The complete genome sequence of *Clostridium difficile* phage ϕ C2 and comparisons to ϕ CD119 and inducible prophages of CD630

Shan Goh,^{1†} Peh Fern Ong,¹ Keang Peng Song,^{1‡} Thomas V. Riley^{2,3} and Barbara J. Chang²

Correspondence
Barbara J. Chang
bchang@cyllene.uwa.edu.au

¹National University of Singapore, Yong Loo Lin School of Medicine, Department of Microbiology, 5 Science Drive 2, #05-03, 117597, Singapore

²Microbiology and Immunology, School of Biomedical, Biomolecular and Chemical Sciences, The University of Western Australia, Queen Elizabeth II Medical Centre, Nedlands, WA 6009, Australia

³Division of Microbiology and Infectious Diseases, The Western Australian Centre for Pathology and Medical Research, Nedlands, WA 6009, Australia

Received 13 September 2006
Revised 19 November 2006
Accepted 27 November 2006

The complete genomic sequence of a previously characterized temperate phage of *Clostridium difficile*, ϕ C2, is reported. The genome is 56 538 bp and organized into 84 putative ORFs in six functional modules. The head and tail structural proteins showed similarities to that of *C. difficile* phage ϕ CD119 and *Streptococcus pneumoniae* phage EJ-1, respectively. Homologues of structural and replication proteins were found in prophages 1 and 2 of the sequenced *C. difficile* CD630 genome. A putative holin appears unique to the *C. difficile* phages and was functional when expressed in *Escherichia coli*. Nucleotide sequence comparisons of ϕ C2 to ϕ CD119 and the CD630 prophage sequences showed relatedness between ϕ C2 and the prophages, but less so to ϕ CD119. ϕ C2 integrated into a gene encoding a putative transcriptional regulator of the *gntR* family. ϕ C2, ϕ CD119 and CD630 prophage 1 genomes had a *Cdu1-attP*-integrase arrangement, suggesting that the pathogenicity locus (PaLoc) of *C. difficile*, flanked by *cdu1*, has phage origins. The *attP* sequences of ϕ C2, ϕ CD119 and CD630 prophages were dissimilar. ϕ C2-related sequences were found in 84 % of 37 clinical *C. difficile* isolates and typed reference strains.

INTRODUCTION

Clostridium difficile has emerged as an important intestinal pathogen since the 1970s, continuing to plague hospital settings worldwide and causing recent epidemics in the USA (McDonald *et al.*, 2005) and Canada (Loo *et al.*, 2005; Warny *et al.*, 2005). The major virulence factors of *C. difficile* are toxins A and B encoded by *tcdA* and *tcdB* respectively, which are located on a 19 kb genomic region termed the pathogenicity locus (PaLoc) (Braun *et al.*, 1996; Hammond

& Johnson, 1995). The toxins are positively regulated by TcdR (Mani & Dupuy, 2001; Rupnik *et al.*, 2005) and are negatively regulated by TcdC (Hundsberger *et al.*, 1997; Matamouros *et al.*, 2006); they are encoded by *tcdR* and *tcdC* respectively, also on the PaLoc. Another toxin-associated gene, *tcdE*, appears phage related but its function is unknown (Tan *et al.*, 2001). *C. difficile* acquires antibiotic resistance and virulence genes through plasmids and transposons (Bruggemann, 2005) shown to significantly contribute to genome plasticity (Sebahia *et al.*, 2006). It is possible that phages also contribute to variance in virulence-associated genes (Lemee *et al.*, 2005) and to the emergence of outbreak strains (McDonald *et al.*, 2005). However, the prevalence of phage genes within *C. difficile* genomes is not known. In comparison to phages of *Escherichia coli*, *Staphylococcus aureus* and *Lactobacillus* species, the study of clostridial phages is in its infancy. Only one phage specific for *C. difficile*, temperate phage ϕ CD119, has been sequenced (Govind *et al.*, 2006), while two putative prophage sequences were detected in the recently sequenced genome of *C. difficile* CD630 (Sebahia *et al.*, 2006).

[†]Present address: Department of Cell and Molecular Biology, Programme for Genomics and Bioinformatics, Karolinska Institutet, Berzelius väg 35, SE-171 77 Stockholm, Sweden.

[‡]Present address: School of Arts and Sciences, Monash University Malaysia, No. 2 Jalan Kolej, Bandar Sunway, Petaling Jaya 46150, Selangor Darul Ehsan, Malaysia.

Abbreviation: PaLoc, pathogenicity locus.

The GenBank accession number for the sequence reported in this paper is DQ466086.

A supplementary table and figure are available with the online version of this paper.

Previously, we induced a temperate phage, ϕ C2, from a clinical *C. difficile* isolate. The phage was partially sequenced and characterized (Goh *et al.*, 2005b) and was shown to increase toxin B levels in lysogens (Goh *et al.*, 2005a). Two other induced temperate phages, ϕ C6 and ϕ C8, were also shown to have this effect in *C. difficile* lysogens (Goh *et al.*, 2005a). While *C. difficile* toxin production itself was not phage mediated, phages may have some other role in host physiology. In this study, we compared the ϕ C2 genome with the genomes of ϕ CD119 and *C. difficile* CD630 (Sebaihia *et al.*, 2006), as well as the unfinished sequence of *C. difficile* QCD-32g58. We found ϕ C2 genes to be prevalent in most of the clinical *C. difficile* isolates tested. Two phages, designated ϕ C630-1 and ϕ C630-2, were induced from CD630 in the course of this study.

METHODS

Bacterial strains, phage and growth conditions. Thirty-four clinical *C. difficile* strains from Singapore General Hospital, and Sir Charles Gairdner Hospital, Western Australia, were used in this study (Table 1). Reference strains used were CCUG 37782, CCUG 20309 and CCUG 16126, which were purchased from the Culture Collection, University of Göteborg, Sweden. VPI 10463 was generously provided by Dr M. Rupnik, University of Ljubljana, Slovenia, and CD630 was kindly provided by Dr P. Mullany, University College London, UK. For induction of temperate phage, an overnight culture of CD630 in Brain Heart Infusion broth (BHIB, Oxoid) supplemented with 5% horse blood, 50 μ g erythromycin ml^{-1} (Sigma) and 10 μ g tetracycline ml^{-1} (Sigma) was induced with 3 μ g mitomycin C ml^{-1} (Sigma) and incubated for another 8 h. The culture supernatant was filtered through a 2 μ m membrane (Pall), then assayed for phage against 10 randomly chosen clinical *C. difficile* strains (Table 1) as previously described (Goh *et al.*, 2005b). Propagation of ϕ C630-1 and ϕ C630-2 with a 4 h culture of CD843 was as previously described for ϕ C6 (Goh *et al.*, 2005b). ϕ C2 was propagated on CD062 as previously described (Goh *et al.*, 2005b). *C. difficile* strains were maintained in Cooked Meat Medium (Oxoid), from which 37 °C overnight cultures in BHIB were prepared. Exponential-phase cultures were prepared by subculturing 1 ml of an overnight culture in 9 ml BHIB and incubating at 37 °C for 4 h.

