

The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage

Gill *et al.*

RESEARCH ARTICLE

Open Access

The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage

Jason J Gill^{1,2}, Joel D Berry^{1,2}, William K Russell³, Lauren Lessor^{1,2}, Diego A Escobar-Garcia^{1,2}, Daniel Hernandez², Ashley Kane², Jennifer Keene², Matthew Maddox², Rebecca Martin², Sheba Mohan², Ashlyn M Thorn², David H Russell³ and Ry Young^{1,2*}

Abstract

Background: The bacterium *Caulobacter crescentus* is a popular model for the study of cell cycle regulation and senescence. The large prolate siphophage phiCbK has been an important tool in *C. crescentus* biology, and has been studied in its own right as a model for viral morphogenesis. Although a system of some interest, to date little genomic information is available on phiCbK or its relatives.

Results: Five novel phiCbK-like *C. crescentus* bacteriophages, CcrMagneto, CcrSwift, CcrKarma, CcrRogue and CcrColossus, were isolated from the environment. The genomes of phage phiCbK and these five environmental phage isolates were obtained by 454 pyrosequencing. The phiCbK-like phage genomes range in size from 205 kb encoding 318 proteins (phiCbK) to 280 kb encoding 448 proteins (CcrColossus), and were found to contain nonpermuted terminal redundancies of 10 to 17 kb. A novel method of terminal ligation was developed to map genomic termini, which confirmed termini predicted by coverage analysis. This suggests that sequence coverage discontinuities may be useable as predictors of genomic termini in phage genomes. Genomic modules encoding virion morphogenesis, lysis and DNA replication proteins were identified. The phiCbK-like phages were also found to encode a number of intriguing proteins; all contain a clearly T7-like DNA polymerase, and five of the six encode a possible homolog of the *C. crescentus* cell cycle regulator GcrA, which may allow the phage to alter the host cell's replicative state. The structural proteome of phage phiCbK was determined, identifying the portal, major and minor capsid proteins, the tail tape measure and possible tail fiber proteins. All six phage genomes are clearly related; phiCbK, CcrMagneto, CcrSwift, CcrKarma and CcrRogue form a group related at the DNA level, while CcrColossus is more diverged but retains significant similarity at the protein level.

Conclusions: Due to their lack of any apparent relationship to other described phages, this group is proposed as the founding cohort of a new phage type, the phiCbK-like phages. This work will serve as a foundation for future studies on morphogenesis, infection and phage-host interactions in *C. crescentus*.

Keywords: Bacteriophage, Genomics, *Caulobacter crescentus*, phiCbK

Background

For over 40 years, the alpha-proteobacterium *Caulobacter crescentus* has been an important model organism for the study of bacterial development, physiology and cell cycle biology. *C. crescentus* exhibits a cyclical, dimorphic lifestyle that is atypical among prokaryotes [1,2].

Its sessile form displays an adhesive polar holdfast, or stalk, and this cell type is exclusively capable of DNA replication and cell division. Cell division in *C. crescentus* is asymmetrical, and stalked cells divide to produce motile daughter "swarmer" cells with a single polar flagellum and multiple polar pili. The swarmer cells are unable to divide or replicate their DNA until they shed their flagellum and pili and undergo a physical transformation to the stalked cell morphotype. *C. crescentus* is also unusual in that its DNA replication is closely coordinated with cell division, resulting in the production of a single copy

* Correspondence: ryland@tamu.edu

¹Center for Phage Technology, 2128 TAMU, Texas A&M University, College Station, Texas, TX 77843, USA

²Dept. of Biochemistry & Biophysics, 2128 TAMU, Texas A&M University, College Station, Texas, TX 77843, USA

Full list of author information is available at the end of the article

of the bacterial chromosome per division cycle. The regulatory networks that control differentiation and division have been well characterized [3-5].

Caulobacter phages were first isolated nearly 50 years ago [6] and have been instrumental as tools for genetic transduction [7,8] and as probes for the presence of cell-cycle specific markers [9-11]. Among the phages of *C. crescentus*, the most extensively studied are large, virulent siphophages with prolate heads, among which phiCbK is the archetype. First reported in 1970 [12], phiCbK has been used as a cell-cycle and morphological indicator [10] because it uses the flagellum and polar pili of the swarmer cell for adsorption [13]. In addition, phiCbK, due to its large size, was one of the earliest phages for which the fine structure of the capsid was determined by electron microscopic image reconstruction [14-16] and was the first instance where the phage head-tail symmetry mismatch was demonstrated [17]. Phage phiCbK and related phages have been shown to have a unique filamentous structure emanating from the top vertex of the capsid [14,18]. This head filament was found to mediate primary attachment of the virion to the flagellum and thus more efficient adsorption to the swarmer cell [18]. However, despite the biological and structural significance of *C. crescentus* phages, little is known about the biology of the phages themselves. Here we report the complete genomes of phage phiCbK and five related *Caulobacter* siphophages. The results are discussed in terms of the unique structure of these phages and the biological imperatives facing phages that infect bacterial species with dimorphic cell types.

Methods

Phage isolation and culture

Phage phiCbK and *C. crescentus* strain CB15 were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (Université Laval, QC, Canada), and strain CB15 was used for the enrichment and propagation of all phage isolates. *C. crescentus* was cultured at 30°C with aeration in PYE broth (2 g/L peptone (Oxoid), 1 g/L yeast extract (Difco), 0.1 g/L anhydrous MgSO₄) or PYE agar (PYE broth plus 15 g/L Bacto agar). Phages were propagated and enumerated on PYE plates by the soft agar overlay method [19] using lawns consisting of 4 ml PYE top agar (PYE broth plus 5 g/L Bacto agar) and inoculated with 100 µl of an overnight PYE culture of *C. crescentus* CB15. After plating, lawns were incubated for 42–48 h at 30°C prior to plaque enumeration or harvesting.

