JOURNAL OF BACTERIOLOGY, Apr. 2006, p. 2568–2577 Vol. 188, No. 7
0021-9193/06/$08.00+0 doi:10.1128/JB.188.7.2568–2577.2006
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

Genomic Organization and Molecular Characterization of *Clostridium difficile* Bacteriophage ΦCD119

Revathi Govind, Joe A. Fralick, and Rial D. Rolfe*

Department of Microbiology and Immunology, Texas Tech University Health Sciences Center, Lubbock, Texas 79430

Received 10 October 2005/Accepted 13 January 2006

In this study, we have isolated a temperate phage (ΦCD119) from a pathogenic *Clostridium difficile* strain and sequenced and annotated its genome. This virus has an icosahedral capsid and a contractile tail covered by a sheath and contains a double-stranded DNA genome. It belongs to the *Myoviridae* family of the tailed phages and the order *Caudovirales*. The genome was circularly permuted, with no physical ends detected by sequencing or restriction enzyme digestion analysis, and lacked a cos site. The DNA sequence of this phage consists of 53,325 bp, which carries 79 putative open reading frames (ORFs). A function could be assigned to 23 putative gene products, based upon bioinformatic analyses. The ΦCD119 genome is organized in a modular format, which includes modules for lysogeny, DNA replication, DNA packaging, structural proteins, and host cell lysis. The ΦCD119 attachment site *attP* lies in a noncoding region close to the putative integrase (*int*) gene. We have identified the phage integration site on the *C. difficile* chromosome (*attB*) located in a noncoding region just upstream of gene *gltP*, which encodes a carrier protein for glutamate and aspartate. This genetic analysis represents the first complete DNA sequence and annotation of a *C. difficile* phage.

Clostridium difficile, a gram-positive, spore-forming, anaerobic bacillus, is the leading cause of nosocomial diarrhea associated with antibiotic therapy (2). *C. difficile* causes a variety of diarrheal syndromes, including diarrhea, nonspecific colitis, and pseudomembranous colitis, all of which vary widely in severity (2). Pathogenic *C. difficile* can produce two major toxins, toxin A, an enterotoxin, and toxin B, a cytotoxin, that are causative agents of diarrhea and colitis (4). Variation in the severity of symptoms of *C. difficile*-associated disease has been attributed in part to the level of toxin production by the infecting strain(s) (4). The toxin genes, *tcdA* and *tcdB*, are part of a 19.6-kb pathogenicity locus (PaLoc), which is present at identical locations in the chromosomes of pathogenic *C. difficile* strains but is missing from the nontoxinogenic strains. This observation has led to the suggestion that the presence of the PaLoc may be associated with a transposable element (5). In other clostridial species, toxins are known to be encoded by mobile elements such as bacteriophages and plasmids (10, 11). However, while there is no direct evidence of lysogenic conversion in *C. difficile* strains, Tan et al. have demonstrated homology between *tcdE*, a gene located within the PaLoc of *C. difficile*, and phage holin genes (33). In another study, Goh et al. analyzed the effect of bacteriophage infection on toxin production and found an increased toxin B production in some lysogens (12). The evolutionary aspects of the PaLoc and its relationship with *C. difficile* phages are not known. Detailed characterization of *C. difficile* phages is necessary to understand their genetics and their potential relationship with the PaLoc of *C. difficile*. In this study, one of our goals has been to sequence the genome of a lysogenic *C. difficile* phage so that such an analysis could begin. This study represents the first detailed characterization of a *C. difficile* phage with a complete DNA sequence and annotation.

(This work is part of the doctoral dissertation of R. Govind.)

* Corresponding author. Mailing address: Department of Microbiology and Immunology, Health Sciences Center, Texas Tech University, School of Medicine, Lubbock, TX 79430. Phone: (806) 743-2905. Fax: (806) 743-2334. E-mail: rial.rolfe@ttuhsc.edu.

MATERIALS AND METHODS

Bacterial growth conditions and media. The *C. difficile* CD119 lysogen F10 and the ΦCD119 phage host *C. difficile* strain 602 were obtained from Rosanna Dei, Universitá degli Studi di Firenze, Italy. Bacterial strains were stored in chopped meat broth (Carr Scarborough Microbiologicals, Inc., Decatur, GA) at room temperature. When required, the cultures were subcultured on brain heart infusion (BHI) agar and incubated anaerobically (anaerobic system; Forma Scientific, Inc., Marietta, OH) at 37°C. Bacteriophage ΦCD119 was induced by mitomycin C treatment from ΦCD119 lysogen F10 and was isolated by techniques described by Mahony et al. (21, 22).

Bacteriophage production and titration. A single colony of host strain 602 was inoculated into BHI broth and incubated at 37°C overnight. One milliliter of the overnight culture was used to inoculate 50 ml of BHI broth and allowed to grow for 2 to 3 h until the optical density at 550 nm reached 0.4. A 0.5-ml volume of 10^8 PFU/ml of phage stock was added to the bacterial culture and incubated anaerobically at 37°C for 20 h. Clearing of the bacterial cultures was monitored spectrometrically at the optical density at 550 nm at regular intervals. The lysed bacterial cultures were centrifuged, and the supernatants were collected and filtered through a 0.4-μm filter. This method of propagation yielded phage titers as high as 10^8 to 10^9 PFU/ml. Phage titers were determined by mixing different serial dilutions of phage lysates with 600 μl of an exponential culture of indicator strain 602 in 3 ml molten BHI top agar (7%) which was poured into BHI plates and incubated anaerobically overnight at 37°C.

Purification of phage. Filtered phage lysates were treated with 10 μg/ml of DNase and RNase cocktail for 1 to 2 h at 37°C. NaCl was then added to a final concentration of 1 M and stirred slowly on ice for an hour. Cell debris was removed by centrifugation at 11,000 × *g* for 10 min at 4°C. The phage were then collected from the supernatant by precipitation with 10% polyethylene glycol 8000 for 2 h on ice and centrifugation as described above. The phage pellets were suspended in 1 ml BHI broth and filter sterilized using 0.4-μm filters.

Library preparation and shotgun sequencing. DNA was isolated from purified bacteriophage with the High Pure lambda isolation kit (Roche). Bacteriophage DNA was sheared by passing it through a 25-gauge needle four times and end repaired using the DNA terminator end repair kit (Lucigen). Phage DNA fragments of sizes from 2 to 4 kb were gel purified and ligated into the pSmart HC vector (Lucigen). The ligation reaction was transformed by electroporation into "E. cloni" 10G electrocompetent cells (Lucigen), and transformants were selected on LB agar containing carbenicillin (100 mg/ml). Plasmids were isolated from 350 randomly picked transformants using the Qiaspin miniprep plasmid