Molecular cloning and DNA sequencing and analysis. Phage purification by CsCl density gradient, DNA extraction by phenol/chloroform/isoamyl alcohol and purification using the Wizard DNA Clean-up System (Promega) was as previously described (Goh *et al.*, 2005b). Phage DNA was also extracted directly from crude lysate using the Qiagen Lambda Midi Kit, according to the manufacturer's instructions. ϕ C2 DNA was digested with *AccI*, *HincII*, *HindIII* or *XbaI* and cloned into pUC19. Positive clones were selected for by blue-white colony screening, PCR using M13 primers and restriction enzyme digestion of recombinant plasmids. Sequencing was performed by the dideoxy chain-termination method with an automated ABI Prism 3100 DNA Sequencer (Applied Biosystems). Primers were designed and sequences were assembled using the Lasergene version 5.05 software (DNASTAR). A mean coverage of $3.53 \times$ (293 sequencing runs) and a minimum of $2 \times$ coverage (at least once in each strand) were obtained from sequencing the phage library. Gaps between contigs of ϕ C2 were filled and ends of the genome were sequenced by primer walking using phage genomic DNA as template. Probable protein-encoding genes (ORFs) were predicted using GeneMark.hmm VIOLIN and GeneMark.hmm for prokaryotes version 2.4 programs (Besemer & Borodovsky, 1999).

Predicted ORFs were searched for similarity to proteins in databases by BLASTP (Altschul *et al.*, 1990). Nucleotide similarity between ϕ C2 and *C. difficile* CD630 or QCD-32g58 was detected by BLASTN (Altschul *et al.*, 1990) at http://www.sanger.ac.uk/Projects/C_difficile/ and <http://www.ncbi.nlm.nih.gov/genomes/geblast.cgi?gi=5410>, respectively. Alignments between two sequences were carried out with bl2seq (Tatusova & Madden, 1999) and the genome was searched for tRNA genes using the tRNAscan web server (Lowe & Eddy, 1997). Cumulative GC skew was carried out with GenSkew at <http://mips.gsf.de/services/analysis/genskew>, and a hydrophobicity plot was generated using Hydrophobicity grapher <http://athena.bioc.uvic.ca/techDoc/hydrophobicity/> using the Kyte–Doolittle scaling system (Kyte & Doolittle, 1982). Transmembrane regions and beta-turns in an ORF were predicted by TMPRED at http://www.ch.embnet.org/software/TMPRED_form.html and BTPRED (Shepherd *et al.*, 1999), respectively. Tandem Repeats Finder (Benson, 1999) was used to detect direct or inverted repeats in the genome.

Protein expression. Plasmid pQE-hol and pQE-AbiF contained ϕ C2 putative ORF 36 and ORF 37, respectively, cloned between the *Bam*HI and *Pst*I sites of the pQE-30 expression vector (Qiagen). ORF 36 was amplified with Vent polymerase (NEB) using *Hol*FBam (5'-CGCGGATCCATGGATAATTTAATAAG-3')/*Hol*RPst (5'-AAC-TGCAGTTACTTTTCACCATCCT-3') with cycling conditions of 95 °C for 4 min, 30 cycles of 95 °C for 30 s, 53 °C for 30 s, 72 °C for 1 min and 72 °C for 10 min. ORF37 was amplified with Vent polymerase (NEB) and *Abi*FBam (5'-CGCGGATCCATGGTTGAAG-TAAAAGA-3')/*Abi*FRPst (5'-AACTGCAGTTATTTAGCCAATATC-TC-3') primers with cycling conditions of 95 °C for 2 min, 30 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 2 min and 72 °C for 10 min. The PCR product was digested with *Bam*HI and *Pst*I and ligated into pQE-30 with complementary ends, then transformed into M15[pREP4] by electroporation (Bio-Rad Gene Pulser II). Recombinant pUC19 (NEB) plasmids harbouring ϕ C2 DNA inserts were grown at 37 °C in LB (Invitrogen) supplemented with 100 μ g ampicillin ml^{-1} (ICN Biomedicals), 62 ng X-Gal ml^{-1} (Bio-Rad) and 0.625 mM IPTG (Sigma). Recombinant pQE-30 (Qiagen) expression plasmids were grown at 37 °C in LB supplemented with 100 μ g ampicillin ml^{-1} and 25 μ g kanamycin ml^{-1} (Sigma). Protein expression in M15[pREP4] cells was induced with 1 mM IPTG when growth of culture at 37 °C with shaking at 250 r.p.m. had reached an OD₆₀₀ of 0.6.

Southern hybridization and dot blot. ϕ C2 genomic probe was prepared by DIG-labelling of *Hinc*II-digested phage DNA according to the DIG High Prime DNA Labelling and Detection Starter Kit 1 (Roche) instructions. *C. difficile* and phage DNA were digested by *Hinc*II or *Hind*III and *Xba*I, separated in 1% TBE agarose and transferred to a nylon membrane (Amersham) as previously described (Sambrook *et al.*, 1989). Hybridization was carried out at 37–39 °C for 16–20 h, followed by stringent washing and colour development of membranes as in the system manual. Dot blots were carried out by spotting 120–200 ng chromosomal DNA on nylon membranes, fixing the DNA by microwave on high power for 2.5 min (Angeletti *et al.*, 1995) followed by hybridization to ϕ C2 genomic probe, according to the DIG system manual (Roche).

