The complete genome sequence of Clostridium difficile phage \( \phi \)C2 and comparisons to \( \phi \)CD119 and inducible prophages of CD630

Shan Goh,† Peh Fern Ong,† Keang Peng Song,†‡ Thomas V. Riley‡,‡ and Barbara J. Chang

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

The complete genomic sequence of a previously characterized temperate phage of Clostridium difficile, \( \phi \)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 \( \phi \)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 \( \phi \)C2 to \( \phi \)CD119 and the CD630 prophage sequences showed relatedness between \( \phi \)C2 and the prophages, but less so to \( \phi \)CD119. \( \phi \)C2 integrated into a gene encoding a putative transcriptional regulator of the gntR family. \( \phi \)C2, \( \phi \)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 \( \phi \)C2, \( \phi \)CD119 and CD630 prophages were dissimilar. \( \phi \)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 (Sebaihia 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 \( \phi \)CD119, has been sequenced (Govind et al., 2006), while two putative prophage sequences were detected in the recently sequenced genome of C. difficile CD630 (Sebaihia et al., 2006).
Previously, we induced a temperate phage, \( \phi C2 \), 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, \( \phi C6 \) and \( \phi C8 \), 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 \( \phi C2 \) genome with the genomes of \( \phi CD119 \) and \( C. \) difficile CD630 (Sebaihia et al., 2006), as well as the unfinished sequence of \( C. \) difficile QCD-32g58. We found \( \phi C2 \) genes to be prevalent in most of the clinical \( C. \) difficile isolates tested. Two phages, designated \( \phi C630-1 \) and \( \phi C630-2 \), 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 \( \mu \)g erythromycin ml\(^{-1} \) (Sigma) and 10 \( \mu \)g tetracycline ml\(^{-1} \) (Sigma) was induced with 3 \( \mu \)g mitomycin C ml\(^{-1} \) (Sigma) and incubated for another 8 h. The culture supernatant was filtered through a 2 \( \mu \)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 \( \phi C630-1 \) and \( \phi C630-2 \) with a 4 h culture of CD843 was as previously described for \( \phi C6 \) (Goh et al., 2005b). \( \phi C2 \) 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 CaCl\(_2\) density gradient, DNA extraction by phenol/ chloroform/isooamyl 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. \( \phi C2 \) DNA was digested with AclI, Hinfl, 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 dyeodeoxy 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 3.33 x (293 sequencing runs) and a minimum of 2 x coverage (at least once in each strand) were obtained from sequencing the phage library. Gaps between contigs of \( \phi C2 \) 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 \( \phi C2 \) and \( C. \) difficile CD630 or QCD-32g58 was detected by BLASTN (Altschul et al., 1990) at http://www.sanger.ac.uk/Projects/C_dificile/ and http://www.ncbi.nlm.nih.gov/genomes/geblast.cgi?gi=5410, respectively. Alignments between two sequences were carried out with b2seq (Tatusova & Madden, 1999) and the genome was searched for rRNA genes using the rRNAscan 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 \( \phi C2 \) putative ORF 36 and ORF 37, respectively, cloned between the BamHI and PstI sites of the pQE-30 expression vector (Qiagen). ORF 36 was amplified with Vent polymerase (NEB) using HolFbam (5'-CGCGGATCCATGGTTGAAG-3') and HolRpt (5'-AAC-TGGATCTTTCATCATCCT-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 2 min. ORF37 was amplified with Vent polymerase (NEB) and AbiFbam (5'-CGCGGATCCATGGTTGAAG-3') and AbiRPT (5'-AATCCTCCTTTGATGCGG-3') with 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 BamHI and PstI 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 \( \phi C2 \) DNA inserts were grown at 37° C in LB (Invitrogen) supplemented with 100 \( \mu \)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 \( \mu \)g ampicillin ml\(^{-1} \) and 25 \( \mu \)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 OD\(_{600}\) of 0.6.