Phages other than phiCbK were isolated in early 2010 from surface water samples collected in Bryan and College Station, TX, USA by students enrolled in the Phage Genomics for Undergraduates program run at Texas A&M University. Phages were isolated following culture

enrichment or direct concentration methods. In culture enrichment of water samples, 40 ml of filter-sterilized water sample (0.22 µm, Millipore) was added to 10 ml of 5X strength PYE broth, inoculated with 100 µl of a fresh *C. crescentus* CB15 overnight PYE culture and incubated with aeration overnight at 30°C. Enrichment cultures were centrifuged (8,000 x g, 10 min, 4°C), the supernatants filter sterilized (0.22 µm) and plated to lawns of *C. crescentus* CB15, and observed for plaque formation. The direct concentration method was modified from a technique kindly provided by R. Hendrix, University of Pittsburgh (personal communication). Briefly, 1 L of water sample was clarified by filtration through Whatman 597½ paper (Whatman). Five grams of Whatman DE-52 anion exchange resin was added and incubated at 22°C for 30 min with shaking. The resin was collected in a 50 ml centrifuge tube and centrifuged at 2,000 x g, 2 min, 22°C and the supernatant discarded. The resin was washed twice by resuspension in 45 ml wash buffer (25 mM Tris-HCl, pH 7.5, 10 mM NaCl, 5 mM MgSO₄) followed by centrifugation as above. Phage were eluted from the resin by addition of 30 ml elution buffer (25 mM Tris-HCl, pH 7.5, 600 mM NaCl, 5 mM MgSO₄) followed by centrifugation as above, and the supernatant was retained and concentrated to a final volume of ~500 µl in a 100 kDa NWCO centrifugal ultrafiltration device (Millipore). This total phage concentrate was then plated to lawns of *C. crescentus* CB15 and observed for plaque formation. Individual plaques were picked and subcultured three times, then propagated to high-titer lysates in soft agar overlays [19].

Phage DNA preparation and sequencing

Bacteriophage genomic DNA was prepared from 10–20 ml of filter-sterilized, high-titer (> 1 x 10⁹ PFU/ml) phage lysates using a modified form of the Promega Wizard DNA clean-up kit (Promega) as described previously [20]. DNA integrity was verified by running on a 0.8% agarose gel and staining with ethidium bromide and DNA was quantified by band densitometry. Phage genome size was estimated by pulsed field gel electrophoresis (PFGE) analysis of genomic DNA on a 1% agarose gel (Pulsed-Field agarose, BioRad) and comparison to a size marker (Lambda Ladder PFG Marker, New England Biolabs).

Phage genomic DNA was mixed in equimolar amounts and sequenced as MID-labeled pools by 454 pyrosequencing (Roche) at the Emory GRA Genome Center (Emory University, GA, USA). Trimmed FLX Titanium flowgram outputs were assembled using the Newbler assembler version 2.0.01.14 or 2.5.3 (454 Life Sciences) at default settings. Contigs were confirmed to be complete by PCR, using primers that faced off each end of the contigs and sequencing of the resulting products. Phage

phiCbK was sequenced to 181-fold average coverage, Colossus to 74-fold coverage, Rogue to 38-fold coverage, Swift to 59-fold coverage, Karma to 291-fold coverage and Magneto to 83-fold coverage. In all cases, the phage genomes produced circular assemblies. Breakpoints in sequencing coverage were determined by manual inspection of contigs in CLC Workbench version 6.2 (CLC bio).

Terminal labeling of genomic DNA

In order to locate the physical termini of the phiCbK chromosome, whole phage genomic DNA was ligated with a short oligonucleotide tag of known sequence and this product resequenced by pyrosequencing as described above. In principle, the oligo tag will ligate to the physical chromosomal termini and the boundary between the tag sequence and the phage genomic sequence will indicate the original physical ends of the phage chromosome. Genomic DNA of phage phiCbK was end-repaired with the NEBNext End Repair Module (NEB) using 1 µg of DNA and 1 µl of enzyme in a 100 µl reaction according to the manufacturer's protocol. End-repaired DNA was precipitated by addition of 200 mM NaCl and 3 volumes of ethanol, and resuspended in 20 µl water. A 49 bp dsDNA oligonucleotide (5' - TTA CTTACAATCCTTGG CCGTTTTGCTGCGCGCCCATGATGGACTGGAC - 3') was added to the genomic DNA at a 25:1 molar ratio and ligated at 16°C for 18 h with T4 DNA ligase (NEB) in a 50 µl reaction volume. The ligase was heat inactivated (65°C, 10 min), DNA precipitated with NaCl and ethanol as described above, resuspended in 20 µl water and submitted for sequencing by 454 pyrosequencing. All reads from this resequencing run containing the complete 49 bp oligo sequence were extracted from the sequencing output and assembled onto the previously established phiCbK genomic sequence to locate hotspots of tag ligation.

Genome annotation

Genes were predicted using GeneMark.hmm [21] and Glimmer 3 [22] and gene starts manually edited in Artemis [23]. tRNA genes were predicted by tRNAscan-SE 1.21 [24] and Rfam [25], and Rho-independent terminators predicted by TransTermHP [26]. Proteins were exported and analyzed in batch by BLASTp [27]; protein functional characteristics were predicted by batch analysis of all proteins in InterProScan version 4.7 [28]. Proteins of particular interest were additionally analyzed by HHpred searches against the pdb70_18Aug12 database [29], TMHMM (www.cbs.dtu.dk/services/TMHMM), LipoP 1.0 [30], and the EMBOSS package [31]. Inter-genomic protein comparisons were conducted via RAST [32] using phiCbK or Colossus as the reference genome and a similarity cutoff of 25%, and variably conserved proteins were grouped by NCBI BLASTClust at settings

of $L=0.9$, $S=30$. Figures were generated using Circos version 0.55 [33] and DNA Master (cobamide2.bio.pitt.edu/computer.htm).

Sequence deposition

The names of novel phage isolates were prefixed with Ccr, which is the ReBase species acronym for *C. crescentus* (rebase.neb.com). Completed phage genomes were deposited in GenBank under the following accession numbers: phiCbK, JX100813; CcrMagneto, JX100812; CcrSwift, JX100809; CcrKarma, JX100811; CcrRogue, JX100814; CcrColossus, JX100810. For simplicity, the phage names will be used without the Ccr prefix.