CHEF electrophoresis. Undigested phage DNA (150–200 ng) was added to an equal volume of molten 1% low-melting-point Agarose (Sigma), then loaded into a 1% TBE Pulse Field Certified Agarose (Bio-Rad). Electrophoresis was carried out in 0.5 $\times$ TBE running buffer using the CHEF-DR II Pulse Field Electrophoresis System (Bio-Rad) at pulse times of 3–13 s, 200 V for 22 h.

SDS-PAGE and N-terminal sequencing. SDS-PAGE analysis of phage was carried out as previously described (Ford *et al.*, 1998) with modifications. ϕ C2 was purified through a CsCl density gradient as previously described (Goh *et al.*, 2005b); 20 μ l

Table 1. Bacterial strains, phages and plasmids

Bacterial strain(s), phage or plasmid	Relevant properties*	Purpose in study	Source/reference
CD3, 7, 13, 32, 43, 25, 28, 46, 56, 66	<i>tcdA</i> ⁺ <i>tcdB</i> ⁺	Dot blot	Singapore General Hospital (SGH), Singapore
CD1, 4, 33, 39, 40, 55, 57, 58	<i>tcdA</i> ⁻ <i>tcdB</i> ⁻		
CD12	<i>tcdA</i> ^{tr} <i>tcdB</i> ⁻		
CD19	<i>tcdA</i> ⁻ <i>tcdB</i> ⁺		
CD22, 70	<i>tcdA</i> ^{tr} <i>tcdB</i> ⁺		
CD630	<i>tcdA</i> ^{tr} <i>tcdB</i> ^{tr}	Phage induction	P. Mullany†
CD61, CD62	Serogroup X <i>tcdA</i> ⁺ <i>tcdB</i> ⁺ (toxintype 0)	Dot blot	M. Rupnik†
CD38	<i>tcdA</i> ^{tr} <i>tcdB</i> ⁺	Dot blot	Sir Charles Gardner Hospital (SCGH), Western Australia
CD60	<i>tcdA</i> ⁺ <i>tcdB</i> ⁺		
CD63	<i>tcdA</i> ⁺ <i>tcdB</i> ⁺		
CD371	<i>tcdA</i> ⁺ <i>tcdB</i> ⁺		
CD578	<i>tcdA</i> ⁻ <i>tcdB</i> ^{tr}		
CD242	<i>tcdA</i> ⁻ <i>tcdB</i> ⁻	φC2 induction, stable lysogen	
CD062	<i>tcdA</i> ^{tr} <i>tcdB</i> ⁻	φC2 propagation, dot blot	
CD80	<i>tcdA</i> ⁻ <i>tcdB</i> ⁻	Stable lysogen, dot blot	
CD843	<i>tcdA</i> ⁺ <i>tcdB</i> ^{tr}	Stable lysogen, dot blot	
CCUG 37782	Serogroup F, <i>tcdA</i> ⁻ <i>tcdB</i> ⁺ (toxintype VIII)	Dot blot	Barbut <i>et al.</i> (1993); Rupnik <i>et al.</i> (1998)
CCUG 20309	Serogroup A, <i>tcdA</i> ^{tr} <i>tcdB</i> ⁺ (toxintype X)		Haslam <i>et al.</i> (1986); Rupnik <i>et al.</i> (1998)
CCUG 19126 VPI 10463 <i>E. coli</i> JM109	Serogroup G, <i>tcdA</i> ⁺ <i>tcdB</i> ⁺ (toxintype 0) <i>endA1 recA1 gyrA96 thi-1 hsdR17(r_K⁻ m_K⁺) relA1 supE44 Δ(lac-proAB) [F' traD36 proAB lacI^qZΔM15]</i>	Cloning	Rupnik <i>et al.</i> (1998) Promega
<i>E. coli</i> M15[pREP4]	Nal ^s Str ^s Rif ^r Km ^r RecA ⁺ Uvr ⁺ Lon ⁺ <i>lac ara gal mtl, lac</i> repressor plasmid	Protein expression	Qiagen
pQE-30	Ap ^r expression vector, C-terminal His tag	Protein expression	Qiagen
pUC19	Ap ^r <i>lacZα</i> , cloning vector	Cloning	Promega

**tcdA*^{tr} indicates a truncated gene.

†P. Mullany, University College London, UK; M. Rupnik, Max Planck Institute of Biophysical Chemistry, Göttingen, Germany.

(10⁷ p.f.u. ml⁻¹) was repeatedly frozen in liquid nitrogen and thawed at 37 °C three times, heated at 75 °C for 4 min, mixed with 5 × sample buffer (0.255 M Tris pH 6.8, 50% v/v, glycerol, 5% w/v SDS, 0.05% w/v, bromophenol blue, 0.25 M DTT) and heated again at 95 °C for 5 min before electrophoresis. Electrophoresis was carried out as previously described (Laemmli, 1970) using the Mini-PROTEAN II cell (Bio-Rad) and 12% acrylamide gels, which were stained with BLUPRINT Fast-PAGE Stain (Gibco BRL). Proteins were electroblotted to PVDF membranes (Bio-Rad) as described by Sambrook *et al.* (1989) using cold tank blotting transfer buffer (25 mM Tris pH 8.3, 150 mM glycine, 20% v/v methanol). The PVDF membrane was stained with Coomassie blue R250 and two major bands were excised from the membrane. Ten amino acids from the N-terminus of each protein were determined by an automated sequencer (Applied Biosystems 477 Protein Sequencer).

Determination of the *attP* region. The *attP* region was predicted to be downstream of the integrase, and primer pair patt-1/int-1 was used to generate a 311 bp PCR product which was DIG labelled and used as an *attP* probe for hybridization at 39 °C. Primer sequences

are: patt-1, 5'-GTAAAGATGATGAAGTGGATGAAG-3'; int-1, 5'-GCATTTTACAATAATTTGCCACCG-3'. The cycling conditions were 95 °C for 5 min, 30 cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min and 72 °C for 10 min. Genomic DNA of CD242 (5 µg) was digested with *AccI*, *HincII* or *HindIII* and self-ligated with T4 DNA ligase. Inverse PCR (Ochman *et al.*, 1988) was carried out with divergent primer pairs patt-0/Cdu1-1 and int-3/int-5. Primer sequences were as follows: patt-0, 5'-CTGTGAATGTAG-TTGATTCATTC-3'; Cdu1-1, 5'-GGTTAGAGCGAAGGGAGTTT-3'; int-3, 5'-CAACCACTATGGACACATATTC-3'; int-5, 5'-CGCA-AGCAATGAAAATTAAAG-3'. The PCR products were sequenced with the same primers. To determine *attBP'*, an additional primer, gntR-F (5'-GGATTTAGAAGTAAATTCC-3') was used.

