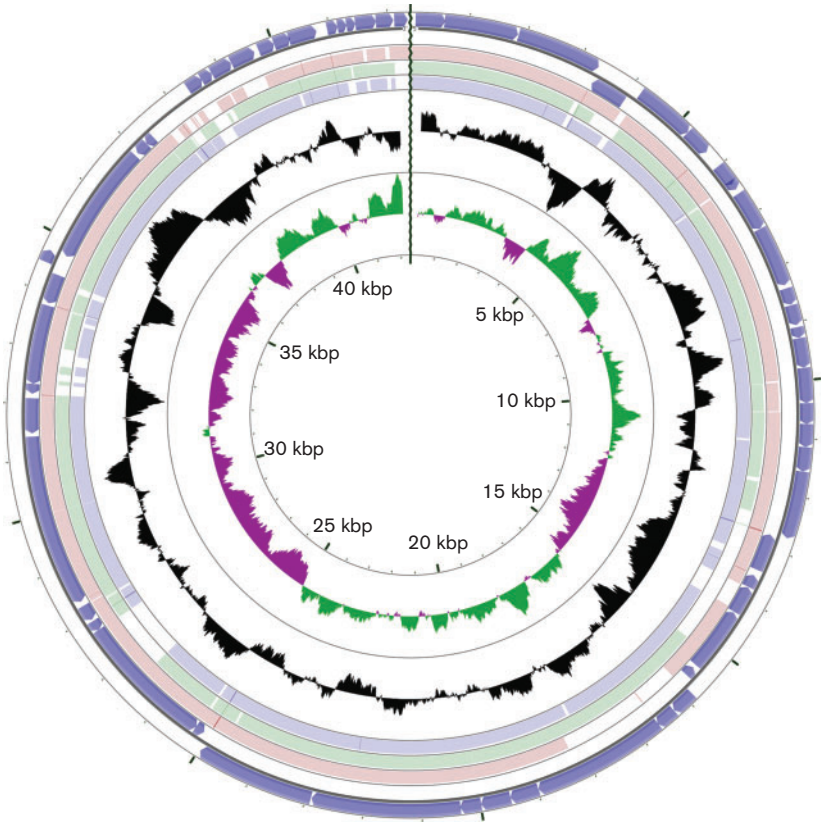

Fig. 2. Circular map of the vB_SenS-Ent1 genome prepared using CGView. The outer ring denotes the vB_SenS-Ent1 genome and ORFs on the positive strand. The next ring illustrates ORFs on the complementary strand. Opaque inner rings correspond to BLASTN identity to *Salmonella* phages SETP3 (red), KS7 (green) and SE2 (purple), respectively. The inner rings show G+C content and G+C skew, where peaks represent the positive (outward) and negative (inward) deviation from the mean G+C content and G+C skew, respectively.