Proteomic analysis

Phage phiCbK was propagated in 2 L PYE liquid culture at 30°C to a titer of $\sim 10^{10}$ PFU/ml. Phage lysate was clarified by two rounds of centrifugation (8,000 x g, 10 min, 4°C) and the phage-containing supernatant was digested with 1 µg/ml DNase I and RNase A (Sigma) for 2 h at 22°C. Phage was concentrated by pelleting in a centrifuge (6,000 x g, 18 h, 4°C) and the pellet was resuspended in λ -dil buffer (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM MgSO₄). Residual cell debris was removed from the phage suspension by centrifugation (12,000 x g, 15 min, 4°C) and the phage were then banded by isopycnic CsCl gradient ultracentrifugation, dialyzed and concentrated as described previously [34]. The phage preparation was diluted in λ -dil, boiled for 5 min and incubated with 2 µg/ml DNase I for 1 h at 37°C to reduce sample viscosity. Samples were then mixed with Laemmli sample buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 5% β -mercaptoethanol, 0.001% bromophenol blue), boiled for 5 min and the equivalent of 1×10^{10} to 5×10^{10} PFU/lane were run on a 10% Tris-glycine SDS-PAGE gel by standard methods [35]. Bands were visualized with Coomassie blue, and the entire gel lane with the exception of the dye front was segmented based on band boundaries. All gel segments, containing visible bands or unstained inter-band regions, were subjected to proteomic analysis as described previously [34]. Alternatively, digested samples were injected into a Waters nanoACQUITY system for UPLC separation of peptic peptides. Separation was achieved after the peptides were trapped and desalted on a VanGuard Pre-Column trap (2.1 x 5 mm, ACQUITY UPLC BEH C18, 1.7 µm) for 3 min. Peptides were eluted from the trap using a 2%–40% linear gradient of acetonitrile over 32 min at a flow rate of 0.40 µl/min and were separated using an ACQUITY UPLC BEH C18 1.7 µm 1.0 x 100 mm column. Peptides that were produced from the enzymatic cleavage were identified from the using Waters MS^E technology on a Waters Synapte G2

instrument and ProteinLynx Global Server (PLGS) searches of a customized database.

Transmission electron microscopy

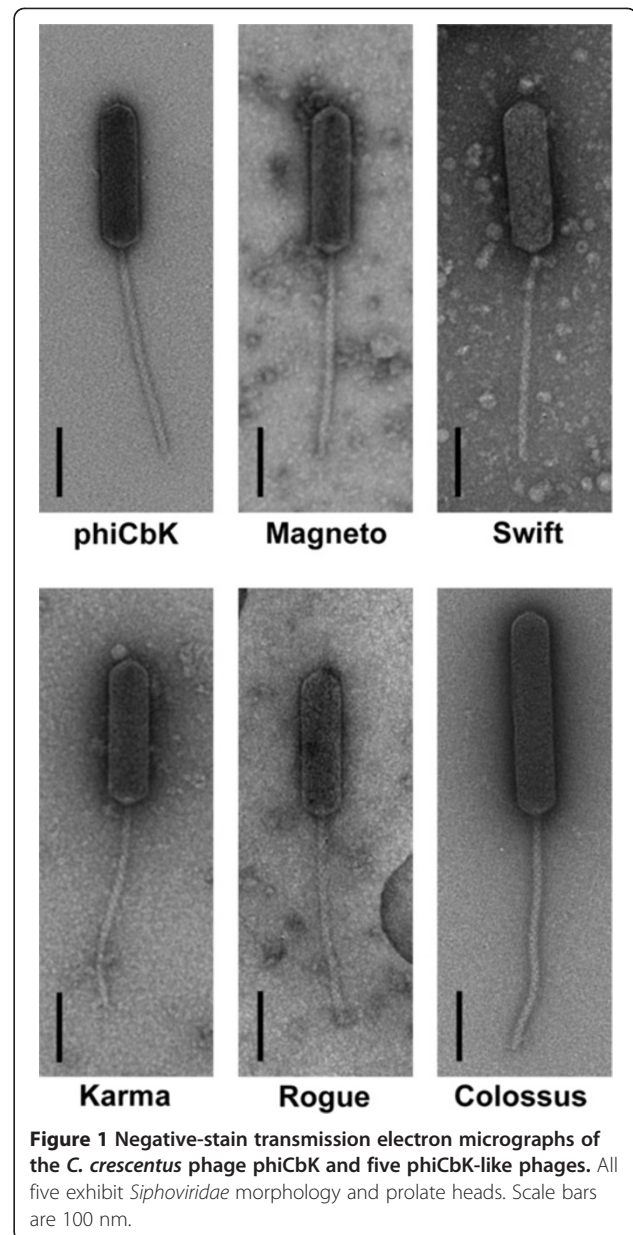
Phages were prepared for microscopy by the Valentine method [36] and stained with 2% (w/v) uranyl acetate. Grids were viewed in a JEOL 1200 EX transmission electron microscope under 100 kV accelerating voltage. Five virions of each phage were measured and these data used to calculate mean dimensions.