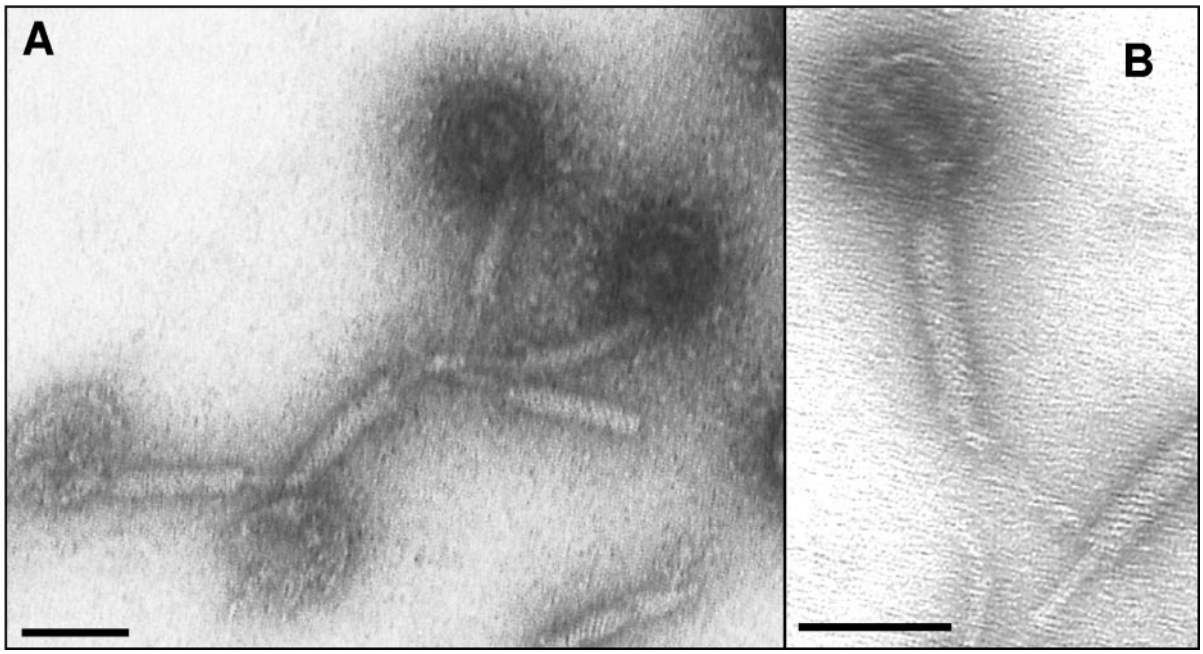

FIG. 1. Electron microscopy of phage ΦCD119 showing its icosahedral capsid and a flexible tail. Bar, 50 nm. Purified phage at a concentration of 1×10^{10} (5 μl) were placed on the top of a carbon film fixed on a copper disk for 5 min. Excess solution was removed, and the grid was washed with water and then negatively stained with 2% uranyl acetate. Pictures of the virus were taken with a transmission electron microscope at magnifications of ×40,000 (A) and ×60,000 (B).

purification kit. Inserts in plasmids were sequenced with primers AmpL1 and AmpR1 by using an ABI PRISM 370 automated DNA sequencer (Center for Biotechnology and Genomics, Texas Tech University).

Sequence assembly and analysis. The sequences obtained were edited and aligned using the software SeqMan (DNASTAR, Inc.). Gaps were filled by direct sequencing of ΦCD119 DNA with specific primers designed from the contigs. The final consensus sequence was analyzed for the presence of protein coding regions using GeneMark (http://opal.biology.gatech.edu/GeneMark/). The predicted proteins were then compared to the NCBI protein database with Blastp (http://www.ncbi.nlm.nih.gov/BLAST/). Structural features of the proteins were determined with the proteomic tools at ExPASy (http://us.expasy.org/). Comparisons of phage sequences with the host genome were performed using the BLAST server at the *C. difficile* sequencing project (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_difficile). The complete DNA sequence of bacteriophage ΦCD119 can be found in GenBank under accession number AY855346.

Generating 602/ΦCD119 lysogens. Phage ΦCD119 was spotted on a lawn of *C. difficile* strain 602 on BHI agar plates and incubated overnight at 37°C under anaerobic conditions. Bacterial colonies within the lysis zone were then picked with sterile toothpicks and tested for phage production following mitomycin C (10 μg/ml) treatment.

Preparation of phage proteins, SDS-PAGE, and N-terminal sequencing. Polyethylene glycol-precipitated bacteriophage was further purified by CsCl density gradient as described by Sambrook et al. (31). Purified phage preparation (1 ml) was precipitated by adding 4 volumes of ice-cold acetone. Samples were centrifuged at 20,000 × *g* for 10 min, the supernatant was discarded, and the pellet was allowed to air dry. The pellet was then resuspended in 100 μl of sample buffer (2 ml of 10% sodium dodecyl sulfate [SDS], 0.2 ml of 0.5% bromophenol blue, 1.25 ml of 0.5 M Tris-HCl [pH 6.8], and 2.5 ml of glycerol, made up to 9.5 ml with deionized water; 50 μl of β-mercaptoethanol was added to 950 μl of this solution prior to use). Samples were boiled for 5 min before being loaded onto SDS-polyacrylamide gel electrophoresis (PAGE) gels. Proteins were electrotransferred from polyacrylamide gels onto polyvinylidene difluoride membranes (Bio-Rad Corp., Richmond, Calif.) in buffer A (25 mM Tris, 192 mM glycine, 20% methanol [pH 8.3]), using a Trans-Blot cell (Bio-Rad, Alpha Technologies, Dublin, Ireland), according to the manufacturer's instructions. Proteins were stained with Coomassie brilliant blue R250, cut out of the membrane, and sequenced on a Porton Instruments 2020 sequencer with online Beckman 32-karat analysis system (Center for Biotechnology and Genomics, Texas Tech University).

Identification of *attPP′* and *attBB′* site. The chromosomal DNA from 602/ΦCD119 lysogens was extracted using DNAZOL reagent (Invitrogen) and used as a template for the identification of the attachment site by inverse PCR (26). The attachment site was expected to be located in a noncoding region immediately downstream of the integrase gene (*int*). A Tsp45I restriction site is present within the *int* gene, and this enzyme was used for complete digestion of the lysogen DNA. Fragments were then treated with T4 DNA ligase to obtain self-ligated circular molecules. Divergent primers INTEG-UP (5′-GCATCTGA AAATTTGAGCAAA-3′) and INTEG-DOWN (5′-TTTTGTTGTGTCCAAAT CTGAA-3′), complementary to a region within the *int* gene, were used for PCR amplification of ligated fragments. The reaction yielded an 840-bp product, which was later purified and sequenced using the same primers. The obtained sequence contained the *attBP′* site, and the nonprophage part of the sequence displayed 100% identity over 639 nucleotides to a sequence of the *C. difficile* strain 630 genome available from the Sanger Institute, United Kingdom (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_difficile) (J. Parkhill, personal communication). Two more primers, attCD-UP (5′-TCTCCGTCAACAATTT AACCA-3′) and attCD-DOWN (5′-AATCGGAAGTTATGCACCAGA-3′), were designed from the bacterial part of the *attBP′* sequence. Inverse PCR was repeated using Bst1007I restriction enzyme-digested and ligated 602/ΦCD119 lysogen DNA templates. This reaction gave an *attPB′* sequence of 1,054 bp, 860 of which were from the bacterial chromosome.

Confirmation of ΦCD119 attachment site by Southern blot hybridization. *C. difficile* 602 and its ΦCD119 lysogens were used to confirm the *attP* site. Chromosomal DNA (10 μg) from the above strains was digested with Tsp45I restriction enzyme and separated on a 0.8% agarose gel by electrophoresis. The separated DNA was then transferred to a positively charged IMMOBILON–NY^+ nylon membrane (Millipore, Bedford, MA) by the capillary transfer method (31). The sequence near the phage integration site in the bacterial chromosome was PCR amplified using primers HyP-forward (5′-AAAATGCTAAATTTGGTTT GT-3′) and GltP-reverse (5′-GCTAACATTCCTGCCTCTGG-3′). The PCR product was radiolabeled with ^{32}P using the Random prime kit (Roche Applied Sciences). The membrane containing the transferred DNA was hybridized with radiolabeled probe as described previously (31) and the ^{32}P detected with the Typhoon 9410 (Amersham Pharmacia Biotech, NJ).

Nucleotide sequence accession number. The genome from phage ΦCD119 was deposited in GenBank under accession number AY855346.