Electron microscopy. Carbon-coated copper Formvar grids (400 mesh, SPI supplies) were floated on 20 µl 0.01% poly-L-lysine for 30 s and excess liquid was wicked off with a filter paper. The grid was floated on 20 µl phage suspension (10⁷–10⁹ p.f.u. ml⁻¹) for 3 min and dried as above, then fixed with 20 µl 1% glutaraldehyde for 1 min and negatively stained with 1% phosphotungstic acid

pH 7.4 for 30–90 s. The grids were air-dried for 30 min and viewed under a Philips 2085 transmission electron microscope at 57 000–89 000 $\times$ magnification, operating at 100 kV. Dimensions of a minimum of four phage particles for each phage were measured, and results expressed as mean $\pm$ SD.

RESULTS

General features of ϕ C2 genome

The dsDNA linear genome was 59.7 kb as estimated by CHEF electrophoresis (Fig. 1a) while the unit genome length was 56 538 bp as determined by sequencing, indicating terminal redundancy of approximately 3.2 kb in the packaged genome. The G+C content was 28.72 mol%, which is identical to the recently sequenced ϕ CD119 (Govind *et al.*, 2006) and slightly lower than that of the *C. difficile* CD630 genome (29.06 %). A total of 84 putative ORFs were identified, of which seven (8 %) showed no homology to proteins in the NCBI and EBI databases; 37 ORFs (44 %) had sequence similarity to proteins with unknown functions, of which 13 ORFs (15 %) were found only in *C. difficile* phage/prophage genomes. A supplementary table detailing the predicted ORFs and similarities to other proteins is included with the online version of this paper. In general, the predicted ORFs in the left and right arms of the genome had sequence similarity to ϕ CD119 and CD630 prophages 1 and 2, while ORFs in the middle section were similar to various bacterial or phage proteins. No programmed frameshift signals for translation were detected. The proteins of ϕ C2 were separated by 12 % SDS-PAGE (Fig. 1b) and two major proteins of ϕ C2 were identified as products of ORF 7, a putative capsid, and ORF 13, a putative tail sheath with the N-terminal methionine removed as expected (Ben-Bassat *et al.*, 1987; Romero *et al.*, 2004). Cumulative GC skew analysis of the genome sequence, included as a supplementary figure with the online version of this paper, revealed a putative origin of

replication in the region of nt 37929 and a putative terminus of replication in the region of nt 30390.

Mosaic structure of ϕ C2 genome

The genome is organized into six modules starting with a DNA packaging module from the left end of the genome, which has been assigned an arbitrary start from the putative terminase (Fig. 2). There was consistently high homology to ORFs of CD630 prophage 1 and 2, while sequence similarity to ϕ CD119 was sporadic. Of interest within the tail assembly module is a cassette of genes found in phage EJ-1 of *Streptococcus pneumoniae* (Romero *et al.*, 2004). The cell lysis module consisting of ORFs 36, 37 and 38 was atypical in the relative positions of the putative endolysin and holin, which are usually adjacent to each other (Young, 1992). Putative holins of the *C. difficile* phages had high sequence similarity to one another and good alignment with holin of another phage, despite low sequence homology (Fig. 3). A hydrophobicity plot (not shown) of ORF 36 revealed transmembrane regions separated by β -turns and outside N- and C-termini typical of type II holins (Wang *et al.*, 2000). Constitutive expression of ORF 36 inhibited growth and induced expression of ORF 36 decreased turbidity in *E. coli*, while expression of ORF 37 did not have an effect on cell growth (Fig. 4). Similar results have been shown for other phage holins (Muyombwe *et al.*, 1999; Sheehan *et al.*, 1999) and for *C. difficile* *tcdE* (Tan *et al.*, 2001), which had good sequence alignment with holins (Fig. 3).

Surprisingly, ORFs 41 and 42 were homologous to ParA and ATPase/ParB of the *Leptospira biflexa* phage LE1 (Bourhy *et al.*, 2005). These enzymes, together with a centromere-like sequence, *parS*, are required for maintenance of DNA stability (Austin & Abeles, 1983; Yamaichi & Niki, 2000); therefore ORFs 41 and 42 could be part of the lysogeny module. Possible *parS* sequences in the form of short direct or inverted repeats (Dam & Gerdes, 1994; Gallie & Kado, 1987; Radnedge *et al.*, 1996) around ORFs 41 and 42 were not detected, as was the case for LE1 (Bourhy *et al.*, 2005). Sequence similarities of ORF 41 to TcdB (22 % of 279 amino acids, *E*-value 1.1) and ORF 46 to Cdu1 (29 % of 125 amino acids, *E*-value 4×10^{-12}) are noteworthy. The ϕ C2 lysogeny module is unusual in being extended compared to that of λ (Birge, 2000); there appear to be two sets of repressors (ORFs 49, 50, 58 and ORFs 51, 53, 61) and perhaps as a consequence, two antirepressors (ORFs 52 and 55). ORF 59 had low percentage sequence similarity to an excisionase, the first to be detected in *C. difficile* phages. Although not shown, homologues of ORFs 10–14, 16, 20–22 and 24–26 have also been detected in the *Clostridium* sp. strain OhILAs sequence with high percentage sequence similarity. The *attP* of ϕ C2 in CD242 was predicted to be between ORF 46 (transcriptional regulator/Cdu1 homologue) and ORF 47 (integrase), and was confirmed by Southern hybridization (Fig. 5a). The *attP* region was flanked by *AccI*, *HincII* and *HindIII* sites; digestion of lysogenic DNA with either enzyme was used for inverse PCR; however, only an *attPB'* PCR product was obtained (Fig. 5b, c). As the

Fig. 1. CHEF electrophoresis of undigested ϕ C2 DNA (a) and SDS-PAGE of phage proteins (b). (a) The genome of ϕ C2 was estimated to be 59.7 kb. M1, 5 kb molecular mass marker (Bio-Rad). (b) The major proteins were sequenced from the N-terminus for 12 residues. The sequences correspond to capsid and sheath proteins. M2, protein molecular mass markers (Fermentas).