Results and discussion

Phage and genome characteristics

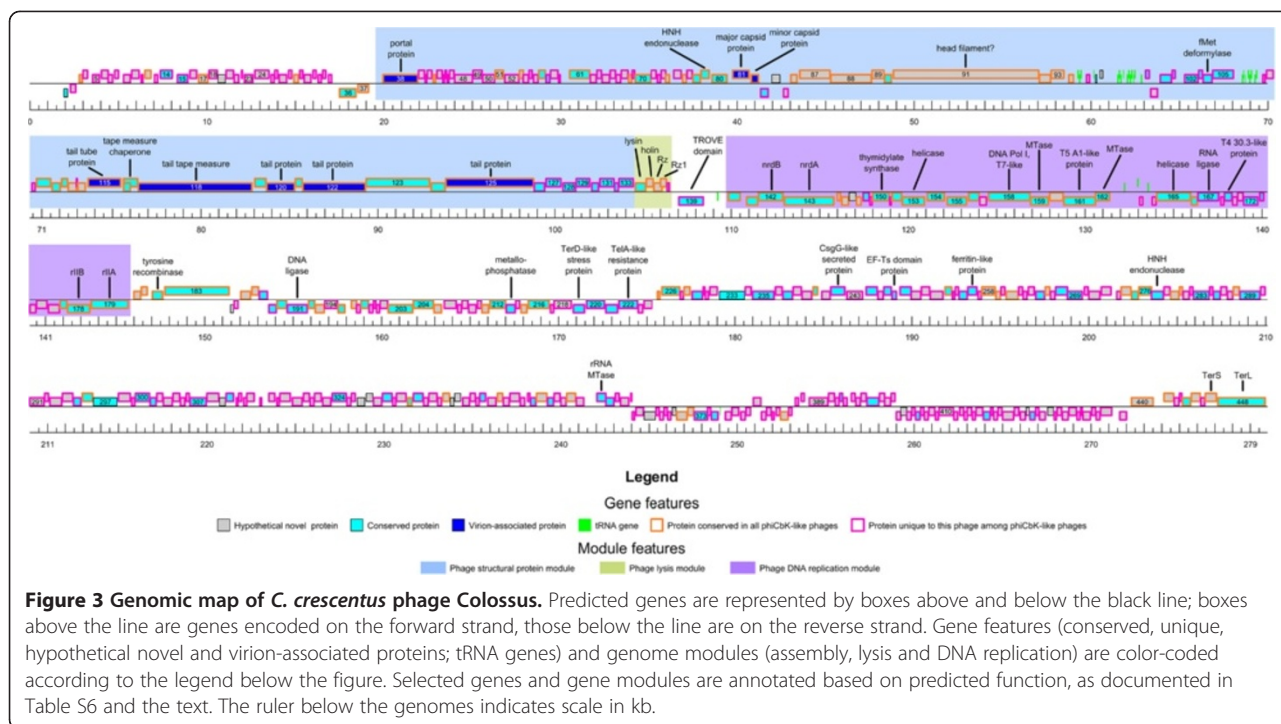
Phages Magneto, Swift, Karma, Rogue and Colossus were isolated from surface waters; Magneto, Swift, Rogue and Colossus were obtained by the culture enrichment method, and phage Karma was isolated by the direct concentration method. Electron microscopic analysis of these phages revealed that all five possessed similar *Siphoviridae* morphology to the previously described phiCbK, with long, non-contractile tails and large prolate heads of varying lengths (Figure 1). The phage phiCbK head measured 205 nm long and 56 nm wide with a tail length of 300 nm, dimensions which are in good agreement with previously reported measurements [14,15,18]. Phage dimensions are summarized in Table 1; all of the phages exhibited similar dimensions to that of phage phiCbK except for Colossus, which had a considerably longer head and slightly longer tail than the other phages.

The genomes of these six phages were sequenced to completion by 454 pyrosequencing followed by manual closure, representing a total combined DNA sequence of 1.33 Mb. The general characteristics of the phage genomes are summarized in Table 1, complete annotations and supporting evidence are provided as supplementary data in Additional file 1: Tables S1 - S6. The unit genomes of these phages ranged from 205.4 kb encoding 318 protein-coding genes (phiCbK) to 280 kb encoding 448 protein-coding genes (Colossus), with terminal direct repeats of ~10 kb in all phages except Colossus, in which the terminal redundancy appeared to be closer to ~17 kb. Three major modules can be identified in these genomes (Figure 2, Figure 3). The phage structural protein module begins with the gene encoding the portal protein (gp42 in phiCbK, gp38 in Colossus) at an apparent divergent transcription site and extends to slightly after a tail protein gene on the plus strand. A lysis cassette, encoding an endolysin, a holin and a two-component spanin protein immediately follows the structural module on the same strand. A DNA metabolism and replication module is encoded on the minus strand, beginning at another divergent transcription site with an *rIIA/rIIB* gene pair (gp138/137 in phiCbK,



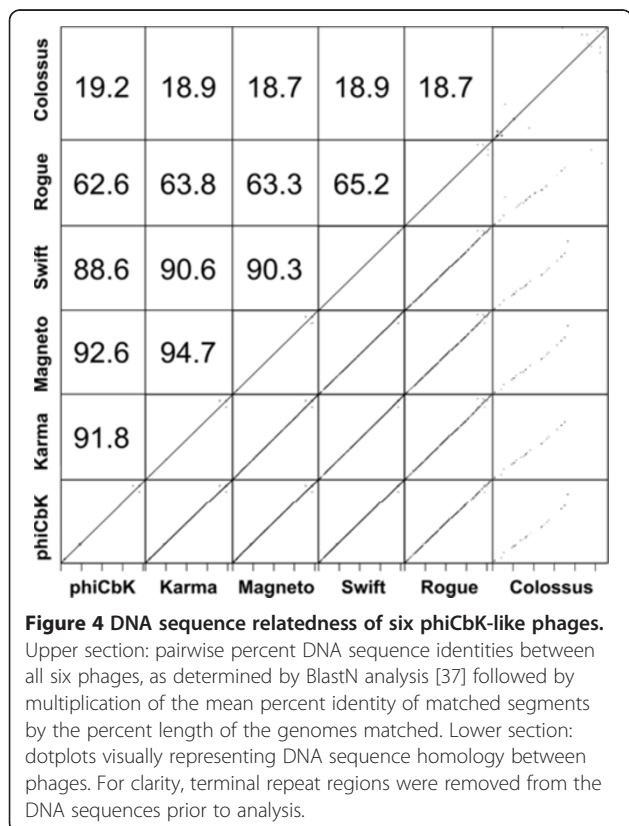
gp179/178 in Colossus) and extending to the end of the lysis cassette. These genomic features are described in greater detail in the following sections.