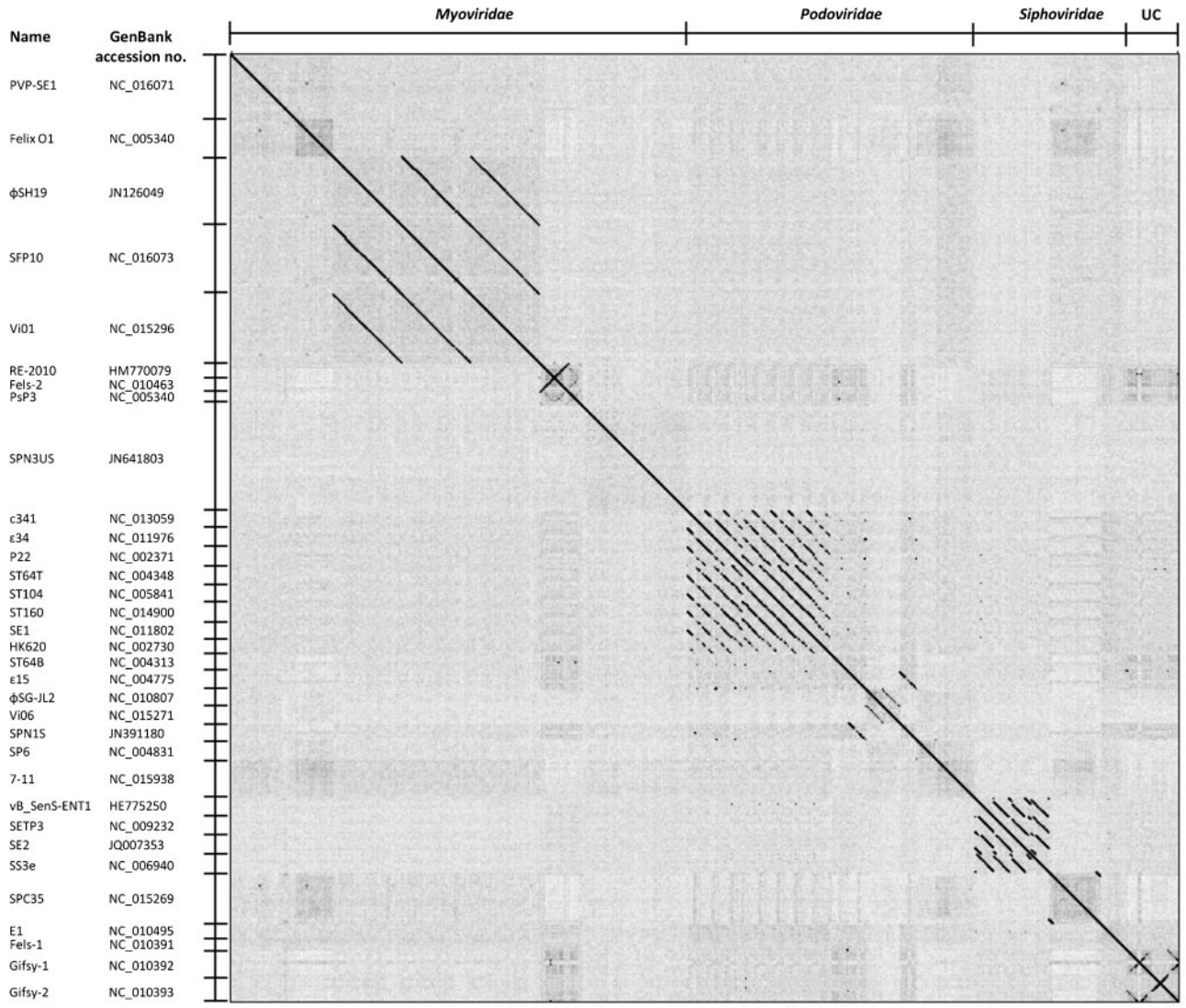

Fig. 3. Clustering of 33 *Salmonella* bacteriophage genomes by taxonomic family and nucleotide sequence similarity. Dot-plots were generated using Gepard. The KS7 genome was reverse-complemented to conform to the convention of presenting structural genes on the positive (rightward) strand. UC, Unclassified prophages.

strand. A relatively large non-coding region of 957 bp is located between gp45 and gp46 and this region contains the minimum G+C skew at position 37843.

Packaging, morphogenesis and structural proteins

Nearly half (48 %) of the Ent1 virion genome is devoted to encoding proteins involved in packaging, structure and morphogenesis. These genes comprise two gene clusters totalling 27 ORFs, separated by an immunity region spanning genes gp20–gp24. The arrangement of genes encoding virion structural and morphogenesis proteins generally follows a conserved organization in the family *Siphoviridae* (Casjens, 2005; Hatfull, 2008) and have been characterized in detail for *E. coli* phages λ and T5, *Bacillus subtilis* phage SPP1 and *Lactococcus* phages TP901-1 and p2.

One-dimensional (1D) SDS-PAGE of CsCl-purified virions yielded 10 bands consisting of four major bands at 72, 42, 38 and 13.5 kDa and five minor bands of 90, 56, 20, 18 and 9 kDa (Fig. 4). Four further proteins were identified on 2D SDS-PAGE gels, and genes corresponding to spots were annotated on the basis of predicted molecular mass, isoelectric point and protein homologues from BLASTP and PSI-BLAST queries.

Gene gp01 was assigned as the small terminase subunit due to its position immediately upstream of the large terminase subunit, gp02. BLASTP results showed limited protein sequence identity (37 %) to gp01 of *Sodalis* phage SO-1,

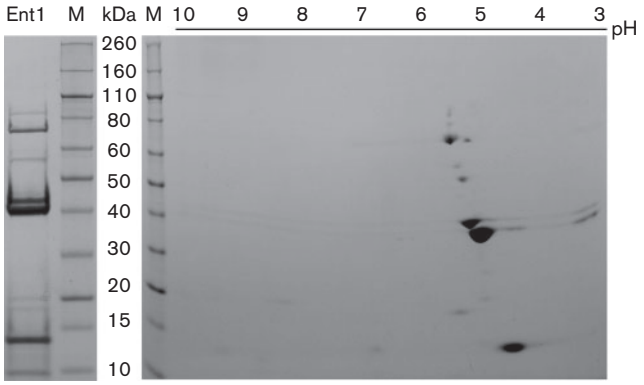

Fig. 4. SDS-PAGE of vB_SenS-ENT1 structural proteins. Phage structural proteins (Ent1) and standard marker (M) resolved on 4–12% Bis-Tris 1D SDS-PAGE and 2D SDS-PAGE gels (linear pH 3.0 to 10.0).

whilst PSI-BLAST analysis yielded hits to small terminase subunits of phage infecting both Gram-positive and Gram-negative bacteria. A terminase 6 family domain (PF03237) spanning residues 24–406 was identified by Pfam analysis of gp02. gp03 was identified as the putative portal protein due to its position immediately downstream of the large terminase subunit. Supporting this assignment, a spot of 58 kDa, slightly larger than the predicted mass of 54 kDa, was observed on 1D and 2D SDS gels.

gp05 was predicted to belong to the phage Mu protein F-like family (PF04233). A member of this Pfam family is the phage SPP1 gp07, a non-essential minor protein participating in capsid assembly and DNA packaging (Dröge *et al.*, 2000). The presence of an immunoglobulin-like I-set domain (PF13895) in gp06 suggests that this gene encodes a head-decoration protein. Head-decoration proteins have been described for phages L, λ and ES18 and are thought to aid stabilization of the capsid structure against disruption by chelating agents (Gilcrease *et al.*, 2005). Immunoglobulin-like domains are found widely within the order *Caudovirales* and are predominantly associated with structural proteins. The precise role of such domains remains unclear, but it has been suggested that they facilitate weak, non-specific binding to the cell surface or act to stabilize the virion structure (Fraser *et al.*, 2007).

No putative function could be assigned to either gp07 or gp08. However, gp08 has multiple potential start sites, which would yield significant overlap with gp07, and gene-calling routines showed disagreement as to the correct start codon. A scaffold protein with a predicted mass of 25.7 kDa is encoded by gp09 immediately upstream of the major coat protein (MCP) gp10. The MCP appears as a strong band/spot of 38 kDa on SDS gels.

PSI-BLAST results suggest that gp11 is a distant homologue of the gp8.5 head fibre of *Bacillus* phage PZA and Phi-29, respectively. A second I-set domain [InterPro records (IPR)007110] is found in gp12 and PSI-BLAST hits included Hoc and Wac proteins. The proximity to the MCP suggests that both these proteins are involved in head completion.