Fig. 2. Genome of ϕ C2. Direction of transcription and predicted ORFs are represented by block arrows which are labelled with corresponding putative functional assignments. Organization of the genome is broadly categorized into six genetic modules. Nucleotide homology to prophage 1 and 2 in CD630 and corresponding positions are shown with green and yellow bars, respectively. Orange and blue bars represent regions of nucleotide similarity to ϕ CD119 and QCD-32g58, respectively. ORFs unique to ϕ C2 are in red, ORFs common to ϕ C2, ϕ CD119 and prophages in CD630 are in brown.

bacterial sequence of *attPB'* had high percentage similarity to a *gntR* transcriptional regulator in CD630 (92 % of 93 amino acids, E -value 1.4×10^{-38}) and QCD-32g58, a forward primer specific for *gntR* (*gntR*-F) and *int*-3 were used to determine the sequence of *attBP'* (Fig. 5c). The ϕ C2 attachment site sequence CTGTGAGAAAT is different from that of ϕ CD119 (TTTATATGTGTTAT), CD630 prophage 1 (TAAAGATGA) and prophage 2 (TCCACTAGG).

Interestingly, the translated 3' end of the integrated CD630 prophage 1 (nt 1143302–1143688) was similar to Cdu1 (36 % of 125 amino acids, E -value 3×10^{-11}) and *attP* of ϕ CD119 was 248 nt downstream of a Cdu1 homologue (Govind *et al.*, 2006).

ϕ CD119 and prophage 1 and 2 homologues were mostly found in the DNA replication, recombination and

Fig. 3. Multiple sequence alignment between the putative holin of ϕ C2 (ORF36) and putative holins of ϕ CD119, CD630 prophage 1 (ORF CD0971), CD630 prophage 2 (ORF CD2894A), *Bacillus clarkii* phage BCJA1c and TcdE of *Clostridium difficile* 88864. Identical amino acids are shaded in black while conserved substitutions are shaded in grey. Predicted transmembrane (TM), hydrophobic and hydrophilic regions are indicated, as well as their predicted topology in italics.

modification module of the ϕ C2 genome. ORF 65, common to the *C. difficile* phages, is likely an essential recombination function (Erf) protein that is a member of a superfamily of single-strand annealing proteins involved in phage genome circularization via homologous recombination following DNA entry (Iyer *et al.*, 2002). Also common was ORF 78,

with high sequence similarity to RusA, an enzyme thought to have phage origins (Sharples *et al.*, 2002). RusA is a DNA endonuclease that resolves Holliday junctions in DNA replication, recombination and repair (Mahdi *et al.*, 1996).

Relatedness of ϕ C2, ϕ C630-1, ϕ C630-2 and ϕ CD119 and prevalence of prophage genes in *C. difficile* isolates

Mitomycin C induction of CD630 resulted in two plaque types which corresponded to phage particles of the same morphology but having slightly different head sizes. Since the genome of prophage 1 (55 850 bp) in CD630 is larger than prophage 2 (49 178 bp), it is likely that prophage 1 produced larger particles (ϕ C630-1) compared to prophage 2 (ϕ C630-2). ϕ C630-1 particles measured 31.7 ± 0.7 nm in head diameter and 62.4 ± 5.1 nm in tail length, while ϕ C630-2 particle dimensions were 28.1 ± 1.3 nm (head) and 39.5 ± 5.8 nm (tail) (Fig. 6). BLASTN identified regions of nucleotide sequence similarity throughout the ϕ C2 genome to regions of CD630 that indicated the ϕ C630-1 and ϕ C630-2 prophages have similar genome organization (Fig. 2). Dot plots (not shown) of ϕ C2 and the hypothetical genome sequences of ϕ C630-1 and ϕ C630-2 were virtually identical. Hence, common frameshift regions in ϕ C2 were detected at nt 26290–26996, nt 32933–33469, nt 47350–50177 (ϕ C630-1) and nt 42546–53713 (ϕ C630-2). Nucleotide comparison of ϕ C2 and ϕ CD119 genomes by dot plot (not shown) revealed some regions of similarity (Fig. 2) but were a less

Fig. 4. Expression of putative phage holin gene and *tcdE* of *C. difficile* in M15[pREP4] cells was induced with 1 mM IPTG. OD₆₀₀ of the cultures were measured over time. A decrease in OD₆₀₀ was observed in the pQE-hol culture (●) up to 4 h post-induction. A similar trend was observed for *tcdE* expression (pQE-tcdE, ▲). Expression of another phage gene (*abiF*) did not result in cell growth retardation (pQE-abiF, ◆).

Fig. 5. Integration of ϕ C2 DNA into the *C. difficile* genome. (a) Southern hybridization of *attP* probe to phage and lysogenic DNA digested with *HincII*. Although cross-hybridizations of the *attP* probe to other fragments of ϕ C2 were observed, the brightest signal was a 2.4 kb band, which was absent in lysogen CD242. Two new observed bands of 1.1 kb and 5.7 kb may contain the *attPB'* and *attBP'*. M, DIG-labelled/*HindIII* marker (Roche). (b) ϕ C2 DNA before integration, where the *attPP'* sequence (bold) was between ORF46 (black arrow) and ORF47 (grey arrow). The *attPP'* aligned with a homologous *attBB'* sequence (bold) in a gene encoding a *gntR*-like transcriptional regulator, present in the host genome. Divergent PCR primers patt-0/Cdu1-1 and int-3/int-5 were used to amplify and sequence the *attPB'* and *attBP'*, respectively. Primers patt-1/int-1 were used to amplify the *attPP'* region to use as a probe for Southern hybridization in (a). (c) Right side (*attPB'*) of integrated phage DNA in the lysogenic *C. difficile* genome by inverse PCR and left side (*attBP'*) by PCR with gntR-F and int-3 primers. ϕ C2 integration resulted in a disrupted *gntR*. *C. difficile* sequences are in lower case; ϕ C2 sequences are in upper case; the homologous *att* site is in bold.

similar pair compared to ϕ C2/ ϕ C630-1 and ϕ C2/ ϕ C630-2. Dot-plot comparisons of ϕ CD119 to ϕ C630-1 and ϕ C630-2 revealed sequence similarity in parts of the cell lysis, lysogeny control and DNA replication and modification modules of the three phages.