The GC content of the genomes were all close to the 67.2% GC content of *C. crescentus* CB15. Each phage also encodes a large number of tRNA genes, ranging from 23 in Rogue to 28 found in Colossus; Rfam analysis did not detect any functional RNA elements other than tRNAs. Each phage encodes tRNAs specific for 13 to 16 amino acids; Thr and Tyr tRNAs are not found in any phage, Asn and His are found only in Colossus, and Ala is found only in Magneto, Swift and Colossus. Phage phiCbK is the only phage to lack an Arg tRNA, while Rogue lacks Asp and Ser. Codon usage was



to the other phages, the protein-coding genes of Colossus were compared to phage phiCbK separately. Of 448 predicted proteins in phage Colossus, 307 (68%) do not have any homologs in the other five phiCbK-like phages. Nevertheless, as shown in Figure 6, Colossus is still largely syntenic with phiCbK. The regions of greatest protein conservation occur in the central portion of the genomes, which contain the predicted structural and DNA replication proteins (Figure 3). Large sections at the left and right ends of the genomes, primarily containing proteins of unknown function, are less well conserved. These sections contain the majority of gene insertions and deletions, and also several apparent gene translocations and duplications (Figure 6). Phage phiCbK genes 68, 99 and 176 appear to be directly duplicated in Colossus, and Colossus genes 212 and 346 are duplicated in phiCbK. In the case of phiCbK gp68 one of the duplicates, Colossus gene 36, is also significantly diverged and translocated to a position ~21 kb upstream of its paralog, the Colossus major capsid protein gene 81 (see below).

While these phages are related to one another, a striking feature of the phiCbK-like phages is the lack of any meaningful relationship to other described organisms, including other phages. In the case of phage phiCbK, 239 of its 318 predicted proteins, or 75%, have no matches (E value >1 x 10⁻⁵) to proteins in the NCBI nr database. Similarly, in phage Colossus, 310 proteins (69.2%) do not have any detectable homologs in the nr database. The phage most closely related to phiCbK is PhiJL001 (YP_224010), a slightly prolate siphophage that



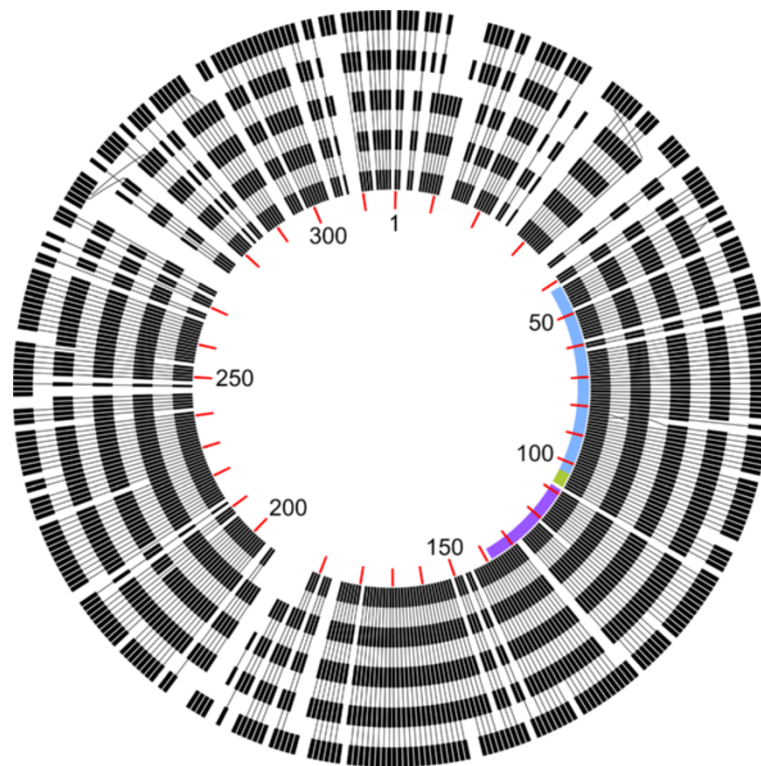


Figure 5 Non-proportional synteny map showing the relationships between related *C. crescentus* phages at the protein level. The black blocks represent protein-coding genes in the order they appear in each phage genome, starting with gp1 at the top and running clockwise, with red tick marks indicating 10-gene intervals in phage phiCbK. Black lines connecting blocks indicate similarity of proteins between phages. From innermost to outermost, tracks represent phages phiCbK, Karma, Magneto, Swift and Rogue. Blue, green and purple arcs on the inside track indicate the boundaries of the phage morphogenesis, lysis and DNA replication modules, respectively. Terminal repeat regions were excluded from this figure for clarity.

infects an uncharacterized marine *Alphaproteobacterium* strain [38]. This relationship however is extremely distant, as even in this case only seven phiCbK proteins, (gp42, gp96, gp97, gp98, gp99, gp118, gp126) mostly located in the phage tail structural region, are detectably related to PhiJL001 proteins, with 14.3 – 40.3% identity. These proteins in themselves do not appear to form a cohesive evolutionary module; for example the PhiJL001 portal homolog gp60 (YP_223984) is most closely related to proteins found in *Bordetella* genomes, while the tail protein homolog gp84 (YP_224008) has a homolog located in *Polymorphum gilvum*. Three of these proteins, gp96, gp97 and gp99, are related to gene transfer agent proteins orfg12 (ABK27260), orfg13 (ABK27261) and orfg15 (ABK27263) of *Rhodobacter capsulatus* [39] (with 42.3%, 17.2% and 15.9% similarity, respectively), suggesting some very distant relationship between the phiCbK-like phages and these phage-like gene transfer agents.

Due to frequent horizontal gene transfer events, the concept of biological species or hierarchical lineages in the classical Linnaean sense are difficult to apply to

phages [40,41]. The model of “phage type” has been proposed, which groups phages based on shared features of genome organization and function [42]. Protein or DNA sequence relationships have been used to organize phages into groups that appear to have biological meaning (e.g., [43-45]), however in all of these cases the enormous diversity of phages results in numerous phage groups with no obvious evolutionary linkage to each other. While the phiCbK-like phages lack obvious relationships to any other described organisms, they are obviously related to each other; 119 proteins were found to be conserved in all six genomes. These conserved proteins form syntenic blocks and are primarily located in the phage structural, lysis and DNA replication modules (Figure 2, Figure 3), suggesting these constitute the core genes of the phiCbK phage type. Given the lack of any clear relationship to any other described phages, we propose phage phiCbK as the founder of a novel phage type, the phiCbK-like phages, which includes phages Magneto, Swift, Karma, Rogue and Colossus as its inaugural members, with the latter perhaps the founder of a sub-type.