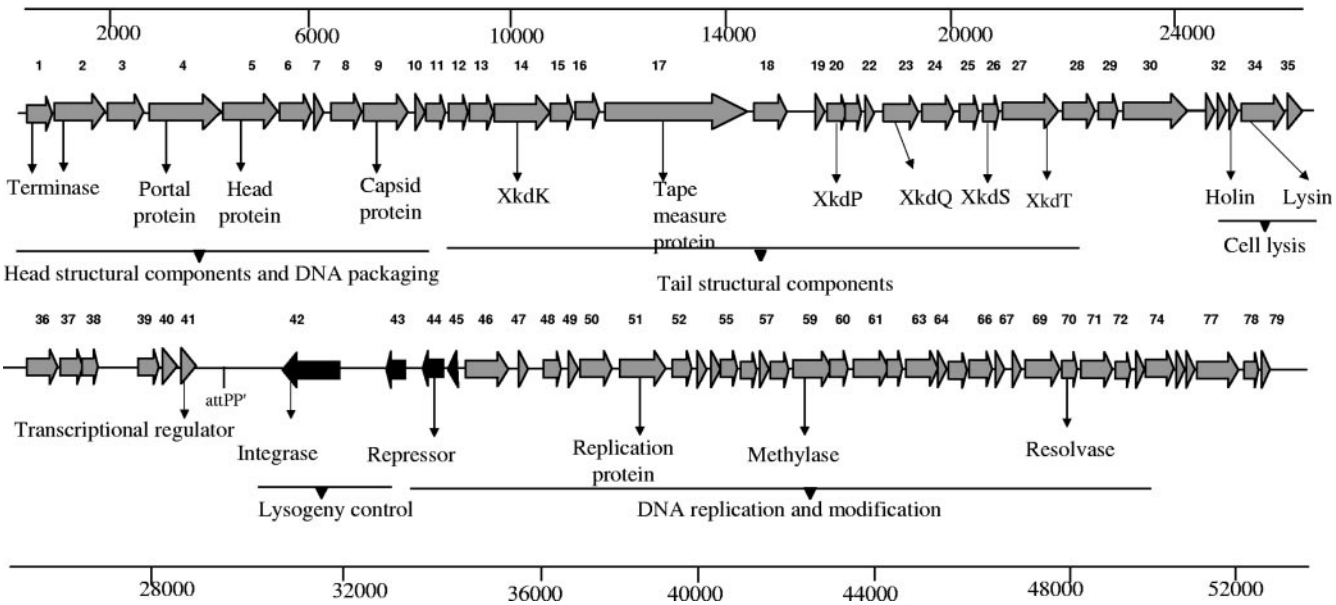

FIG. 2. Genetic and physical organization of ΦCD119 genome with predicted ORFs and some functional assignments. The ORFs (1 to 79) are indicated by arrows or arrowheads pointing in the direction of transcription. The relative positions of the ORFs and the *attPP′* site in the genome are marked.

RESULTS

General features of phage ΦCD119 and its genome. Electron microscopy revealed that the ΦCD119 virion has an icosahedral capsid (diameter, 50 nm) with a contractile tail (length, approximately 110 nm) (Fig. 1). Purified nucleic acid contents of the phage were treated with DNase, RNase, or various restriction enzymes to determine its biochemical nature. It was found to be RNase resistant and DNase susceptible (data not shown) and could be digested with restriction enzymes. Hence, we have classified this phage under the *Myoviridae* family of double-stranded DNA bacterial viruses in the order *Caudovirales* (1). Based on sequence analysis, the genome of ΦCD119 is a double-stranded DNA molecule containing 53,325 bp. It has an average GC content of 28.7%, which is similar to the reported 29.06% GC content of the *C. difficile* genome (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_difficile). No physical terminus of the genome was detected by multiple rounds of primer walking (the ends of the phage genome depicted in Fig. 2 and Table 1 are arbitrary). No evidence of the presence of cohesive ends (cos sites) on ΦCD119 DNA was found when restriction enzyme digestions were followed by heating to 80°C and rapid cooling prior to electrophoresis (Fig. 3A). A circularly permuted and terminally redundant linear phage chromosome behaves as a circular chromosome with respect to restriction analysis (3). Restriction analysis of the ΦCD119 DNA showed behavior of a circular genome. For example, the BsmI digest should produce fragments of sizes of 14,561, 11,791, 10,002, 8,341, 4,035, 2,788, and 1,807 bp, assuming a circularly permuted genome (Fig. 3C). We could see all seven fragments in Fig. 3A, lanes 3 and 4. Undigested phage DNA ran as a single, sharp band on 0.7% agarose gels (Fig. 3B, lane 4). When restriction enzymes that cut once (SphI and MscI) were used to digest the genome, the DNA ran similarly to the undigested DNA. Double digestion with SphI and MscI produced two DNA fragments. These observations suggest that the ΦCD119 genome is circularly permuted. In bacteriophages that carry circularly permuted linear chromosomes, the replicated phage concatemeric DNA is recognized at a *pac* site by the phage terminase, a cut is made in the DNA at or near that point, and a series of packaging events proceeds in one direction from the DNA break thus produced (3). When such virion DNA is cleaved by a restriction enzyme, a unique fragment, one of whose ends is the packaging series initiation cut, is generated, and this fragment is thus present in submolar amounts relative to the true restriction fragments. No apparent submolar DNA fragment could be seen in the ethidium bromide-stained electrophoresis gels of ΦCD119 restriction digests. Hence, further studies will be needed to identify the *pac* initiation site and direction of packaging. Similar behavior has been reported for other circularly permuted phage genomes, such as A118 of *Listeria monocytogenes* (20), the coliphage 933W (27), and the pneumococcal phage of EJ-1 (30). Time-limited treatment of ΦCD119 DNA with the exonuclease BAL-31, followed by complete digestion with restriction enzymes, revealed that all fragments were simultaneously degraded, in contrast to the specific truncation of fragments observed in the control, λ DNA (data not shown). These results taken together suggested that there are no invariable ends in the mature ΦCD119 DNA molecules, that is, the packaged DNA is circularly permuted.

Predicted ORFs and their features. The DNA sequence of ΦCD119 was analyzed for the presence of open reading frames (ORFs), and the putative products were compared with the nonredundant protein database (http://www.ncbi.nlm.nih.gov/BLAST/). A total of 79 ORFs were predicted from the DNA sequence (Table 1 and Fig. 2), some of which code for unique products, with little or no homology to proteins from the database, and others which code for proteins with a high degree of homology to known phage proteins. Generally, phage genomes are organized in modular structures, with each mod-

TABLE 1. Features of bacteriophage ΦCD119 ORFs, gene products, and their functional assignments