To evaluate the prevalence of ϕ C2-related genes in clinical *C. difficile* strains, dot blots and Southern blots were carried out with a ϕ C2 genomic probe. The results showed that out of 37 strains tested (Table 1), only six (CD13, CD57, CD62, CD062, CD843 and CCUG 37782) did not contain prophage genes; 31 strains including four typed reference strains exhibited homology to varying degrees. Differences

in dot-blot signals were further analysed by Southern hybridization (not shown). The most common homologous bands found in almost all strains correlated to ORFs 7–13 of the head structural module and ORFs 78–84 of the DNA replication, recombination and modification module, respectively. Regions representing modules of lysogeny control, lysis, head and DNA methylase were also found in some strains. To determine whether ϕ C2-related phage genes were present in another sequenced *C. difficile* strain, QCD-32g58, a BLASTN search was carried out. Only three regions of high sequence similarity were found between QCD-32g58 and ϕ C2 (Fig. 2). However, a whole prophage genome was not detected in QCD-32g58.

Fig. 6. Particle morphology of CD630 phages. (a) ϕ C630-1, (b) ϕ C630-2 with contracted sheath. Bars, 50 nm.

DISCUSSION

In the last few years there has been an exponential increase in sequenced phage genomes made available for comparison. This has resulted in the currently accepted view of divergent phages being related by virtue of genetic modules (Hendrix, 2002; Juhala *et al.*, 2000) and co-evolution of bacterial hosts and phage (Kwan *et al.*, 2005; Pedulla *et al.*, 2003). An aim of this study was to compare the genome of our previously partially characterized phage ϕ C2 to other phage genomes to provide an insight into phages of *C. difficile*.

The genome organization of ϕ C2 is typical of phages with a lysogenic cycle (Canchaya *et al.*, 2003). Its genome length was previously underestimated by addition of restriction fragments separated under normal electrophoresis (Goh *et al.*, 2005b). It was previously shown not to possess cohesive ends (Goh *et al.*, 2005b) and genome sequencing revealed it to have terminally redundant ends. The packaging mechanism of a terminase may be predicted from the large subunit amino acid sequence (Casjens *et al.*, 2005); the ϕ C2 terminase large subunit had high percentage sequence similarity to SPP1 (26 % of 426 amino acids, *E*-value 3×10^{-25}) of the P22-like headful subgroup. This suggests the ends of ϕ C2 were likely to be generated by a headful packaging mechanism. Based on the number of protein homologues and unique hypothetical proteins, ϕ C2 is closely related to the other *C. difficile* phages, ϕ CD119, ϕ C630-1 and ϕ C630-2, and demonstrates modular mosaicism (Casjens, 2005). Gene divergence appears greatest within the lysogeny control module, followed by tail structural proteins, which is common for tailed temperate phages as a method of diversifying infectable hosts (Casjens, 2005). Tail structural components of ϕ C2 may have been derived from an ancestral phage of EJ-1, while a putative LysM within tail-associated proteins may indicate lytic enzymes are used for local cell wall degradation and hence penetration of host wall for injection of phage DNA, similar to tailspike proteins of P22, Sf6 (Freiberg *et al.*, 2003) and T4

(Kanamaru *et al.*, 2005) and tail fibre proteins of anti-K1 phages (Muhlenhoff *et al.*, 2003). Homologues of extra-chromosomal replicative proteins ParA and ParB found close to the lysogeny module and their relative direction of transcription suggest their expression is associated with lysogenic conversion. Interesting possibilities include ϕ C2 switching to an LE1-like replicative prophage state for stability and perhaps having a pseudolysogeny phase (i.e. genome does not integrate into host chromosome but rather exists as a circular intermediate), or that it was once a replicative prophage. Low percentage sequence homology between ORF 41 and TcdB suggests phage origins of the toxin and may explain the genetic variability in *tcdB*, which has been observed in *C. difficile* toxinotypes (Rupnik *et al.*, 2001), more commonly in *tcdB* than *tcdA* (Rupnik *et al.*, 1998). Sequence similarity of ORF 46 to Cdu1, which borders the PaLoc, and the holin having a similar effect on *E. coli* as TcdE also point toward the PaLoc as a collection of genes transferred by phages, which have evolved to become the current virulence genes of *C. difficile*. ORF 46 (Cdu1 homologue) and its downstream non-coding region appear to be involved in integration because the Cdu1-*attP*-integrase arrangement is conserved in ϕ C2, ϕ C630-1 and ϕ CD119. The *attPs* of ϕ C2 and ϕ CD119 are 197 nt and 248 nt downstream of a Cdu1 homologue (Govind *et al.*, 2006), respectively, while *cdu1* contains an *attP* for ϕ C630-1 (Sebaihia *et al.*, 2006). There does not appear to be a preferred site for the integration of this group of phages in *C. difficile*, as the attachment sites are different for each phage. The presence of an *AbiF* protein on the phage genome is unusual; it is normally carried on a plasmid and confers phage resistance to bacteria, resulting in an abortive phage infection at the level of phage DNA replication (Garvey *et al.*, 1995). The advantage of carrying *abiF* is unknown and has not been found on other phage genomes; one possibility could be to prevent superinfection of C2 lysogens by unrelated phages that are susceptible to *AbiF*.

Comparative DNA sequence analysis of ϕ C2, ϕ C630-1, ϕ C630-2 and ϕ CD119 showed the degree of pairwise relatedness to be ϕ C630-1/ ϕ C2 > ϕ C2/ ϕ C630-2 > ϕ C630-1/ ϕ CD119 > ϕ CD119/ ϕ C630-2 > ϕ CD119/ ϕ C2. This indicates that ϕ C630-1 and ϕ C630-2 are intermediates of ϕ C2 and ϕ CD119 in the evolutionary sense. The brief sampling of randomly chosen clinical *C. difficile* isolates revealed that ϕ C2-related prophage genes are prevalent in *C. difficile*. Although these may represent only phage remnants, the apparently low prevalence of inducible phage (Goh *et al.*, 2005b) may simply be due to the lack of appropriate indicator strains. Isolates possessing both *tcdA* and *tcdB* (e.g. CD61) were as likely to be devoid of ϕ C2-related genes as isolates possessing truncated versions of *tcdA* and/or *tcdB* (e.g. CCUG 20309, CD843), or lacking either or both toxin genes (CCUG 37782, CD55). In general, there was no correlation between the presence of *tcdA* and *tcdB* and prevalence of ϕ C2-related prophage genes in *C. difficile* strains. Therefore, the current role of ϕ C2 is not in generating genetic diversity within the PaLoc but perhaps