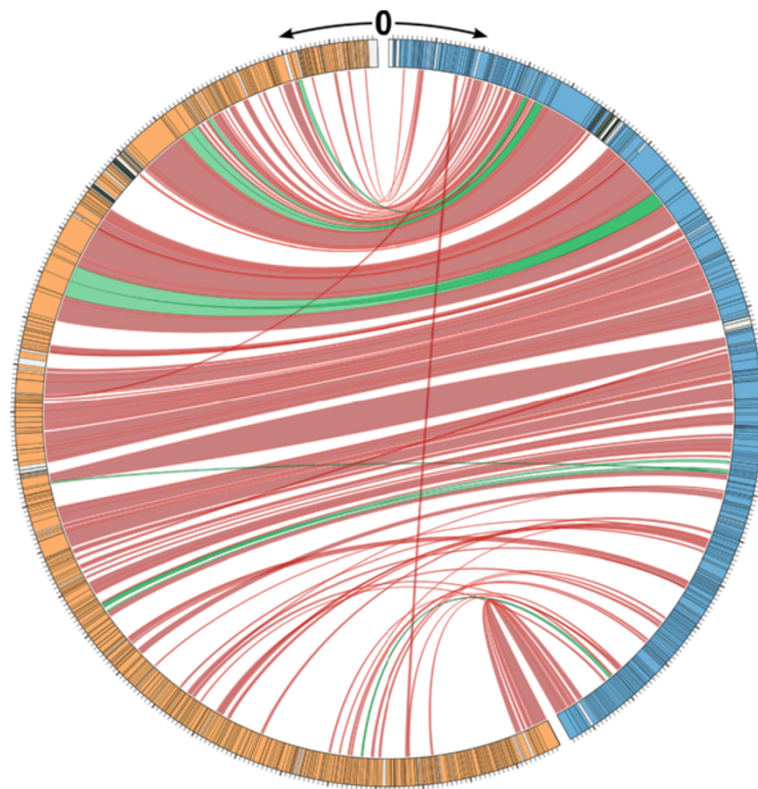


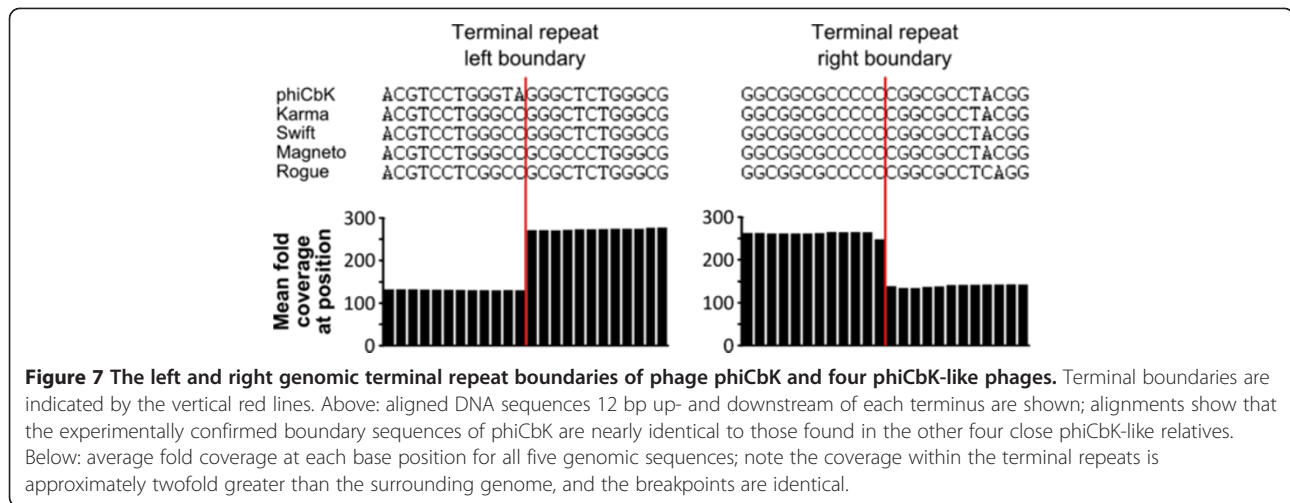
Figure 6 Map showing the relationship between *C. crescentus* phages phiCbK and Colossus at the protein level. Both genomes start at position 0 at the top of the figure and move down, with Colossus in orange on the left and phiCbK in blue on the right. Black bands in each track denote the boundaries of protein-coding genes, grey areas represent non-coding regions. Grey tick marks on the outside edge represent 1 kb of sequence, with heavier ticks representing 10 kb. Proteins present in both phages are connected by red ribbons between the two genomes; green ribbons mark proteins with more than one homolog in the other phage, suggesting gene duplications. For clarity, terminal repeat regions were excluded from this figure.

Determination of genomic termini

Following closure, the phage phiCbK genome produced a circular assembly of 205,423 bp. However, the sequence contigs assembled from pyrosequencing reads possessed a 10,287 bp region of significantly higher sequence coverage than the rest of the genome. At the boundaries of this region, sequence coverage abruptly transitioned from approximately 170-fold to over 360-fold between two base positions, and coverage of this region was on average 2.2-fold greater than the rest of the sequence. This suggested that the high coverage region was a large terminal repeat of ~10 kbp, as found in the classical coliphage T5 [46]. To test this notion, the phiCbK genomic DNA was subjected to terminal labeling analysis to determine the physical ends of the phage genome. This procedure retrieved 52 pyrosequencing reads containing the complete oligo tag sequence, of which 41 assembled to the established phiCbK genome. Fifteen of these 41 reads assembled to two distinct loci on the phiCbK genome, with the boundaries of the oligo sequence and the phage genomic sequence corresponding exactly to

the boundaries of the high-coverage region identified in the phiCbK assembly. The remaining 26 sequencing reads assembled to various loci across the genome, which we interpret as tag ligation to double-stranded DNA breaks generated during the DNA isolation procedure, or possibly termini generated by aberrant packaging events. These experimental data were interpreted as a confirmation of the presence and location of long direct terminal repeats in the phiCbK genome, and the genome was therefore reopened accordingly as shown in Figure 2.