ORF	Start position	Stop position	No. of aa[a]	Predicted function	Accession no.	Significant match(es) (source, E value)[b]
1	201	692	163	Terminase	NP_815686.1	Terminase, large subunit, putative (prophage in *Enterococcus faecalis* V583, 1e−52)
2	778	1746	323	Terminase	NP_815686.1	Terminase, large subunit, putative (prophage in *E. faecalis* V583, 2e−61)
3	1897	2595	232			
4	2610	3965	451	Portal protein	NP_814126.1	Portal protein (prophage in *E. faecalis* V583, 1e−23)
5	3978	5018	346	Head protein	NP_814127.1	Minor head protein (prophage in *E. faecalis* V583, 7e−14)
6	5086	5715	209		NP_607551.1	Hypothetical phage protein (*Streptococcus pyogenes* MGAS8232, 1e−08)
7	5737	5931	64			
8	6323	6919	198	Scaffold protein	NP_814130.1	Scaffold protein (prophage in *E. faecalis* V583, 4e−08)
9	6943	7881	312	Capsid protein	ZP_00234864.1	Main capsid protein gp34 (prophage in *L. monocytogenes* F6854, 5e−15)
10	8143	8424	93			
11	8497	8847	116			
12	8910	9260	116			
13	9271	9723	150			
14	9724	10794	356		NP_782684.1	Phage-like element PBSX protein XkdK (*C. tetani* E88, 3e−72)
15	10809	11240	143		NP_782683.1	Phage-like element PBSX protein XkdM (*C. tetani* E88, 1e−25)
16	11272	11742	156		NP_389149.1	PBSX phage protein XkdN (*B. subtilis* 168, 3e−04)
17	11919	14753	944	Tape measure protein	NP_562046.1	Phage-related hypothetical protein (*Clostridium perfringens* strain 13, 3e−17)
18	14958	15578	207			
19	16507	16752	81			
20	16770	17156	128		G69732	PBSX prophage ORF XkdP (*B. subtilis*, 9e−09)
21	17135	17443	102			
22	17464	17706	80			
23	17965	18651	228		NP_782678.1	Phage-like element PBSX protein XkdQ (*C. tetani* E88, 7e−16)
24	18695	19318	207		NP_780938.1	Putative cell wall-associated hydrolase (*C. tetani* E88, 2e−26)
25	19501	19827	108			
26	19827	20195	122		NP_782677.1	Phage-like element PBSX protein XkdS (*C. tetani* E88, 5e−16)
27	20249	21301	351		NP_782676.1	Phage-like element PBSX protein XkdT (*C. tetani* E88, 7e−38)
28	21926	22324	132	Tail fiber protein	NP_900088.1	Probable tail fiber-related protein (*Chromobacterium violaceum* ATCC 12472, 1e−24)
29	22383	22706	108			
30	22724	23995	423			
31	23995	24177	60			
32	24213	24443	76			
33	24463	24720	85	Holin		
34	24720	25535	272	Lysin	ZP_00162412.2	*N*-acetylmuramoyl-L-alanine amidase (*Anabaena variabilis* ATCC 29413, 5e−23)
35	25552	25848	99			
36	26361	26954	198			
37	26972	27403	143			
38	27405	27731	108			
39	27703	28110	135			
40	29563	29844	93			
41	29884	30156	90	Transcriptional regulator	CAA63560.1 cdu1	(*C. difficile*, 3e−06)
42	31674	30568	368	Integrase	ZP_00510128.1	Phage integrase (*Clostridium thermocellum* ATCC 27405, 4e−40)
43	32134	31733	133			
44	33177	32782	131	Repressor	YP_175240.1	Transcriptional repressor of PBSX phage (*Bacillus clausii* KSM-K16, 3e−10)
45	33912	33694	73	Cro/CI like	NP_689001.1	Transcriptional regulator, Cro/CI family (*Streptococcus agalactiae* 2603V/R, 4e−09)
46	34261	35094	278		ZP_00063048.2	COG3561: phage anti-repressor protein (*Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC 8293, 5e−36)
47	35138	35332	65			
48	35955	36311	118			
49	37274	37516	80			
50	37526	38134	202			
51	38135	39025	296	DNA replication	NP_833429.1	Phage replication protein (*Bacillus cereus* ATCC 14579, 2e−16)
52	39275	39682	135	DNA replication	NP_348542.1	Phage-related SSB-like protein (*Clostridium acetobutylicum* ATCC 824, 1e−16)
53	39757	40011	84			
54	40058	40348	96			
55	40345	40710	121			
56	40775	41095	106			
57	41355	41528	57			
58	41528	41863	111			
59	41949	42731	260	DNA methylase	ZP_00314461.1	Site-specific DNA methylase (*C. thermocellum* ATCC 27405, 6e−70)
60	42712	43044	110	DNA methylase	ZP_00314461.1	Site-specific DNA methylase (*C. thermocellum* ATCC 27405, 2e−25)
61	43343	43999	218			
62	44004	44339	111			
63	44370	45074	234	Recombination	YP_215329.1	Lambda Nin-like protein (*Salmonella enterica* subsp. *enterica* serovar Choleraesuis strain SC-B67, 4e−04)
64	45071	45244	57			
65	45237	45608	124			
66	45611	46078	155			
67	46157	46342	61			

Continued on following page

TABLE 1—*Continued*

ORF	Start position	Stop position	No. of aa[a]	Predicted function	Accession no.	Significant match(es) (source, E value)[b]
68	46356	46595	79			
69	46724	47530	268	Methyltransferase	BAA11514.1	Methyltransferase (*Curtobacterium albidum*, 1e−53)
70	47544	47897	117	Holliday junction resolvase	ZP_00303454.1	Holliday junction resolvase (*Novosphingobium aromaticivorans* DSM 12444, 4e−16)
71	47986	48693	235	Antirepressor	ZP_00089317.1	Phage antirepressor protein (*Azotobacter vinelandii*, 1e−21)
72	48787	49275	162			
73	50002	50196	64			
74	50196	50804	203			
75	50826	51071	81			
76	51051	51248	65			
77	51468	52412	314			
78	52720	53019	99			
79	53085	53315	76			

[a] aa, amino acids.
[b] Predicted by computer analysis.

ule containing clusters of genes with specific functions (6). The ΦCD119 genome is no exception and is organized into four modules containing gene clusters for lysogeny control, DNA replication and packaging, structural proteins, and host cell lysis.

Lysogeny module. ORFs 42 and 44 are transcribed divergently from the other ORFs of ΦCD119 and share sequence similarities with an integrase and an XRE family repressor, respectively. ORF 42 contains an integrase-like domain found in the integrase gene of the *Escherichia coli* P4 phage (accession no. gnl CDD 27722; E value, 9e−05). ORF 42 lies close to the identified *attP* site, an organizational arrangement common to other temperate phages (38), and its product may play a role in the site-specific integration of the ΦCD119 genome into the *C. difficile* chromosome. ORF 44 contains a helix-turn-helix domain (IPR001387) which belongs to the XRE family of repressors and displays N-terminal sequence similarities to a repressor of a *Bacillus clausii* phage (PBSX) (38). Hence, ORF 44 may play a role in the maintenance of lysogeny of ΦCD119.

DNA replication, recombination, and DNA packaging module. ORFs coding for putative DNA methylases (ORFs 59, 60, 69), single-stranded DNA binding protein (ORF 52), and Holliday junction resolvase (ORF 65) could be identified in the ΦCD119 genome based on protein sequence similarities. DNA methylases are known to participate in regulatory events of DNA replication, methyl-directed mismatch repair, and transposition (23). These enzymes are also known to be associated with bacterial DNA restriction modification systems that are responsible for the degradation of foreign DNA, such as conjugative plasmids, transposons, and phage DNA. It has been speculated that some bacteriophages express their own DNA methylases to overcome this bacterial protection (23). ORFs 1 and 2 are possibly coding for the terminase enzymes but show no similarity with any well-characterized terminase proteins in the database. Blastp matches for ORFs 1 and 2 are series of uncharacterized terminase proteins. Terminase proteins are required for packing of the phage genomic DNA into the preassembled empty capsid shells (8, 29). ORF 4 shows a high sequence similarity (44% to 55% similarity) to phage portal proteins, and the conserved domain search found the presence of a phage SPP1 portal protein gp6-like domain (pfam05133; E value, 7e−46). Portal proteins are known to form a hole, or portal, that enables phage DNA passage during packaging and ejection. It also forms the junction between the phage head (capsid) and the tail proteins (9). Portal proteins, such as gp6 in phage SPP1, may also participate in procapsid assembly during phage morphogenesis (9). Many of the ORFs in this module encode unique products which shared no homologies with proteins present in the microbial database. Interestingly, the nucleotide sequence of ΦCD119 from bp 41,800 to bp 51,400 (nearly 1/5 of the genome) containing ORFs 59 to 75 is present (100% identical) in the genome of *C. difficile* strain 630 (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_difficile) (see Fig. 7A).