The six ORFs (gp13–gp18) located downstream of gp12 probably constitute head-to-tail joining and tail-scaffold proteins and, of these, gp15, gp16 and gp17 are candidates for spots on 2D SDS-PAGE gels. Two ORFs, gp14 and gp16, can be linked by PSI-BLAST searches to proteins found in mature virions of the *Pseudomonas aeruginosa* phage vB_PaeS-Kakheti25 (Karumidze *et al.*, 2012). No ORFs encoding candidate ejection proteins such as those reported for P22 and SPP1 were identified through database searches. gp19 was assigned as the major tail protein on the basis of PSI-BLAST results and presence as a strong band of 41 kDa on SDS gels.

Genes comprising the second gene cluster encode the structural and assembly proteins forming the virion tail. We were unable to identify a candidate slippery sequence associated with a programmed translational frameshift within coding regions of gp25 and gp26 (Xu *et al.*, 2004). gp27 was assigned as the tail tape measure protein and a strong spot at 83 kDa corresponding to the predicted size of this gene product was observed on 2D gels. The length of the tape measure protein agreed precisely with the measured length of the virion tail from TEM images, assuming a ratio of 0.15 nm per amino acid residue as reported for λ (Katsura & Hendrix, 1984) and, like other tape measure proteins, gp27 is predicted to form a predominantly helical structure. Three ORFs separate gp27 from the putative tail fibre gene, gp31. Referring to the organization of the structural cassette in SPP1, P22 and phages infecting *Lactococcus* species, it is plausible that these genes encode the distal tail baseplate (Casjens & Thuman-Commike, 2011; Mc Grath *et al.*, 2006; Veesler & Cambillau, 2011). PSI-BLAST analysis of gp31 revealed similarities to the GpJ host specificity protein of λ and p33 of phage T1. A tailspike, gp32, was identified with confidence due to the presence of a complete Pfam P22 tailspike domain (PF09251), indicating that adsorption of Ent1 to host cells involves recognition of carbohydrate moieties of the LPS O-antigen. BLASTN and BLASTP analyses revealed significant homologues in the families *Siphoviridae*: SE2, KS7, SETP3, SETP5, SETP7, SETP12 and SETP13; and *Podoviridae*: P22, SE1, ST104, ST64T and SETP1, 14 and 15. A CLUSTAL W alignment of gp32 revealed an N-terminal sequence similar to other members of the *Siphoviridae* infecting *Salmonella*: SE2, KS7 and SETP phages 3, 5, 7 and 12, but retaining significant conservation in the catalytic and C-terminal regions relative to P22. N-terminal sequences of P22-like tailspikes are involved in attachment to the virion tail structure and exhibit conservation between related phages within the families *Podoviridae*, *Siphoviridae* and *Myoviridae* (Hooton *et al.*, 2011). Spots and bands corresponding to the predicted mass of the putative tape measure, tail fibre and tailspike were apparent on 1D and 2D SDS gels.

Regulatory proteins

Phage genomes usually encode a number of regulatory proteins that, in concert with host proteins, co-ordinate the expression of early and late genes. Ent1 possesses a regulatory module consisting of five early genes encoded on the complementary strand, gp20–gp24, dividing the array of genes responsible for virion morphogenesis. gp20 showed strong sequence similarity to DNA-binding proteins from a number of different phage and bacteria and contains matches to two Pfam family domains: ANT (PF03374) and pRha (PF09669). In P22 *ant* encodes an anti-repressor that inhibits binding of the *c2* repressor to the P_L and P_R operators, enabling the expression of genes necessary for lytic development (Byl & Kropinski, 2000). The pRha domain represents a family of proteins whose expression is detrimental for lytic growth in the absence of integration host-factor function (Henthorn & Friedman, 1995). gp22 is predicted to encode a free-standing HNH endonuclease (PF13392) and PSI-BLAST analysis of gp23 revealed similarities to endonuclease recombination subunits. gp24 shares extended similarity to the immunity (Imm) proteins effecting exclusion of superinfecting phage (Lu *et al.*, 1993). BLASTP homologues included phages JS98 (Zuber *et al.*, 2007), vB_EcoM-VR7 (Kaliniene *et al.*, 2011) and IME08 (Jiang *et al.*, 2011). Supporting the protein homologue evidence, gp24 is predicted to localize at the cytoplasmic membrane (PSortb) and to contain two transmembrane domains (TMHMM).

In addition to the regulatory gene cluster, two ORFs, gp04 and gp42, interrupt the sequence of the structural and replication gene modules, in each instance encoded in the opposite orientation to the transcriptional direction of the respective module. The first separates the structural genes gp03 and gp05. ORF gp04 exhibits negative G+C skew in relation to neighbouring coding sequences, suggesting a historical horizontal acquisition event. Protein homologues detected by PSI-BLAST suggest that gp05 encodes a KilA-like protein. gp42 interrupts the sequence of the replicative gene cluster and is predicted to act as a transcriptional regulator, possibly involved in control of late gene expression, due to the presence of an N-terminal helix–turn–helix DNA-binding domain similar to cro/cI (PF12844).