in other areas of the host genome related to virulence. The following ORFs may affect host fitness: ORF 19 (putative TerD), ORF 9 (sequence similarity to the 'Alkaliphilus metalliredigenes' sigma-54 interaction region, 28 % of 114 amino acids, *E*-value 0.054) and ORF 15 (low sequence similarity to *Clostridium thermocellum* S-layer protein). ϕ C2 disruption of a *gntR*-like transcriptional regulator potentially affects expression of host genes, which may lead to altered fitness. In CD630, *gntR* was upstream of genes encoding the mannose-specific phosphotransferase system (PTS) (Sebahia *et al.*, 2006). The mannose PTS is involved in sugar transport and global regulation of gene expression, in a number of Gram-positive genera (Abranches *et al.*, 2003; Arous *et al.*, 2004; Chaillou *et al.*, 2001; Reizer *et al.*, 1999), including the regulation of energy metabolism and virulence genes in *Streptococcus mutans* (Abranches *et al.*, 2006). Hypothetically, integration of ϕ C2 into CD630 could lead to significant changes in the *C. difficile* phenotype through mannose PTS deregulation. The contribution of ϕ C2 and related temperate phages to the physiology of *C. difficile* and their potential roles in gene transfer and as genetic tools for this species are worthy of further investigation.

ACKNOWLEDGEMENTS

We thank Mark Schreiber for bioinformatics advice and support throughout the course of this work and Josephine Howe for help with electron microscopy.

REFERENCES

- Abranches, J., Chen, Y. Y. & Burne, R. A. (2003). Characterization of *Streptococcus mutans* strains deficient in EIIAB Man of the sugar phosphotransferase system. *Appl Environ Microbiol* **69**, 4760–4769.
- Abranches, J., Candella, M. M., Wen, Z. T., Baker, H. V. & Burne, R. A. (2006). Different roles of EIIABMan and EIIGlc in regulation of energy metabolism, biofilm development, and competence in *Streptococcus mutans*. *J Bacteriol* **188**, 3748–3756.
- 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.
- Angeletti, B., Battiloro, E., Pascale, E. & D'Ambrosio, E. (1995). Southern and Northern blot fixing by microwave oven. *Nucleic Acids Res* **23**, 879–880.
- Arous, S., Dalet, K. & Hechard, Y. (2004). Involvement of the *mpo* operon in resistance to class IIa bacteriocins in *Listeria monocytogenes*. *FEMS Microbiol Lett* **238**, 37–41.
- Austin, S. & Abeles, A. (1983). Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. *J Mol Biol* **169**, 373–387.
- Barbut, F., Mario, N., Delmee, M., Gozian, J. & Petit, J. C. (1993). Genomic fingerprinting of *Clostridium difficile* isolates by using a random amplified polymorphic DNA (RAPD) assay. *FEMS Microbiol Lett* **114**, 161–166.
- Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo, K., Boosman, A. & Chang, S. (1987). Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J Bacteriol* **169**, 751–757.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580.
- Besemer, J. & Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* **27**, 3911–3920.
- Birge, E. A. (2000). *Bacterial and Bacteriophage Genetics*, 4th edn. New York: Springer.
- Bourhy, P., Frangeul, L., Couve, E., Glaser, P., Saint Girons, I. & Picardeau, M. (2005). Complete nucleotide sequence of the LE1 prophage from the spirochete *Leptospira biflexa* and characterization of its replication and partition functions. *J Bacteriol* **187**, 3931–3940.
- Braun, V., Hundsberger, T., Leukel, P., Sauerborn, M. & von Eichel-Streiber, C. (1996). Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. *Gene* **181**, 29–38.
- Bruggemann, H. (2005). Genomics of clostridial pathogens: implication of extrachromosomal elements in pathogenicity. *Curr Opin Microbiol* **8**, 601–605.
- Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brussow, H. (2003). Prophage genomics. *Microbiol Mol Biol Rev* **67**, 238–276.
- Casjens, S. R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* **8**, 451–458.
- Casjens, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schicklmaier, P., Schmieger, H., Pedulla, M. L., Ford, M. E., Houtz, J. M., Hatfull, G. F. & Hendrix, R. W. (2005). The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol* **187**, 1091–1104.
- Chaillou, S., Postma, P. W. & Pouwels, P. H. (2001). Contribution of the phosphoenolpyruvate:mannose phosphotransferase system to carbon catabolite repression in *Lactobacillus pentosus*. *Microbiology* **147**, 671–679.
- Dam, M. & Gerdes, K. (1994). Partitioning of plasmid R1. Ten direct repeats flanking the *parA* promoter constitute a centromere-like partition site *parC*, that expresses incompatibility. *J Mol Biol* **236**, 1289–1298.
- Ford, M. E., Stenstrom, C., Hendrix, R. W. & Hatfull, G. F. (1998). Mycobacteriophage TM4: genome structure and gene expression. *Tuber Lung Dis* **79**, 63–73.
- Freiberg, A., Morona, R., Van den Bosch, L., Jung, C., Behlke, J., Carlin, N., Seckler, R. & Baxa, U. (2003). The tailspike protein of *Shigella* phage Sf6. A structural homolog of *Salmonella* phage P22 tailspike protein without sequence similarity in the beta-helix domain. *J Biol Chem* **278**, 1542–1548.
- Gallie, D. R. & Kado, C. I. (1987). *Agrobacterium tumefaciens* pTAR *parA* promoter region involved in autoregulation, incompatibility and plasmid partitioning. *J Mol Biol* **193**, 465–478.
- Garvey, P., Fitzgerald, G. F. & Hill, C. (1995). Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Appl Environ Microbiol* **61**, 4321–4328.
- Goh, S., Chang, B. J. & Riley, T. V. (2005a). Effect of phage infection on toxin production by *Clostridium difficile*. *J Med Microbiol* **54**, 129–135.
- Goh, S., Riley, T. V. & Chang, B. J. (2005b). Isolation and characterization of temperate bacteriophages of *Clostridium difficile*. *Appl Environ Microbiol* **71**, 1079–1083.
- Govind, R., Fralick, J. A. & Rolfe, R. D. (2006). Genomic organization and molecular characterization of *Clostridium difficile* bacteriophage ϕ CD119. *J Bacteriol* **188**, 2568–2577.
- Hammond, G. A. & Johnson, J. L. (1995). The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microb Pathog* **19**, 203–213.
- Haslam, S. C., Ketley, J. M., Mitchell, T. J., Stephen, J., Burdon, D. W. & Candy, D. C. (1986). Growth of *Clostridium difficile* and production of toxins A and B in complex and defined media. *J Med Microbiol* **21**, 293–297.