Structural module. Analogous to other double-stranded DNA bacteriophages, the structural module in phage ΦCD119 is located next to the DNA replication module (38). Structural proteins of phage ΦCD119 were examined by SDS-PAGE (Fig. 4), and N-terminal sequencing identified three proteins that correspond to the predicted proteins of ORFs 9, 14, and 15. The apparent molecular weights of these proteins are in agreement with the predicted molecular weight from DNA sequence analysis. The N-terminal sequences (Asn-Thr-Leu-Ala-Tyr-Gly-Gln-Val-Leu-Gln-Gln-Gly-Leu-Asp) for the 34-kDa protein in SDS-PAGE (Fig. 4) matched with the predicted N-terminal sequence of ORF 9, which showed sequence similarity with a major capsid protein in the *L. monocytogenes* prophage (Table 1). N-terminal sequences of the 38-kDa and 16-kDa proteins from SDS-PAGE were identified as Ala-Gly-Leu-Val-Asn-Leu-Asn-Ile-Glu and Ala-Thr-Ser-Phe-Glu-Ser-Lys-Asn-Val-Ile-Asn and matched with predicted amino acids of ORF 14 and ORF 15, respectively. ORFs 14 and 15 share high sequence similarity with *Clostridium tetani* PBSX-like prophage proteins XkdK and XkdM, respectively. Based on the migration patterns of these proteins and also by comparing results from other *Myoviridae* phages (30), XkdK and XkdM may code for sheath and core tail proteins, respectively. PBSX phage is a chromosomally based element which encodes a noninfectious defective myovirus with bactericidal activity in *Bacillus subtilis* strain 168 (32). In the ΦCD119 phage structural module, seven ORFs display strong sequence similarities to genes XkdK, XkdM, XkdN, XkdP, XkdQ, XkdS, and XkdT from the tail morphogenesis region of PBSX phage (Table 1). Similar PBSX-like genes have been identified in the *C. difficile* strain 630 genome (24) as well as in the high toxin-producing *C. difficile* strain VPI

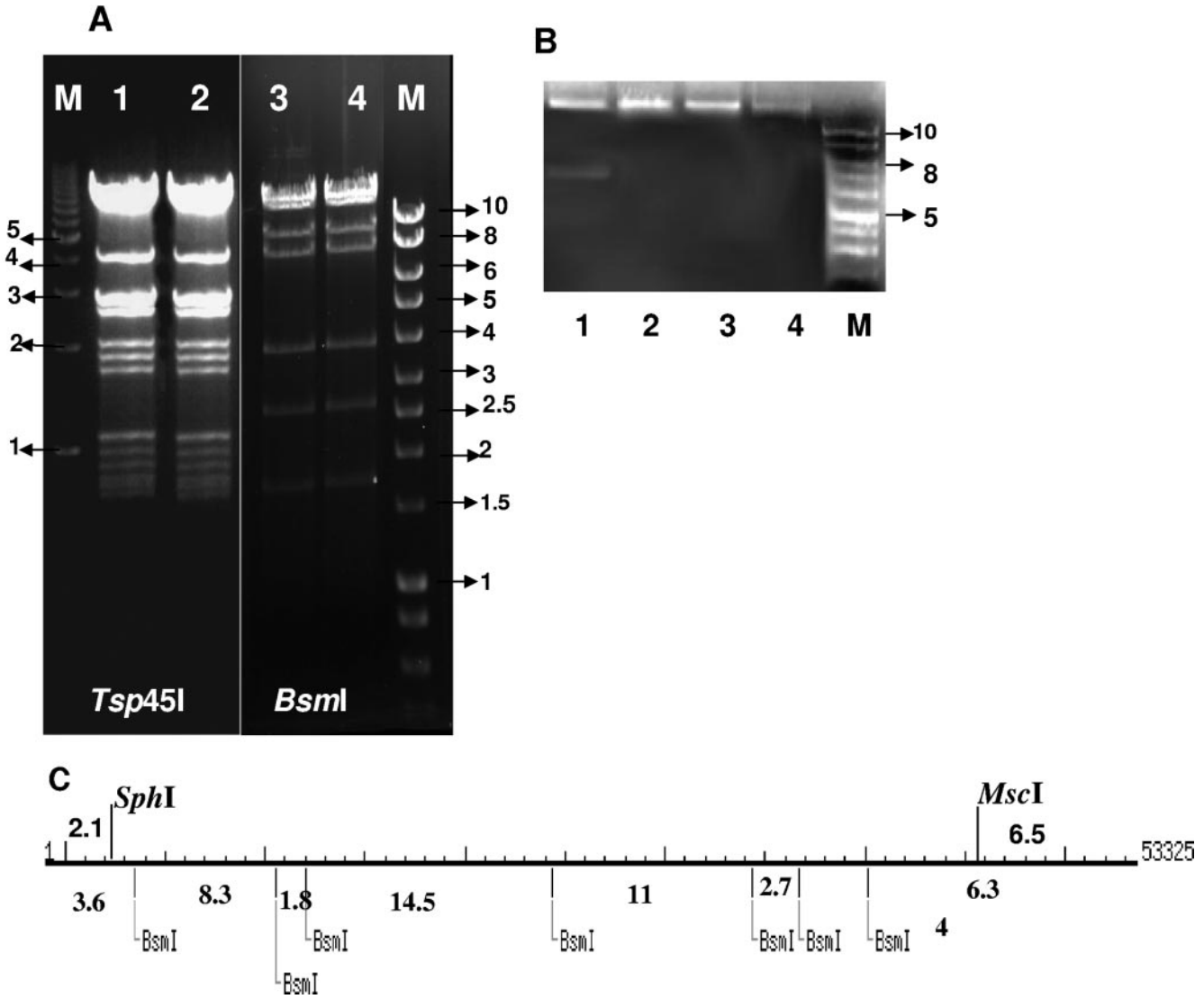

FIG. 3. Restriction digestion analysis of the ΦCD119 genome. (A) Lanes 1 and 2, Tsp45I-digested ΦCD119 DNA; lanes 3 and 4, BsmI-digested ΦCD119 DNA. Lanes 2 and 4 contain DNA that was digested, heated to 80°C, and then chilled on ice before electrophoresis. (B) Lane 1, SphI and MscI double-digested ΦCD119 DNA; lane 2, SphI digest; lane 3, MscI digest; lane 4, undigested ΦCD119 DNA. Lanes M, DNA molecular size markers (in kilobases). (C) Restriction map for SphI, MscI, and BsmI in the ΦCD119 genome. Sizes of expected fragments are marked.

10463 (24). The PBSX phage-like genes in genome 630 are similar but not identical to the PBSX phage-like genes in ΦCD119. The prophage present in *C. difficile* genome 630 possess sequences from a partially characterized *C. difficile* phage ΦC2 (see Fig. 7A), which carry some of the PBSX phage-like tail genes (13). ORF 17 is the largest putative gene in ΦCD119 and may encode a "tape measure protein" which is thought to determine tail length in tailed phage (17). The Blastp hit for ORF 17 was a series of uncharacterized phage tail proteins and tape measure proteins.