The vB_SenS-Ent1 DNA replication module contains mobile elements

Ent1 encodes 12 genes co-localized as a distinct module and comprising a primase, helicase, DNA polymerase, a restriction endonuclease, two putative DNA-binding proteins, a uvsX homologue and six hypothetical products. Sequence similarity to other phage-encoded helicases suggested the presence of an in-frame insertion containing a DOD homing endonuclease motif within the sequence of gp34. The insertion sequence was identified as a large intein of 348 aa, satisfying the four criteria outlined by Perler *et al.* (1997). Inteins are defined as internal protein elements that self-excise from their host protein and catalyse ligation of the flanking sequences with a peptide bond, yielding two stable proteins: the mature protein and the intein. Searches conducted using BLASTP and the mature helicase resulted in significantly improved alignments to other bacteriophages. Notably, an intein is also present at an identical position within the helicase of the closely related SETP3 phage. Similarly, the DNA polymerase A encoded by gp37 contains a minimal intein, lacking an endonuclease domain, of 299 aa consistent with those predicted for the DNA polymerase gene of SETP3, SETP5 and SETP12. To date, 290 inteins are documented within the InBase database for Eubacteria, of which 36 are associated with phages or prophages (Perler, 2002).

gp43 encodes a putative primase containing an AAA_28 Pfam domain (PF13481), suggesting that, like *Enterobacteria* phages T4 and T5, Ent1 employs a primosome complex consisting of separate primase and helicase proteins (Ilyina *et al.*, 1992). A DNA-binding protein containing a helix–turn–helix 17 domain (PF12728) with low identity to transcriptional regulators of bacterial and phage origins is encoded by gp45.

Host lysis

Eubacterial phages employ a dual protein system comprising a holin and lysin to effect host-cell lysis (Young *et al.*, 2000). gp52 encodes a lysin with inferred glycoside hydrolyase activity (PF00959) and is immediately preceded by gp51, a protein of 95 aa predicted to contain two transmembrane regions, features consistent with class II holins (Young, 2002). Following the dual lysin system, with the exception of gp56 which encodes a putative ninH-like protein, five ORFs, gp53–58, could not be assigned any putative function.

Promoters and terminators

Analysis of 150 bp regions upstream of ORFs using MEME yielded a single candidate motif of 47 bp width (E value$=9.4\times10^{-8}$) containing regions resembling -10 and -35 elements. The regulatory sequence analysis tools (RSAT) set was used to convert the MEME output file to a position matrix and to search upstream regions for additional instances. Seven locations for the consensus motif were found at a mean distance of 48 bases from the start codon, suggesting a role in transcriptional regulation (Table S4). No candidate sequences were identified upstream of cluster starts gp04, gp05, gp25 or gp45 and the absence of promoter motifs in these areas suggests that additional sequences involved in transcriptional regulation remain to be found.

Seventeen intergenic sequences representing potential rho-independent transcription terminators were identified (Table S5). Candidate terminators were assessed according to location, presence of a U-rich tail and stable predicted stem–loop secondary structure ($\Delta G \leqslant -10$ kcal mol^{-1}) as

calculated by RNAFold (Gruber *et al.*, 2008). Seven of these are predicted to be on early transcripts and ten on late transcripts. Of these, eight form four bidirectional terminators separating the 3′ junctions of opposing transcriptional gene clusters.

Physical genome ends

Heat treatment of restriction fragments followed by either rapid or slow cooling did not alter restriction patterns (data not shown), excluding the possibility of cohesive genome ends. Time-limited treatment with the exonuclease *Bal*31 resulted in even, simultaneous degradation of all restriction fragments (Fig. S1). These data suggest that the Ent1 genome is terminally redundant and circularly permuted (Loessner *et al.*, 2000). Circularly permuted genomes are characteristic of head-full packaging strategy, where the packaged DNA length is between 102 and 110 % of the total genome length and could account for the 1.5 kb discrepancy between the genome size as estimated by pulsed-field gel electrophoresis (PFGE) and DNA sequencing.

DISCUSSION

The similarity in morphology, genome size, nucleotide identity and number of shared orthologous proteins provides weight to the designation of vB_SenS-Ent1 as a novel SETP3-like siphovirus. Along with KS7 and SE2, these bacteriophages remain unclassified in the *Ninth Report of the International Committee on Taxonomy of Viruses* (King *et al.*, 2011).

The (relatively) broad specificity exhibited by Ent1 for different *Salmonella* serovars is attributed primarily to the presence of the P22-like tailspike. Defining the true specificity of bacteriophages infecting *Salmonella* is no small task; the host genus is serologically complex, with over 2500 serovars described in the latest addition of the White–Kauffmann–Le Minor scheme.

As yet, it is not evident whether Ent1 is capable of lysogeny. No gene products exhibited similarity to characterized integrases, resolvases or excisionases and preliminary results suggest that the growth of small colonies within plaques arises due to host resistance rather than by a phage-related immunity mechanism. In addition, the related family *Siphoviridae* KS7, SETP3 and K1 phages are reported to be lytic (Bull *et al.*, 2010; De Lappe *et al.*, 2009). In contrast, PHACTS analysis (McNair *et al.*, 2012) denotes that the Ent1 lifestyle is temperate, but the result is marginal, with only 0.51 of trees in the forest algorithm yielding this decision. Without absolute certainty as to whether Ent1 possesses a temperate life cycle, the use of this phage as a biocontrol or therapeutic agent *ex vitro* is precluded, a necessary condition due to the association of temperate phage with the horizontal transfer of bacterial virulence factors (Canchaya *et al.*, 2003; Figueroa-Bossi *et al.*, 2001). Notwithstanding, the establishment of an efficient and productive lytic infection cycle depends upon the function and interaction of multiple early and late proteins with both self- and host-encoded proteins (reviewed by Roucourt & Lavigne, 2009). As such, and like other bacteriophage genomes, Ent1 may harbour multiple gene products that have potential for exploitation as antimicrobial agents (Liu *et al.*, 2004). Our current understanding of phage genomes, and the molecular mechanisms of the infective cycle, is constrained by the sheer number of predicted proteins for which no putative function has yet been elucidated. Concerted efforts to unravel familial and ancestral relationships between bacteriophage proteins are under way, particularly represented by PHANTOME, ACLAME (Leplae *et al.*, 2010) and Phamerator, supplemented by an ever-increasing number of available sequences.