**Southern hybridization and dot blot.** \( \phi C2 \) genomic probe was prepared by DIG-labelling of Hinfl-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 HinflIII or HindIII and XbaI, separated in 1% TBE agarose and transferred to 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 \( \phi C2 \) 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 x 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. \( \phi C2 \) was purified through a CaCl2 density gradient as previously described (Goh et al., 2005b); 20 \( \mu \)l
**Table 1. Bacterial strains, phages and plasmids**

<table>
<thead>
<tr>
<th>Bacterial strain(s), phage or plasmid</th>
<th>Relevant properties*</th>
<th>Purpose in study</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3, 7, 13, 32, 43, 25, 28, 46, 56, 66</td>
<td>tcdA+ tcdB+</td>
<td>Dot blot</td>
<td>Singapore General Hospital (SGH), Singapore</td>
</tr>
<tr>
<td>CD1, 4, 33, 39, 40, 55, 57, 58</td>
<td>tcdA− tcdB−</td>
<td>Dot blot</td>
<td>M. Rupnik†</td>
</tr>
<tr>
<td>CD12</td>
<td>tcdA+ tcdB+</td>
<td>Phage induction</td>
<td>P. Mullany†</td>
</tr>
<tr>
<td>CD19</td>
<td>tcdA− tcdB−</td>
<td>Stable lysogen, dot blot</td>
<td>Sir Charles Gardner Hospital (SCGH), Western Australia</td>
</tr>
<tr>
<td>CD22, 70</td>
<td>tcdA+ tcdB+</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD630</td>
<td>tcdA+ tcdB+</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD61, CD62</td>
<td>Serogroup X tcdA+ tcdB+ (toxinotype 0)</td>
<td>Dot blot</td>
<td>barbut et al. (1993); Rupnik et al. (1998)</td>
</tr>
<tr>
<td>CD38</td>
<td>tcdA+ tcdB+</td>
<td>Dot blot</td>
<td>Haslam et al. (1986); Rupnik et al. (1998)</td>
</tr>
<tr>
<td>CD60</td>
<td>tcdA+ tcdB+</td>
<td>Phage propagation, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD63</td>
<td>tcdA+ tcdB+</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD371</td>
<td>tcdA+ tcdB+</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD578</td>
<td>tcdA− tcdB−</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD242</td>
<td>tcdA− tcdB−</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD062</td>
<td>tcdA+ tcdB+</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CD843</td>
<td>tcdA+ tcdB+</td>
<td>Stable lysogen, dot blot</td>
<td></td>
</tr>
<tr>
<td>CCUG 37782</td>
<td>Serogroup F, tcdA− tcdB+ (toxinotype VIII)</td>
<td>Dot blot</td>
<td>Barbut et al. (1993); Rupnik et al. (1998)</td>
</tr>
<tr>
<td>CCUG 20309</td>
<td>Serogroup A, tcdA+ tcdB+ (toxinotype X)</td>
<td>Dot blot</td>
<td>Haslam et al. (1986); Rupnik et al. (1998)</td>
</tr>
<tr>
<td>CCUG 19126 VPI 10463</td>
<td>Serogroup G, tcdA+ tcdB+ (toxinotype 0)</td>
<td>Cloning</td>
<td>Promega</td>
</tr>
<tr>
<td>E. coli JM109</td>
<td>endA1 recA1 gyrA96 thi-1 hsdRI7(rC m(l)) relA1 supE44 lacZ (proAB lac–proAB lacZD36 phoA lacPZAM15)</td>
<td>Cloning</td>
<td>Qiagen</td>
</tr>
<tr>
<td>E. coli M15[pREP4]</td>
<td>Na+ Str+ Rif+ Km+ RecA+ Uvr+ Lon+ lac ara gal mtl lac repressor plasmid</td>
<td>Protein expression</td>
<td>Qiagen</td>
</tr>
<tr>
<td>pQPE-30</td>
<td>Ap+ expression vector, C-terminal His tag</td>
<td>Protein expression</td>
<td>Qiagen</td>
</tr>
<tr>
<td>pUC19</td>
<td>Ap+ lacZα, cloning vector</td>
<td>Cloning</td>
<td>Promega</td>
</tr>
</tbody>
</table>

*tcdA+ indicates a truncated gene.
†P. Mullany, University College London, UK; M. Rupnik, Max Planck Institute of Biophysical Chemistry, Göttingen, Germany.