Lysis module. The lysis module is located between the structural module and the lysogeny module. ORF 33 and ORF 34 encode a dual lysis system, consisting of a holin and an endolysin responsible for cell lysis and release of phage progeny. Most double-stranded DNA phages require the combination of a holin and an endolysin to achieve host lysis. The disruption of the cell wall is based on peptidoglycan degradation by a phage-encoded muralytic enzyme or endolysin after permeabilization and destabilization of the membrane by a holin, a small membrane protein (36, 37). The endolysin encoded by ORF 34 contains a putative *N*-acetylmuramoyl-L-alanine amidase domain, and enzymes containing this domain digest the peptidoglycan by cleaving the amide bond between *N*-acetylmuramoyl and L-amino acids (34, 36). ORF 33 does not show any homology to known proteins. However, its small size (85 residues) and genome location suggest that it may code for a holin (37). Furthermore, the TMHMM program in ExPASy (http://us.expasy.org/) predicted two transmembrane regions in the protein encoded by ORF 33, which is a hallmark for holins, and the presence of a high number of charged, polar residues in the protein's C terminus is also consistent with known holins (37). Holin accumulation and oligomerization in the cell mem-

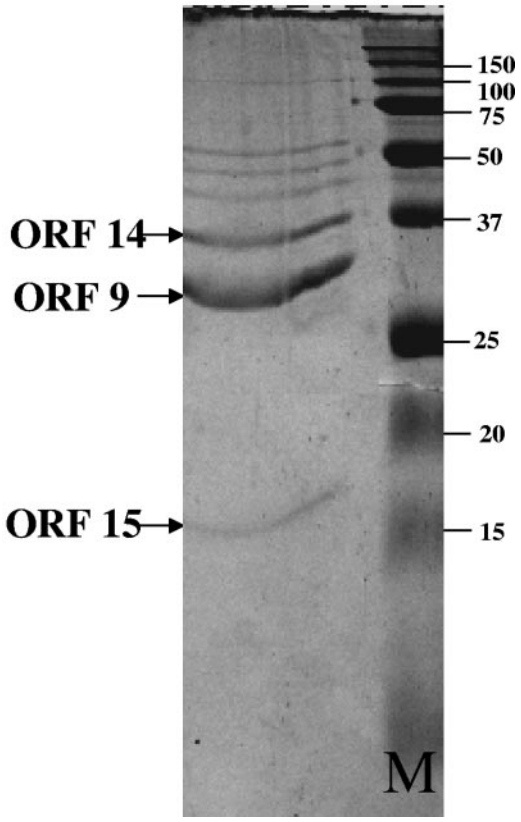

FIG. 4. One-dimensional SDS-PAGE of phage ΦCD119 structural proteins stained with Coomassie brilliant blue. Lane M, precision plus protein marker (Bio-Rad). Protein bands were sequenced, and their corresponding ORFs are marked.

brane during the late gene expression phase is essential for a "clock"-based permeabilization of the membrane (14).

Integration site of ΦCD119. The integration site of the bacteriophage ΦCD119 was identified by using an inverse PCR approach. The divergent primers designed from the integrase gene (*int*) of the phage gave an 840-bp product, and sequencing the product yielded 629 nucleotides of the *C. difficile* sequence. This prophage-host junction was designated *attBP′*, which is the left end junction of phage and bacterial chromosomes. The bacterial *attBP′* sequence was used to design two more divergent primers, and the inverse PCR was repeated. This second PCR product yielded the *attPB′* sequence of the phage-host right end junction. Alignment of the two *att* site flanking sequences revealed a core sequence of 14 nucleotides (Fig. 5). The phage integrase mediates integrative and excisive site-specific recombination between these short homologous sequences located on the phage genome and the bacterial chromosome (19). Further analysis of the integration site revealed the integration of phage in an intergenic region between a hypothetical gene (Hyp) and the *gltP* gene in the bacterial chromosome. The relative position of this site in the *C. difficile* strain 630 genome has been noted (see Fig. 7B). The identified phage integration site was confirmed by Southern blot hybridization. The forward primer Hyp-Forward from the hypothetical gene and the reverse primer GltP-Reverse from the *gltP* gene were used in a PCR using the phage-sensitive strain 602 as a template. The PCR product was labeled with ^{32}P and used as a probe. The hybridization was performed with membrane-immobilized Tsp45I-digested chromosomal DNA isolated from strain 602 and 602/ΦCD119 lysogens. The two DNA-hybridized bands were detected only in DNA isolated from lysogens (Fig. 6). This result confirms the identified ΦCD119 integration site by inverse PCR.

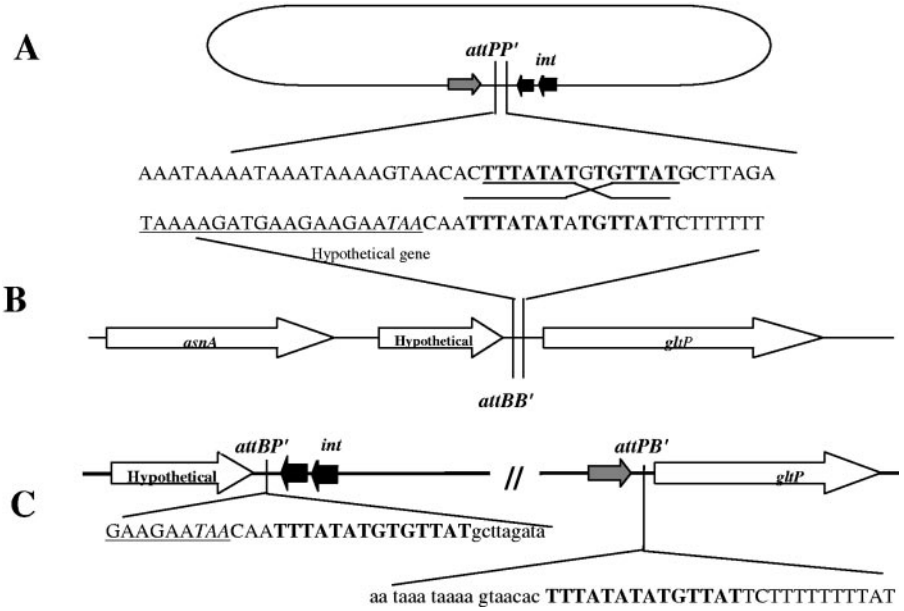

FIG. 5. Organization of bacterial and phage attachment sites. (A) Schematic representation of circularized phage genome with its *attPP′* site and nearby genes. (B) *C. difficile* genome showing *attBB′* site and surrounding genes. (C) Partial sequences of junctions showing the phage sequence in lowercase letters, the bacterial sequence in uppercase letters, and the homologous *att* site in boldface letters. The underlined sequence is the 3′ end of the hypothetical gene, and the stop codon is in italics.

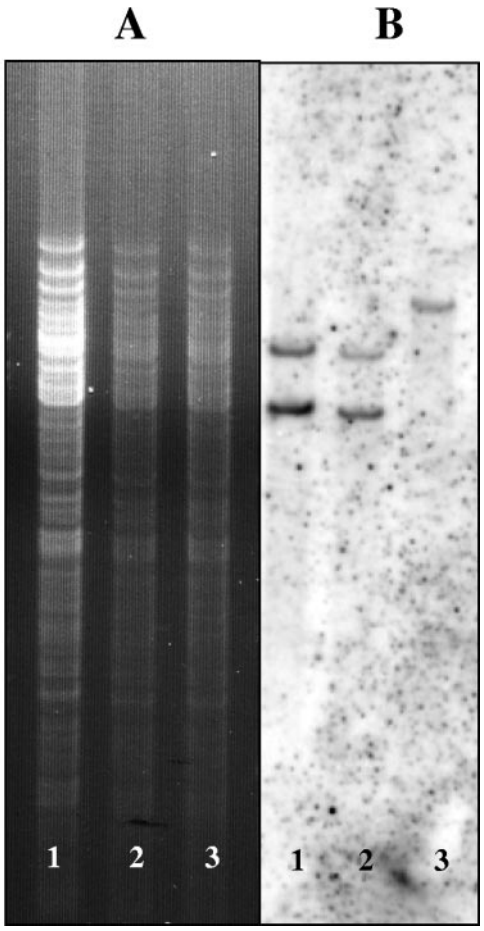

FIG. 6. Southern hybridization to confirm the identified integration site. (A) Ethidium bromide-stained gel with Tsp45I-digested genomic DNA. (B) Southern hybridization with the probe generated by PCR (see Materials and Methods) using primers overlapping the integration site. Lanes 1, 2, 602/ΦCD119 lysogens; lanes 3, 602 strain.