METHODS

Bacterial strains. A complete listing of the bacterial strains used and their relevant culture collection accession numbers are available in Table S1. Where a particular isolate has been serotyped, the accession number refers to the reference number designated by the serotyping laboratory: the Veterinary Laboratories Agency (VLA), UK.

Isolation and purification. Bacteriophage vB_SenS-Ent1 was isolated by incubation of clarified (0.2 μm-filtered) swine effluent, diluted 10-fold in lysogeny broth (LB; Sigma-Aldrich) supplemented with 10 mM $MgSO_4$ and 1 mM $CaCl_2$, at 37 °C for 18 h in the presence of *Salmonella* Enteritidis (VLA S07544-07). After incubation, samples were assessed for the presence of bacteriophages using the soft-agar overlay plaque assay (Kropinski *et al.*, 2009b). In brief, samples were filtered (0.45 μm pore size), serially diluted and 10 μl aliquots were added to 150 μl of an exponential-phase culture ($OD_{540\ nm}$ approx. 0.1–0.2) and mixed in 4 ml LB overlay agar [0.6 % (w/v) agarose]. This mixture was poured onto the surface of 90 mm LB agar plates and allowed to dry prior to overnight incubation at 37 °C. Plaques were picked and subjected to a further three rounds of plating to ensure homogeneity, after which individual plaques were excised and suspended in SM buffer (50 mM Tris-Cl, 100 mM NaCl, 8 mM $MgSO_4$, pH 7.5) for 24 h at 4 °C to elute the bacteriophages.

Standard methods were used to obtain high-titre stocks of bacteriophages. Plaque suspensions were titrated and used to inoculate 1 l early exponential-phase cultures (OD_{540} 0.1) at an approximate m.o.i. of 0.1. Growth and lysis of bacterial cultures was monitored by hourly measurements of A_{540}. Chloroform (1 %, v/v) was added to each flask to terminate enrichment and prevent further microbial growth. Lysates were treated with DNase I and RNase A (Sigma-Aldrich) at 1 μg ml^{-1} prior to removal of bacterial debris by centrifugation at 11 000 ***g*** for 10 min at 4 °C. The supernatant was filtered (3 μm/0.2 μm combination pore size) and bacteriophage particles purified by precipitation with polyethylene glycol followed by two-step caesium chloride (CsCl) density-gradient centrifugation (Sambrook & Russell, 2001). Visual bands containing bacteriophage particles were recovered using a 21 gauge syringe and CsCl was removed by diafiltration against two 500-fold volume changes of SM buffer. Purified stocks were stored in SM buffer at 4 °C and titrated by standard double agar overlay plates.

Electron microscopy. Electron microscopy of bacteriophages was performed using methods described by Ackermann (2009). Bacteriophages were negatively stained with 1 % (w/v) uranyl acetate solution (pH 4.5; Sigma-Aldrich) and examined using a Phillips

CM10 transmission electron microscope operated at 60 kV. Magnification was calibrated using T4 tails and dimensions established by measurement of 20 intact particles. Bacteriophages were assigned to a respective family based upon examination of particle morphology (King *et al.*, 2011).

Adsorption rate constant and one-step growth. Determination of adsorption rate constants followed the method described by Kropinski (2009). Briefly, host strains were grown in LB to an OD_{540} of 0.1 (approx. 5×10^7 c.f.u. ml^{-1}) and enumerated using a spiral plating system (Don Whitely Scientific). Bacteriophages were added to cultures to yield a final concentration of 5×10^4 p.f.u. ml^{-1} (t=0). At 1 min intervals, 50 µl was transferred to 950 µl LB saturated with chloroform and stored on ice. Samples were titrated for unabsorbed bacteriophages using agar overlay plates. Absorption rate constants (k) were calculated as $-m/N$, where m represents the slope of linear regression of the natural logarithm of the free phage titre over time and N the initial bacterial density in c.f.u. ml^{-1}.

Host range and efficiency of plating (EOP). Bacteriophage host range and EOP were determined by the standard double agar layer plating technique (Kutter, 2009). Preparations of bacteriophages were adjusted by dilution to yield a titre of 10^{11} p.f.u. ml^{-1} on their respective propagating host. Square 120 mm plates containing LB agar were subdivided to yield a 6×6 grid, to which was added 8 ml LB overlay agar containing 200 µl from an exponential-phase culture of the bacterial strain to be tested. To each section of the grid, 5 µl aliquots of bacteriophage from a 10-fold dilution series were spotted onto the surface of the overlay agar so that each plate assessed a range of phage concentrations. Phage samples were allowed to absorb into the overlay agar prior to overnight incubation at 37 °C. Plates were visually examined for plaque formation and assigned a score (Fig. 5). The relative EOP was calculated as the titre in p.f.u. ml^{-1} for a given isolate divided by the titre for the relevant propagating host from triplicate overlay plaque assays.

Estimation of phage genome size. PFGE was performed according to the method described by Lingohr *et al.* (2009). Bacteriophages, approximately 10^{10} p.f.u. ml^{-1}, were immobilized in agar plugs [1 % (w/v) agarose] and lysed by treatment with proteinase K [New England Biolabs; 20 mg ml^{-1} in 10 mM Tris pH 8.0, 50 mM EDTA, and 1 % (w/v) SDS] for 2 h at 54 °C. Agarose plugs were subsequently washed three times by soaking in TE (Tris-EDTA) buffer (Sigma-Aldrich) for 1 h. Gels [1 % (w/v) agarose] were run in 0.5× TBE buffer (Bio-Rad) at 6 V cm^{-1} for 15 h at 14 °C with pulses of 2.2–54.2 s using a CHEF-DR II electrophoresis unit (Bio-Rad). DNA size standards were provided by use of low-range PFGE marker DNA (New England Biolabs) and bands were visualized by staining with 1 µg ethidium bromide ml^{-1}. Gel images were acquired using a FluorChem Q (ProteinSimple) and analysed using ImageJ (Abramoff *et al.*, 2004).