(10^7 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 (0.25 M Tris pH 6.8, 50%, w/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 BLUEPRINT 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% w/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'-GTAAGGATGATAGTGATGAAAG-3'; int-1, 5'-GCATTTCACAAATTTGCCCACCG-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'-CTGTGAAATGATTGTGATTCCAT-3'; Cdu1-1, 5'-GGTTAGCGAAGGGAGGT-3'; int-3, 5'-CAAACAGTAGCATGACACAT-3'; int-5, 5'-CGCAAGAATGAAAATTAAAG-3'. The PCR products were sequenced with the same primers. To determine attBP, an additional primer, gntR-F (5'-GAGATTTGAAGTAAATTCC-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 (10^7–10^9 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 × magnification, operating at 100 kV. Dimensions of a minimum of four phage particles for each phage were measured, and results expressed as mean ± SD.

RESULTS

General features of φC2 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 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 φCD119 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 φC2 were separated by 12 % SDS-PAGE (Fig. 1b) and two major proteins of φC2 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 φC2 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 φCD119 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 φC2 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 φC2 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 ActI, HinClI 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
bacterial sequence of \( \text{attPB} \) had high percentage similarity to a \( \text{gntR} \) transcriptional regulator in CD630 (92% of 93 amino acids, \( E \)-value \( 1.4 \times 10^{-38} \)) and QCD-32g58, a forward primer specific for \( \text{gntR} \) (\( \text{gntR}-\text{F} \)) and \text{int-3} were used to determine the sequence of \( \text{attBP} \) (Fig. 5c). The \( \phi \text{C2} \) attachment site sequence CTGTGAGAAAT is different from that of \( \phi \text{CD119} \) (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 \( 3 \times 10^{-11} \)) and \( \text{attP} \) of \( \phi \text{CD119} \) was 248 nt downstream of a Cdu1 homologue (Govind et al., 2006).

\( \phi \text{CD119} \) and prophage 1 and 2 homologues were mostly found in the DNA replication, recombination and
modification module of the \( \phi C2 \) 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 \( \phi C2 \), \( \phi C630-1 \), \( \phi C630-2 \) and \( \phi CD119 \) 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 (\( \phi C630-1 \)) compared to prophage 2 (\( \phi C630-2 \)). \( \phi C630-1 \) particles measured 31.7 ± 0.7 nm in head diameter and 62.4 ± 5.1 nm in tail length, while \( \phi C630-2 \) 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 \( \phi C2 \) genome to regions of CD630 that indicated the \( \phi C630-1 \) and \( \phi C630-2 \) prophages have similar genome organization (Fig. 2). Dot plots (not shown) of \( \phi C2 \) and the hypothetical genome sequences of \( \phi C630-1 \) and \( \phi C630-2 \) prophages have similar genome organization (Fig. 2).
similar pair compared to \( \phi C2/\phi C630-1 \) and \( \phi C2/\phi C630-2 \). Dot-plot comparisons of \( \phi CD119 \) to \( \phi C630-1 \) and \( \phi C630-2 \) 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 \( \phi C2 \)-related genes in clinical \( C.\ difficile \) strains, dot blots and Southern blots were carried out with a \( \phi C2 \) 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 \( \phi C2 \)-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 \( \phi C2 \) (Fig. 2). However, a whole prophage genome was not detected in QCD-32g58.