- Hendrix, R. W. (2002). Bacteriophages: evolution of the majority. *Theor Popul Biol* **61**, 471–480.
- Hundsberger, T., Braun, V., Weidmann, M., Leukel, P., Sauerborn, M. & von Eichel-Streiber, C. (1997). Transcription analysis of the genes *tcdA-E* of the pathogenicity locus of *Clostridium difficile*. *Eur J Biochem* **244**, 735–742.
- Iyer, L. M., Koonin, E. V. & Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**, 8.
- Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F. & Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* **299**, 27–51.
- Kanamaru, S., Ishiwata, Y., Suzuki, T., Rossmann, M. G. & Arisaka, F. (2005). Control of bacteriophage T4 tail lysozyme activity during the infection process. *J Mol Biol* **346**, 1013–1020.
- Kwan, T., Liu, J., DuBow, M., Gros, P. & Pelletier, J. (2005). The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci U S A* **102**, 5174–5179.
- Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Lemee, L., Bourgeois, I., Ruffin, E., Collignon, A., Lemeland, J. F. & Pons, J. L. (2005). Multilocus sequence analysis and comparative evolution of virulence-associated genes and housekeeping genes of *Clostridium difficile*. *Microbiology* **151**, 3171–3180.
- Loo, V. G., Poirier, L., Miller, M. A., Oughton, M., Libman, M. D., Michaud, S., Bourgault, A. M., Nguyen, T., Frenette, C. & other authors (2005). A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med* **353**, 2442–2449.
- Lowe, T. M. & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964.
- Mahdi, A. A., Sharples, G. J., Mandal, T. N. & Lloyd, R. G. (1996). Holliday junction resolvases encoded by homologous *rusA* genes in *Escherichia coli* K-12 and phage 82. *J Mol Biol* **257**, 561–573.
- Mani, N. & Dupuy, B. (2001). Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc Natl Acad Sci U S A* **98**, 5844–5849.
- Matamouros, S., Govind, R. & Dupuy, B. (2006). TcdC inhibits toxin synthesis in *Clostridium difficile*. In *5th International Meeting on the Molecular Biology and Pathogenesis of the Clostridia*, p. 64. Nottingham, UK.
- McDonald, L. C., Killgore, G. E., Thompson, A., Owens, R. C., Jr, Kazakova, S. V., Sambol, S. P., Johnson, S. & Gerding, D. N. (2005). An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N Engl J Med* **353**, 2433–2441.
- Muhlenhoff, M., Stummeyer, K., Grove, M., Sauerborn, M. & Gerardy-Schahn, R. (2003). Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. *J Biol Chem* **278**, 12634–12644.
- Muyombwe, A., Tanji, Y. & Unno, H. (1999). Cloning and expression of a gene encoding the lytic functions of *Bacillus amyloliquefaciens* phage: evidence of an auxiliary lysis system. *J Biosci Bioeng* **88**, 221–225.
- Ochman, H., Gerber, A. S. & Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623.
- Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J. & other authors (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**, 171–182.
- Radnedge, L., Davis, M. A. & Austin, S. J. (1996). P1 and P7 plasmid partition: ParB protein bound to its partition site makes a separate discriminator contact with the DNA that determines species specificity. *EMBO J* **15**, 1155–1162.
- Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M. H., Jr & Stulke, J. (1999). Novel phosphotransferase system genes revealed by genome analysis – the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. *Microbiology* **145**, 3419–3429.
- Romero, P., Lopez, R. & Garcia, E. (2004). Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus. *Virology* **322**, 239–252.
- Rupnik, M., Avesani, V., Janc, M., von Eichel-Streiber, C. & Delmee, M. (1998). A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *J Clin Microbiol* **36**, 2240–2247.
- Rupnik, M., Brazier, J., Duerden, B., Grabnar, M. & Stubbs, S. (2001). Comparison of toxinotyping and PCR ribotyping of *Clostridium difficile* strains and description of novel toxinotypes. *Microbiology* **147**, 439–447.
- Rupnik, M., Dupuy, B., Fairweather, N. F., Gerding, D. N., Johnson, S., Just, I., Lyster, D. M., Popoff, M. R., Rood, J. I. & other authors (2005). Revised nomenclature of *Clostridium difficile* toxins and associated genes. *J Med Microbiol* **54**, 113–117.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sebahia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeno-Tarraga, A. M. & other authors (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* **38**, 779–786.
- Sharples, G. J., Bolt, E. L. & Lloyd, R. G. (2002). *RusA* proteins from the extreme thermophile *Aquifex aeolicus* and lactococcal phage ϕ 11 resolve Holliday junctions. *Mol Microbiol* **44**, 549–559.
- Sheehan, M. M., Stanley, E., Fitzgerald, G. F. & van Sinderen, D. (1999). Identification and characterization of a lysis module present in a large proportion of bacteriophages infecting *Streptococcus thermophilus*. *Appl Environ Microbiol* **65**, 569–577.
- Shepherd, A. J., Gorse, D. & Thornton, J. M. (1999). Prediction of the location and type of beta-turns in proteins using neural networks. *Protein Sci* **8**, 1045–1055.
- Tan, K. S., Wee, B. Y. & Song, K. P. (2001). Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. *J Med Microbiol* **50**, 613–619.
- Tatusova, T. A. & Madden, T. L. (1999). BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**, 247–250.
- Wang, I. N., Smith, D. L. & Young, R. (2000). Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54**, 799–825.
- Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E. & McDonald, L. C. (2005). Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**, 1079–1084.
- Yamaichi, Y. & Niki, H. (2000). Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**, 14656–14661.
- Young, R. (1992). Bacteriophage lysis: mechanism and regulation. *Microbiol Rev* **56**, 430–481.