SDS-PAGE. Structural proteins from CsCl-purified virions were extracted and concentrated using methanol–chloroform and resuspended in lithium dodecyl sulphate (LDS) sample buffer (Invitrogen). For isoelectric focusing, linear immobilized pH gradient (IPG) ZOOM strips from pH 3 to 10 were rehydrated overnight with the solubilized proteins and separated using a ZOOM IPGRunner system according to the manufacturer's instructions (Invitrogen). Prior to SDS-PAGE, IPG strips were equilibrated twice for 15 min in 1× LDS sample buffer containing 1× reducing agent for the first equilibration step and 125 mM iodoacetamide for the second step. Protein separation was conducted alongside Novex Sharp unstained protein standard marker using a NuPAGE mini-gel system and 4–12 % Bis-Tris gels in MES–SDS running buffer (Invitrogen) at 200 V. Proteins were fixed [50 % (v/v) methanol, 10 % (v/v) acetic acid] for 1 h at room temperature, then stained using SimplyBlue Safestain (Invitrogen) and images of gels were acquired and analysed as described for PFGE.

Genome sequencing and bioinformatics. Phenol–chloroform-extracted bacteriophage genomic DNA was sequenced externally (Eurofins MWG Operon) using GS FLX454 sequencing. *Ab initio* prediction of ORFs used a combination of the GeneMark.hmm algorithm for prokaryotes (Besemer & Borodovsky, 1999), Glimmer 3.02 (Delcher *et al.*, 1999) and RAST (Aziz *et al.*, 2008). Overlapping 10 kb sections of the sequenced genome were queried against the non-redundant protein databases using TBLASTX for intrinsic prediction of ORFs. The results of evidence-based and *ab initio* gene predictions were compared in order to evaluate best gene models and, where appropriate, to resolve overlaps among ORFs. All ORFs were inspected for the presence of convincing purine-rich ribosome-binding sites upstream of the start codon. Annotation was performed using Artemis (Rutherford *et al.*, 2000) and physical maps were prepared using CGView (Stothard & Wishart, 2005). Translated sequences from predicted ORFs were queried using BLASTP and PSI-BLAST against the non-redundant database (Altschul *et al.*, 1990, 1997). Functional annotation of gene products was performed by querying translated sequences against the conserved domain database (Marchler-Bauer *et al.*, 2011), Prosite (Sigrist *et al.*, 2010), Pfam (Punta *et al.*, 2012) and InterProScan (Quevillon *et al.*, 2005). Translated ORFs were characterized by number of amino acids, molecular mass and isoelectric point using the ExPASy tool: Compute pI/Mw (Bjellqvist *et al.*, 1993). Prediction of transmembrane helices was performed using TMHMM 2.0 (Krogh *et al.*, 2001) and searches for structural motifs and signal peptides were carried out using COILS, SignalP (Petersen *et al.*, 2011) and PSIPRED (Jones, 1999; Buchan *et al.*, 2010). tRNAscan-SE (Schattner *et al.*, 2005) and ARAGORN (Laslett & Canback, 2004) were used to scan for tRNAs. Putative rho-independent terminators were predicted using TransTermHP (Kingsford *et al.*, 2007) and WebGester (Mitra *et al.*, 2011). Candidate promoter sequences were identified by searches of 150 bp sequences upstream of ORFs using MEME (Bailey *et al.*,

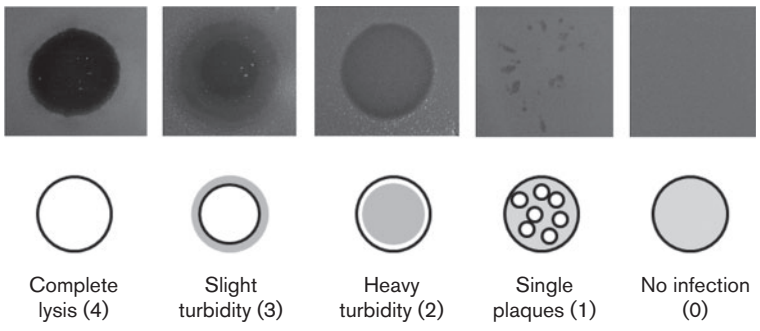

Fig. 5. Score card for the visual assessment of plaques on spot plates. The system is adapted from that described by Rees & Dodd (2006). Photographs of plaques formed by vB_SenS-Ent1 are provided for illustrative purposes.

2006) and the regulatory sequence analysis tools (RSAT) set (Thomas-Chollier *et al.*, 2011). Nucleotide and protein comparisons between *Salmonella* phage genomes were performed using Gepard (Krumsiek *et al.*, 2007), MAUVE (Darling *et al.*, 2010) and CoreGenes (Zafar *et al.*, 2002), using default parameters.

ACKNOWLEDGEMENTS

The authors wish to thank Dr David Patton for assistance with electron microscopy and Dr Helen Lovell for technical advice relating to PFGE. We also thank Dr Robin Thorn and Dr Gareth Robinson for the critical review of draft manuscripts and Dr Rebecca Green for permission to use strains from Campden BRI culture collection. This work was carried out by D. T. in partial fulfilment of the requirements for a PhD from the University of the West of England, Bristol, UK, and was funded by a Higher Education Funding Council for England bursary. The authors declare no competing interests.

REFERENCES

Abramoff, M. D., Magalhaes, P. J. & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics Int* **11**, 36–42.