The left terminus of the physical genome lies within codon 652 of the predicted terminase large subunit gene, while the right terminus resides in a predicted non-coding region between genes 20 and 21, resulting in a 10,287 bp direct repeat. Phages Swift, Magneto, Karma and Rogue exhibited similar abrupt discontinuities in sequencing coverage to that of phage phiCbK, indicating that these phages possess direct terminal repeats similar to that of phiCbK. In all five cases these coverage transition boundaries were of identical or nearly identical nucleotide sequence to that found in phiCbK (Figure 7),



with the left boundary in all cases lying within codon 652 of the large terminase gene, and the right boundary within an intergenic region ~10 kb downstream. Given these similarities, the genomes of phages Swift, Magneto, Karma and Rogue were also reopened and annotated to reflect the presence of these repeats. The sizes of the terminal repeat regions in these phage genomes are summarized in Table 1.

Like the other phiCbK-like phages, Colossus exhibited an abrupt transition at the left end of a high-coverage region, but instead of an equally abrupt reduction at the end of a discrete length of sequence, a more gradual transition was observed, spanning ~ 200 bp from high to normal coverage on the right end. In this case the left boundary is in a non-coding region slightly downstream of the large terminase gene, and the right boundary lies within gene 34, some 16,700 bp downstream. Thus, Colossus possesses a non-permuted terminal repeat like the other phiCbK-like phages, marking the initiation site for genome packaging. However, unlike the more closely related phiCbK-like phages, where packaging terminates at the end of the terminal repeat, genome packaging in Colossus appears to have a more imprecise termination and cleavage mechanism.

While all *Caudovirales* phages package their DNA into the capsid as a linear molecule, the nature of the genomic termini can be markedly different between phage types [47]. Phage genomes may have non-permuted termini with short 3' or 5' overhangs (like phage lambda), circularly permuted terminal redundancies (like phage T4), or non-permuted direct terminal repeats which can be either short (like phage T7) or long (like phage T5). Some phages, like Mu, have host DNA at their termini, and the phi29-like phages have a covalently linked protein. Characterizing these diverse modalities for DNA packaging is crucial for understanding the biology of phages.

In addition, there is practical significance, in that these mechanisms differ greatly in supporting generalized transduction, a feature of great utility for bacterial genetics but highly undesirable in phages being considered for therapeutic use [48]. With the advent of next-generation DNA sequencing, a rate-limiting step in phage genomics is now the determination of the genomic termini, which must still largely be determined experimentally. This is especially true of phages with non-permuted terminal repeats: during sequence assembly, reads from each identical repeat are usually collapsed into a single region in the middle of the assembled contig, resulting in what can appear to be a circularly permuted genome. Furthermore, determination of the true boundaries of direct repeats by traditional methods of restriction mapping and direct sequencing is laborious and time consuming [47]. Here, simple quantification of coverage depth in the assembled contigs revealed the boundaries of non-permuted terminal repeats in these genomes. Moreover, the terminal ligation method developed here confirmed these repeats and the position of the genomic termini in the phiCbK genome, and the close DNA sequence homology and similar coverage patterns (Figure 7) allowed high-confidence assignment of the termini in four of the other five phages. Deep-sequencing coverage has been used to determine copy number variation in other genomic contexts (recently reviewed in [49]), but the small size and low complexity of phage genomes allows for manual inspection of coverage patterns. The terminal ligation method used here is similar in principle to a previously described method based on cloning of restriction fragments following linker ligation [50]. The use of next-generation sequencing, however, removes the limitations imposed by restriction digestion of the DNA, and thus should be applicable to mapping all forms of genomic termini. In principle this method can be run in parallel

with the phage *de novo* sequencing run to obtain terminus information simultaneously with the genomic sequence.

Terminase and DNA packaging

The genes encoding the large subunit terminase (TerL) proteins of the phiCbK-like phages are located at the far right end of the genomes (Figure 2, Figure 3). In all phages except Colossus, the TerL gene overlaps with the terminal repeat, generating an unusual architecture in which the C-terminal fragment of *terL* is duplicated in each repeat. The novel genes immediately upstream of *terL* (gene 317 in phiCbK and 447 in Colossus) are predicted to encode small subunit terminase (TerS) proteins, based on their HHpred similarity to the TerS-based PDB family 3zqp_A (91.6% probability in phiCbK), and their position and size relative to *terL*. These proteins are conserved within the phiCbK-like phages but otherwise do not have any homologs in the NCBI database detectable by BLASTp. Phage TerS proteins are generally less conserved and more difficult to predict by simple sequence homology than TerL proteins.

The phiCbK-like TerL proteins contain an intein in the N-terminal portion of the protein, based on the detection of conserved intein domains by InterProScan, including a Hint domain (IPR003587) and intein splice site (IPR006141). Inteins are self-splicing elements that post-translationally catalyze their own excision from the completed protein. Based on alignments of the phiCbK TerL with its non-intein containing homologs, the putative intein appears to reside between K130 and S472, residues which are adjacent and absolutely conserved in 50 related TerL proteins ($E = 10^{-28}$ to 10^{-43} vs the nr database). The intein sequence itself is predicted to contain the 341 residues inclusive of P131 to N471, resulting in a 568 aa functional TerL protein following splicing; these boundaries are annotated in the phage genome records. The C-terminal boundary of the intein, S472, immediately follows the conserved His-Asn residues often associated with the C-terminal intein-extein boundary [51]. While the N-terminal intein boundary does not possess the Ser or Cys residue involved in typical N-terminal intein cleavage, functional inteins with a Pro at this position have been described [52]. Interestingly, the N-terminal region of the mature TerL protein is not related to other TerL proteins, including those that share similarity in the C-terminal region. The N-terminal domain of the phiCbK-like TerL may be the result of a recombinational event that swapped this portion of the protein for a different domain, perhaps using the intein sequence to provide homologous DNA for recombination.