DISCUSSION

We have isolated a temperate phage from a pathogenic *C. difficile* strain and have sequenced and annotated its genome. ΦCD119 is a member of the *Myoviridae* and is the first *C. difficile* phage to have its genome sequenced. It possesses a circularly permuted double-stranded DNA genome carrying 79 putative ORFs, many of which exhibit similarities with proteins of other phages that infect gram-positive bacteria. A putative integrase (*int*) is present in ΦCD119, and the *attPP′* site is located close to the *int* gene (163 bp transcriptionally downstream). This is a common organization and has been used to develop site-specific integration vectors in some bacteria (19). Very few vector systems (15, 16, 25, 28) are available for *C. difficile*, and construction of an integration vector using ΦCD119 sequence information would be of considerable value for molecular and genetic research on this medically important pathogen. No ORF encoding an excisionase was identified in the ΦCD119 genome. However, the absence of an excisionase gene has been noted in other phages as well (18, 38). Several ORFs were unique to ΦCD119 and their predicted products did not match any of the proteins in the NCBI protein database (http://www.ncbi.nlm.nih.gov/BLAST/).

Blastn analysis, comparing the phage ΦCD119 nucleotide sequence with that of the *C. difficile* 630 genome (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_difficile), found the presence of two ΦCD119 sequence clusters (100% identical) (Fig. 7). One contains the DNA replication and recombination module, including the methylase genes, and the other contains the lysis module of ΦCD119. Located between these ΦCD119 clusters on the *C. difficile* chromosome are the partially characterized structural genes of *C. difficile* phage ΦC2 (13). This finding suggests that the prophage found in *C. difficile* strain 630 may be a mosaic of ΦC2- and ΦCD119-like phages.

It has been shown that genes from the PaLoc of *C. difficile* share homology with phage genes (7, 12, 33). For example, Tan et al. have demonstrated homology between *tcdE* and phage holin genes (33); Goh et al. (12) have also demonstrated cross-

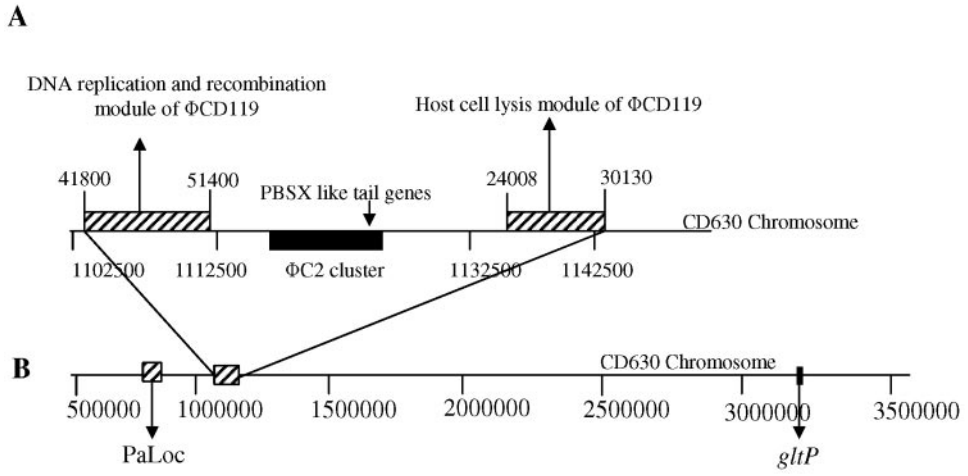

FIG. 7. Phage ΦCD119 nucleotide positions in *C. difficile* CD630 genome. (A) Phage ΦCD119 sequences (striped boxes) were located between nucleotide positions 1102700 and 1112251 and positions 1137425 and 1143549. The phage ΦC2 sequence cluster is marked as a filled box. (B) The PaLoc is shown located between nucleotide positions 786149 and 795379, approximately 308 kb from the ΦCD119 sequence cluster. The ΦCD119 integration site, near the *gltP* gene in strain 602, was not in close proximity to the PaLoc.

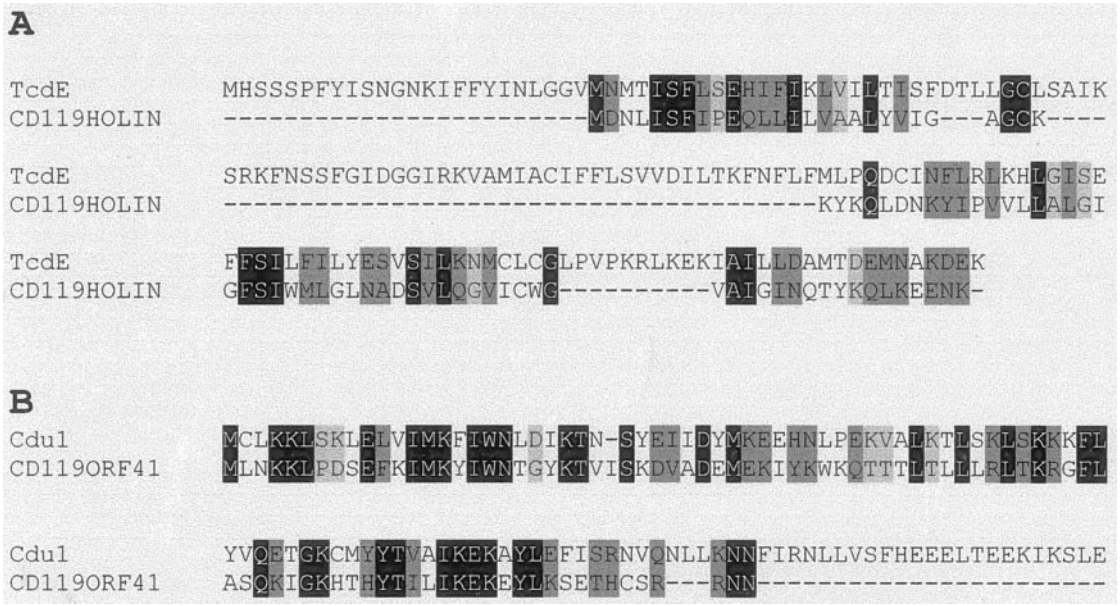

FIG. 8. (A) Similarity of ΦCD119 holin with *C. difficile* TcdE. (B) Alignment of ΦCD119 ORF 41 with Cdu1 of *C. difficile*. The sequences were aligned using ClustalW (http://www.ch.embnet.org/software/ClustalW.html) with default settings. Identical and similar amino acids are marked with black and gray, respectively.

reactivity of p32-labeled *tcdE* probe with *C. difficile* phage DNA. The toxin A gene (*tcdA*) has been reported to be homologous to a gene of phage ϕCT2 of *C. tetani* (7), and *tcdC*, a putative repressor in the *C. difficile* PaLoc, has been reported to have similarities with ORF 22 of *Lactobacillus casei* phage A2 (12). We have compared the ΦCD119 holin (ORF 34) with TcdE (ClustalW analysis) in *C. difficile* and found many common amino acid residues between these two proteins (Fig. 8A). The homology of *C. difficile* PaLoc-encoded *tcdE*, *tcdA*, and *tcdC* to phage sequences suggests that the PaLoc was once carried by phages.

To determine the role of ΦCD119 in the origin of the PaLoc, we compared the nucleotide sequences of ΦCD119 with that of the PaLoc. Our results indicate that no similarities exist between these sequences and neither the integration site of ΦCD119 nor the location of the ΦCD119 sequence cluster are in close proximity to the PaLoc in the *C. difficile* chromosome (Fig. 7B). We did find that a gene of ΦCD119, ORF 41, which resides next to the identified *attPP′*, matched (41% identity and 58% similarity) (Fig. 8B) with a *C. difficile* gene, Cdu1 (a putative penicillinase repressor), which resides next to the PaLoc integration site. However, the significance of this homology is not known. Hopefully, further characterization of *C. difficile* phages will provide a better understanding of the origin of the PaLoc of *C. difficile*.