Acheson, D. & Hohmann, E. L (2001). Nontyphoidal salmonellosis. *Clin Infect Dis* **32**, 263–269.

Ackermann, H.-W. (2007). *Salmonella* phages examined in the electron microscope. *Methods Mol Biol* **394**, 213–234.

Ackermann, H.-W. (2009). Basic phage electron microscopy. *Methods Mol Biol* **501**, 113–126.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. & other authors (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75.

Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34** (web server issue), W369–W373.

Besemer, J. & Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* **27**, 3911–3920.

Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., Frutiger, S. & Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14**, 1023–1031.

Buchan, D. W., Ward, S. M., Lobley, A. E., Nugent, T. C., Bryson, K. & Jones, D. T. (2010). Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38** (Suppl.), W563–W568.

Bull, J. J., Vimr, E. R. & Molineux, I. J. (2010). A tale of tails: sialidase is key to success in a model of phage therapy against K1-capsulated *Escherichia coli. Virology* **398**, 79–86.

Byl, C. V. & Kropinski, A. M. (2000). Sequence of the genome of *Salmonella* bacteriophage P22. *J Bacteriol* **182**, 6472–6481.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L. & Brüssow, H. (2003). Phage as agents of lateral gene transfer. *Curr Opin Microbiol* **6**, 417–424.

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

Casjens, S. R. & Thuman-Commike, P. A. (2011). Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology* **411**, 393–415.

Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W. & Hatfull, G. F. (2011). Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* **12**, 395.

Darling, A. E., Mau, B. & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147.

De Lappe, N., Doran, G., O'Connor, J., O'Hare, C. & Cormican, M. (2009). Characterization of bacteriophages used in the *Salmonella enterica* serovar Enteritidis phage-typing scheme. *J Med Microbiol* **58**, 86–93.

Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**, 4636–4641.

Dröge, A., Santos, M. A., Stiege, A. C., Alonso, J. C., Lurz, R., Trautner, T. A. & Tavares, P. (2000). Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. *J Mol Biol* **296**, 117–132.

ECDC (2010). *Annual Epidemiological Report on Communicable Diseases in Europe 2010*. Stockholm: European Centre for Disease Prevention and Control.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D. & Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella. Mol Microbiol* **39**, 260–271.

Fraser, J. S., Maxwell, K. L. & Davidson, A. R. (2007). Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Curr Opin Microbiol* **10**, 382–387.

Gilcrease, E. B., Winn-Stapley, D. A., Hewitt, F. C., Joss, L. & Casjens, S. R. (2005). Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. *J Bacteriol* **187**, 2050–2057.

Grimont, A. D. & Weill, F.-X. (2007). *Antigenic Formulae of the Salmonella Serovars, World Health Organisation Collaborating Centre for Reference and Research on Salmonella*, 9th edn. Paris; Pasteur Institute.

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Res* **36** (web server issue), W70–W74.

Hanning, I. B., Nutt, J. D. & Ricke, S. C. (2009). Salmonellosis outbreaks in the United States due to fresh produce: sources and potential intervention measures. *Foodborne Pathog Dis* **6**, 635–648.

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

Hatfull, G. F. & Hendrix, R. W. (2011). Bacteriophages and their genomes. *Curr Opin Virol* **1**, 298–303.

Hatfull, G. F., Cresawn, S. G. & Hendrix, R. W. (2008). Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. *Res Microbiol* **159**, 332–339.

Henthorn, K. S. & Friedman, D. I. (1995). Identification of related genes in phages phi 80 and P22 whose products are inhibitory for phage growth in *Escherichia coli* IHF mutants. *J Bacteriol* **177**, 3185–3190.

Hooton, S. P., Timms, A. R., Rowsell, J., Wilson, R. & Connerton, I. F. (2011). *Salmonella* Typhimurium-specific bacteriophage ΦSH19 and the origins of species specificity in the Vi01-like phage family. *Virol J* **8**, 498.

Ilyina, T. V., Gorbalenya, A. E. & Koonin, E. V. (1992). Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J Mol Evol* **34**, 351–357.

Jiang, X., Jiang, H., Li, C., Wang, S., Mi, Z., An, X., Chen, J. & Tong, Y. (2011). Sequence characteristics of T4-like bacteriophage IME08 benome termini revealed by high throughput sequencing. *Virol J* **8**, 194.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195–202.

Kaliniene, L., Klausa, V., Zajančkauskaite, A., Nivinskas, R. & Truncaite, L. (2011). Genome of low-temperature T4-related bacteriophage vB_EcoM-VR7. *Arch Virol* **156**, 1913–1916.

Karumidze, N., Thomas, J. A., Kvatadze, N., Goderdzishvili, M., Hakala, K. W., Weintraub, S. T., Alavidze, Z. & Hardies, S. C. (2012). Characterization of lytic *Pseudomonas aeruginosa* bacteriophages via biological properties and genomic sequences. *Appl Microbiol Biotechnol* **94**, 1609–1617.

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

King, A. M. Q., Adams, M. J., Carstens, E. B. & Lefkowitz, E. J. (editors) (2011). *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier/Academic Press.

Kingsford, C. L., Ayanbule, K. & Salzberg, S. L. (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* **8**, R22.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567–580.

Kropinski, A. M. (2009). Measurement of the rate of attachment of bacteriophage to cells. *Methods Mol Biol* **501**, 151–155.

Kropinski, A. M., Sulakvelidze, A., Konczy, P. & Poppe, C. (2007). *Salmonella* phages and prophages–genomics and practical aspects. *Methods Mol Biol* **394**, 133–175.

Kropinski, A. M., Prangishvili, D. & Lavigne, R. (2009a). Position paper: the creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. *Environ Microbiol* **11**, 2775–2777.