\[ \text{Fig. 5. Integration of } \phi C2 \text{ DNA into the } C.\ difficile \text{ genome. (a) Southern hybridization of } \text{attP} \text{ probe to phage and lysogenic DNA digested with } \text{HincII}. \text{Although cross-hybridizations of the } \text{attP} \text{ probe to other fragments of } \phi C2 \text{ were observed, the brightest signal was a 2.4 kb band, which was absent in lysogen CD242. Two new observed bands of 1.1 kb and 5.7 kb may contain the } \text{attPB'} \text{ and } \text{attBP'}. \text{M, DIG-labelled/HindIII marker (Roche). (b) } \phi C2 \text{ DNA before integration, where the } \text{attPP'} \text{ sequence (bold) was between ORF46 (black arrow) and ORF47 (grey arrow). The } \text{attPP'} \text{ aligned with a homologous } \text{attBB'} \text{ sequence (bold) in a gene encoding a gntR-like transcriptional regulator, present in the host genome. Divergent PCR primers patt-0/Cdu1-1 and int-3/int-5 were used to amplify and sequence the } \text{attPB'} \text{ and } \text{attBP'}, \text{ respectively. Primers patt-1/int-1 were used to amplify the } \text{attPP'} \text{ region to use as a probe for Southern hybridization in (a). (c) Right side (} \text{attPB'} \text{) of integrated phage DNA in the lysogenic } C.\ difficile \text{ genome by inverse PCR and left side (} \text{attBP'} \text{) by PCR with gntR-F and int-3 primers. } \phi C2 \text{ integration resulted in a disrupted gntR. } C.\ difficile \text{ sequences are in lower case; } \phi C2 \text{ sequences are in upper case; the homologous } \text{att} \text{ site is in bold.} \]
partially characterized phage
and phage (Kwan et al., 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 \( \phi C630 \) to other phage genomes to provide an insight into phages of \( C. difficile \).

The genome organization of \( \phi C2 \) 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 \( \phi C2 \) terminase large subunit had high percentage sequence similarity to SPP1 (26% of 426 amino acids, 

\[
E \text{-value} \times 10^{-25}\]

) of the P22-like headful subgroup. This suggests the ends of \( \phi C2 \) were likely to be generated by a headful packaging mechanism. Based on the number of protein homologues and unique hypothetical proteins, \( \phi C2 \) is closely related to the other \( C. difficile \) phages, \( \phi CD119, \phi C630-1 \) and \( \phi C630-2 \), and demonstrates modular mosaicism (Casjens, 2005). Gene divergence appears greatest within the lyssogeny 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 \( \phi C2 \) 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 extrachromosomal replicative proteins ParA and ParB found close to the lysogeny module and their relative direction of transcription suggest their expression is associated with lyssogenic conversion. Interesting possibilities include \( \phi C2 \) 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 \( Cdu1-\text{attP} \)-integrate arrangement is conserved in \( \phi C2, \phi C630-1 \) and \( \phi CD119 \). The \( \text{attP} \)s of \( \phi C2 \) and \( \phi CD119 \) are 197 nt and 248 nt downstream of a \( Cdu1 \) homologue (Govind et al., 2006), respectively, while \( cdu1 \) contains an \( \text{attP} \) for \( \phi C630-1 \) (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 \( \text{abiF} \) 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 \( \phi C2, \phi C630-1, \phi C630-2 \) and \( \phi CD191 \) showed the degree of pairwise relatedness to be \( \phi C630-1/\phi C2 > \phi C2/\phi C630-2 > \phi C630-1/\phi CD119 > \phi CD119/\phi C630-2 > \phi CD119/\phi C2 \). This indicates that \( \phi C630-1 \) and \( \phi C630-2 \) are intermediates of \( \phi C2 \) and \( \phi CD119 \) in the evolutionary sense. The brief sampling of randomly chosen clinical \( C. difficile \) isolates revealed that \( \phi C2 \)-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 \( \phi C2 \)-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 \( \phi C2 \)-related prophage genes in \( C. difficile \) strains. Therefore, the current role of \( \phi C2 \) 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 metallireducens’ 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). φC2 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 mannospecific-phosphotransferase system (PTS) (Sebaihia 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 φC2 into CD630 could lead to significant changes in the C. difficile phenotype through mannos PTS deregulation. The contribution of φC2 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


Edited by: P. R. Herron