Prophage genes of lysogens may control virulence factor production by host bacteria (35). We have identified several potential transcriptional regulators (ORF 41, 44, 45, 46, and 71) in the ΦCD119 genome. We are currently examining the mechanism by which these genes are being regulated and their influence, if any, on gene regulation and pathogenicity of *C. difficile*.

ACKNOWLEDGMENTS

The *C. difficile* CD119 lysogen F10 and the ΦCD119 phage infecting *C. difficile* strain 602 were obtained from Rosanna Dei, Universitá degli Studi di Firenze, Italy. We thank Mary Catherine for electron microscopy work and Susan San-Francisco and Ruwanthi Wettasinghe for help in sequencing. We also thank Julian Parkhill and other members of Sanger Centre for *C. difficile* 630 genome data made available online (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_difficile) before publication.

Sequencing of ΦCD119 was accomplished with support from NIH grant 5R03DK054816-01.

REFERENCES

1. **Ackermann, H. W.** 1998. Tailed bacteriophages: the order Caudovirales. Adv. Virus Res. **51:**135–201.
2. **Bartlett, J. G., T. W. Moon, N. T. Chang, and A. B. Onderdonk.** 1978. Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. Gastroenterology **75:**778–782.
3. **Black, L. W.** 1989. DNA packaging in dsDNA bacteriophages. Annu. Rev. Microbiol. **43:**267–292.
4. **Borriello, S.** 1990. Pathogenesis of *Clostridium difficile* infection of the gut. J. Med. Microbiol. **33:**207–215.
5. **Braun, V., T. Hundsberger, P. Leukel, M. Sauerborn, and C. Von Eichel Streiber.** 1996. Definition of the single integration site of the pathogenicity locus in *Clostridium difficile*. Gene **27:**29–38.
6. **Brussow, H., and R. W. Hendrix.** 2002. Phage genomics: small is beautiful. Cell **108:**13–16.
7. **Canchaya, C., F. Desiere, W. M. McShan, J. J. Ferretti, J. Parkhill, and H. Brussow.** 2002. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. Virology **302:**245–258.
8. **Catalano, C. E.** 2000. The terminase enzyme from bacteriophage lambda: a DNA-packaging machine. Cell. Mol. Life Sci. **57:**128–148.
9. **Droge, A., M. A. Santos, A. C. Stiege, J. C. Alonso, R. Lurz, T. A. Trautner, and P. Tavares.** 2000. Shape and DNA packaging activity of bacteriophage SPP1 procapsid: protein components and interactions during assembly. J. Mol. Biol. **296:**117–132.
10. **Eklund, M. W., F. T. Poysky, S. M. Reed, and C. A. Smith.** 1971. Bacteriophage and the toxicity of *Clostridium botulinum* type C. Science **172:**480–482.
11. **Finn, C. W., R. P. Silver, W. H. Habig, M. C. Hardegree, G. Zen, and C. F. Gardon.** 1984. The structural gene for tetanus neurotoxin is on a plasmid. Science **224:**881–884.

12. **Goh, S., B. J. Chang, and T. V. Riley.** 2005. Effect of phage infection on toxin production by *Clostridium difficile*. J. Med. Microbiol. **54:**129–135.
13. **Goh, S., T. V. Riley, and B. J. Chang.** 2005. Isolation and characterization of temperate bacteriophages of *Clostridium difficile*. Appl. Environ. Microbiol. **71:**1079–1083.
14. **Grundling, A., M. D. Manson, and R. Young.** 2001. Holins kill without warning. Proc. Natl. Acad. Sci. USA **98:**9348–9352.
15. **Haraldsen, J. D., and A. L. Sonenshein.** 2003. Efficient sporulation in *Clostridium difficile* requires disruption of the sigmaK gene. Mol. Microbiol. **48:**811–821.
16. **Herbert, M., T. A. O'Keeffe, D. Purdy, M. Elmore, and N. P. Minton.** 2003. Gene transfer into *Clostridium difficile* CD630 and characterisation of its methylase genes. FEMS Microbiol. Lett. **229:**103–110.
17. **Katsura, I.** 1987. Determination of bacteriophage length by protein ruler. Nature **327:**73–75.
18. **Kropinski, A. M.** 2000. Sequence of the temperate serotype converting *Pseudomonas aeruginosa* bacteriophage D3. J. Bacteriol. **182:**6066–6074.
19. **Lauer, P., M. Y. Chow, M. J. Loessner, D. A. Portnoy, and R. Calendar.** 2002. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. J. Bacteriol. **184:**4177–4186.
20. **Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar.** 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. **35:**324–340.
21. **Mahony, D. E., J. Clow, L. Atkinson, N. Vakharia, and W. F. Schlech.** 1991. Development and application of a multiple typing system for *Clostridium difficile*. Appl. Environ. Microbiol. **57:**1873–1879.
22. **Mahony, D. E., P. D. Bell, and K. B. Easterbrook.** 1985. Two bacteriophages of *Clostridium difficile*. J. Clin. Microbiol. **21:**251–254.
23. **Marinus, M. G.** 1996. Methylation of DNA, p. 697–702. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 1. ASM Press, Washington, D.C.
24. **Mukherjee, K., S. Karlsson, L. G. Burman, and T. Akerlund.** 2002. Proteins released during high toxin production in *Clostridium difficile*. Microbiology **148:**2245–2253.
25. **Mullany, P., M. Wilks, L. Puckey, and S. Tabaqchali.** 1994. Gene cloning in *Clostridium difficile* using Tn916 as a shuttle conjugative transposon. Plasmid **31:**320–323.
26. **Ochman, H., A. S. Gerber, and D. L. Hartl.** 1988. Genetic applications of an inverse polymerase chain reaction. Genetics **120:**621–623.
27. **Plunkett, G., III, D. J. Rose, T. J. Durfee, and F. R. Blattner.** 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. **181:**1767–1778.
28. **Purdy, D., T. A. O'Keeffe, M. Elmore, M. Herbert, A. McLeod, M. Bokori-Brown, A. Ostrowski, and N. P. Minton.** 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Mol. Microbiol. **46:**439–452.
29. **Rentas, F. J., and V. B. Rao.** 2003. Defining the bacteriophage T4 DNA packaging machine: evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17. J. Mol. Biol. **14:**37–52.
30. **Romero, P., R. Lopez, and E. Garcia.** 2004. Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus. Virology **322:**239–252.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 2001. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
32. **Seaman, E., E. Tarmy, and J. Marmur.** 1964. Inducible phages of *Bacillus subtilis*. Biochemistry **3:**607–612.
33. **Tan, K. S., B. Y. Wee, and K. P. Song.** 2001. Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. J. Med. Microbiol. **50:**613–619.
34. **Vasala, A., M. Valkkila, J. Caldentey, and T. Alatossava.** 1995. Genetic and biochemical characterization of the *Lactobacillus delbrueckii* subsp. lactis bacteriophage LL-H lysin. Appl. Environ. Microbiol. **61:**4004–4011.
35. **Wagner, P. L., J. Livny, M. N. Neely, D. W. Acheson, D. I. Friedman, and M. K. Waldor.** 2002. Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. Mol. Microbiol. **44:**957–970.
36. **Young, R.** 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. **56:**430–481.
37. **Young, R., and U. Blasi.** 1995. Holins: form and function in bacteriophage lysis. FEMS Microbiol. Rev. **17:**191–205.
38. **Zimmer, M., S. Scherer, and M. J. Loessner.** 2002. Genomic analysis of *Clostridium perfringens* bacteriophage ϕ3626, which integrates into *guaA* and possibly affects sporulation. J. Bacteriol. **184:**4359–4368.