Kropinski, A. M., Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. (2009b). Enumeration of bacteriophages by double agar overlay plaque assay. *Methods Mol Biol* **501**, 69–76.

Krumsiek, J., Arnold, R. & Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**, 1026–1028.

Kutter, E. (2009). Phage host range and efficiency of plating. *Methods Mol Biol* **501**, 141–149.

Laslett, D. & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* **32**, 11–16.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W. & Kropinski, A. M. (2008). Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res Microbiol* **159**, 406–414.

Lavigne, R., Darius, P., Summer, E. J., Seto, D., Mahadevan, P., Nilsson, A. S., Ackermann, H. W. & Kropinski, A. M. (2009). Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC Microbiol* **9**, 224.

Leplae, R., Lima-Mendez, G. & Toussaint, A. (2010). ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* **38** (database issue), D57–D61.

Lingohr, E., Frost, S. & Johnson, R. P. (2009). Determination of bacteriophage genome size by pulsed-field gel electrophoresis. *Methods Mol Biol* **502**, 19–25.

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N. & other authors (2004). Antimicrobial drug discovery through bacteriophage genomics. *Nat Biotechnol* **22**, 185–191.

Loessner, M. J., Inman, R. B., Lauer, P. & Calendar, R. (2000). Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. *Mol Microbiol* **35**, 324–340.

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

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C. & other authors (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39** (database issue), D225–D229.

Mc Grath, S., Neve, H., Seegers, J. F. M. L., Eijlander, R., Vegge, C. S., Brøndsted, L., Heller, K. J., Fitzgerald, G. F., Vogensen, F. K. & van Sinderen, D. (2006). Anatomy of a lactococcal phage tail. *J Bacteriol* **188**, 3972–3982.

McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M. & other authors (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**, 852–856.

McNair, K., Bailey, B. A. & Edwards, R. A. (2012). PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinformatics* **28**, 614–618.

Mitra, A., Kesarwani, A. K., Pal, D. & Nagaraja, V. (2011). WebGeSTer DB – a transcription terminator database. *Nucleic Acids Res* **39** (database issue), D129–D135.

Perler, F. B. (2002). InBase: the intein database. *Nucleic Acids Res* **30**, 383–384.

Perler, F. B., Olsen, G. J. & Adam, E. (1997). Compilation and analysis of intein sequences. *Nucleic Acids Res* **25**, 1087–1093.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–786.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G. & other authors (2012). The Pfam protein families database. *Nucleic Acids Res* **40** (database issue), D290–D301.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. & Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res* **33** (web server issue), W116–W120.

Reen, F. J., Boyd, E. F., Porwollik, S., Murphy, B. P., Gilroy, D., Fanning, S. & McClelland, M. (2005). Genomic comparisons of *Salmonella enterica* serovar Dublin, Agona, and Typhimurium strains recently isolated from milk filters and bovine samples from Ireland, using a *Salmonella* microarray. *Appl Environ Microbiol* **71**, 1616–1625.

Rees, C. E. D. & Dodd, C. E. R. (2006). Phage for rapid detection and control of bacterial pathogens in food. *Adv Appl Microbiol* **59**, 159–186.

Rohwer, F. & Edwards, R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* **184**, 4529–4535.

Roucourt, B. & Lavigne, R. (2009). The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environ Microbiol* **11**, 2789–2805.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945.

Sambrook, J. & Russell, D. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd edn, vol. 1. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M.-A., Roy, S. L., Jones, J. L. & Griffin, P. M. (2011). Foodborne illness acquired in the United States – major pathogens. *Emerg Infect Dis* **17**, 7–15.

Schattner, P., Brooks, A. N. & Lowe, T. M. (2005). The tRNAscan-SE, snoscan and sno*GPS* web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33** (web server issue), W686–W689.

Sigrist, C. J. A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A. & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38** (database issue), D161–D166.

Stothard, P. & Wishart, D. S. (2005). Circular genome visualization and exploration using CGView. *Bioinformatics* **21**, 537–539.

Suttle, C. (2005). Crystal ball. The viriosphere: the greatest biological diversity on Earth and driver of global processes. *Environ Microbiol* **7**, 481–482.

Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D. & van Helden, J. (2011). RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res* **39** (web server issue), W86–W91.

Thomson, N. R., Clayton, D. J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., Quail, M. A., Stevens, M., Jones, M. A. & other authors (2008). Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**, 1624–1637.

Veesler, D. & Cambillau, C. (2011). A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol Mol Biol Rev* **75**, 423–433.

Voetsch, A. C., Van Gilder, T. J., Angulo, F. J., Farley, M. M., Shallow, S., Marcus, R., Cieslak, P. R., Deneen, V. C., Tauxe, R. V. & Emerging Infections Program FoodNet Working Group (2004). FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. *Clin Infect Dis* **38** (Suppl. 3), S127–S134.

Whitman, W. B., Coleman, D. C. & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**, 6578–6583.

Xu, J., Hendrix, R. W. & Duda, R. L. (2004). Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell* **16**, 11–21.

Young, R. (2002). Bacteriophage holins: deadly diversity. *J Mol Microbiol Biotechnol* **4**, 21–36.

Young, I., Wang, I.-N. & Roof, W. D. (2000). Phages will out: strategies of host cell lysis. *Trends Microbiol* **8**, 120–128.

Zafar, N., Mazumder, R. & Seto, D. (2002). CoreGenes: a computational tool for identifying and cataloging ''core'' genes in a set of small genomes. *BMC Bioinformatics* **3**, 12.

Zuber, S., Ngom-Bru, C., Barretto, C., Bruttin, A., Brüssow, H. & Denou, E. (2007). Genome analysis of phage JS98 defines a fourth major subgroup of T4-like phages in *Escherichia coli*. *J Bacteriol* **189**, 8206–8214.