The TerL proteins of phiCbK, Swift, Magneto and Karma share 100% identity in amino acid sequence. The TerL protein of Rogue is 87% similar, while Colossus is only 55% similar to that of phiCbK. Previous bioinformatic analysis of bacteriophage large subunit terminase proteins has shown that TerL alleles can be grouped into families that correlate to the DNA packaging style of the organism [53]. A mature version of these TerL proteins, with the predicted inteins excised, was used for BLAST analysis and alignment. Among the TerL proteins of phages with known packaging types, the phiCbK TerL protein is most similar to that of phages AaΦ23 (NP_852753, 25.6% similarity) and PY100 (CAJ28416, 24.4% similarity). Phage AaΦ23 has been shown to package DNA that is circularly permuted and terminally redundant by 3.5% of the phage genome [54], and phage PY100 packages its DNA by a headful mechanism initiated at a defined *pac* site analogous to the mechanism employed by phage P22 [55]. Given the low protein similarity of the phiCbK TerL to even its closest relatives, we propose that the TerL proteins of phages phiCbK, Magneto, Swift, Karma and Rogue constitute a novel TerL class that packages phage DNA via a mechanism of long direct terminal repeats in a manner similar to that of phages T5 or SPO1 [46,56]. The TerL homolog of phage Colossus is distinctly diverged from the phiCbK TerL and appears to use a slightly altered packaging mechanism, based on analysis of its genomic termini as described above.

DNA replication

The central portion of the phiCbK genome contains a region of genes encoded predominantly on the minus strand, beginning approximately at position 113,200 and ending at 80,700, encompassing genes 109 through 138. In phage Colossus, this region spans position 146,000 to 109,000, including Colossus genes 140 through 179. This module appears to contain phage DNA metabolism and replication functions, including predicted aerobic ribonucleoside reductase subunits (phiCbK gp111 and gp112), thymidylate synthase (phiCbK gp116), a RecD-like ExoV helicase (phiCbK gp118), a DNA Pol III-like ribonuclease (phiCbK gp121) and a T7-like DNA polymerase (phiCbK gp123). The ribonucleoside reductase alpha subunit (phiCbK gp112) contains an intein that spans C278 to N584 inclusive, featuring the commonly conserved Cys residue at its N terminus and a relatively uncommon C-terminal Gly-Gln [52], resulting in a mature 629 amino acid protein. This intein is also present in the Colossus homolog gp143. The only other organism containing this intein feature is invertebrate iridescent virus 6 (NP_149548). While the alpha subunit possesses significant similarity to other proteins in the database (e.g.,

to the *C. crescentus* CB15 homolog NP_422286, 24.8% identity), the ribonucleoside reductase beta subunit (gp112) exhibits virtually no similarity to previously identified proteins. The beta subunit also contains a C-terminal thioredoxin-like domain (IPR012336, IPR002109), a domain architecture found only in the ribonucleoside reductase of *Vibrio cholerae* phage ICP1 (YP_004251147).

An interesting feature of all six phiCbK-like phages is the presence of a DNA polymerase, gp123 in phiCbK and gp158 in Colossus, which closely resembles that of coliphage T7 and members of the T7-like phage superfamily. Among the most closely related proteins to phiCbK gp123 is the DNA polymerase gp5 of phage T7 itself (NP_041982, $E = 4 \times 10^{-109}$, 35.1% identity) which can be aligned to phiCbK gp123 over its entire length. This relationship suggests that phiCbK-like phages may replicate their DNA in a manner similar to that of phage T7. Coliphage T7 employs a unique mechanism of DNA replication, requiring only four proteins: a DNA polymerase (gp5), a combination helicase-primase (gp4), a single-stranded DNA binding protein (gp2.5), and thioredoxin which serves as DNA polymerase processivity factor and is supplied by the *E. coli* host [57]. Both the phiCbK-like and T7 DNA polymerases possess similar domain architectures, with an N-terminal ribonuclease-like domain (IPR012337) responsible for proofreading and C-terminal DNA polymerase A domain (IPR001098). Unlike other members of the polA family such as the *E. coli* DNA pol I, which are low processivity enzymes involved in DNA repair, T7 gp5 is responsible for replication of the phage chromosome with an associated high processivity of thousands of bp per event [58]. The high processivity of T7 gp5 is conferred in large part by its association with host thioredoxin [58,59]; the thioredoxin-binding loop of T7 gp5 (residues 258–333) [59] is present but extended by 11 aa in phiCbK gp123, thus the phiCbK-like homologs may establish similar interactions with another protein to enhance their processivity. Phage phiCbK possesses a DNA helicase (gp131), but this protein does not appear to contain a DNA primase domain like that of T7 and is not related to T7 gp4 at the primary structure level. A single-stranded DNA binding protein was not detected in the genome of phage phiCbK or its relatives. If these phiCbK-like phages replicate their DNA in a manner similar to that of the T7-like phages, they may use host proteins to complete the DNA replication complex. To our knowledge, this is the first instance of a T7-like DNA polymerase appearing in a phage outside of the T7 superfamily.

PhiCbK gene *I26* (Colossus *I61*) encodes a homolog of the coliphage T5 A1 protein (YP_006832, 34.2% similarity, $E \sim 10^{-89}$), located transcriptionally slightly upstream of the predicted DNA polymerase (Figure 2 and Figure 3).

This is the second time that an A1 homolog has been detected outside of the known T5-like phages, with the only other instance in the genome of phiJL001 (YP_223929, 31.8% similarity). In T5, the A1 protein is transcribed in the pre-early stage of infection as one of the eight genes of the first-step-transfer (FST) DNA, which corresponds to the left large terminal repeat. A1 mutants are defective in multiple processes, including failure to degrade host DNA and to shut off T5 pre-early gene expression [60], and to accomplish the transfer of the rest of the T5 genome into the cell [61]. Oddly, A1 fractionates with the membrane fraction after treatment with RNase and DNase [62]. It also co-immunoprecipitates with host RNA polymerase separated in glycerol gradients, and A1-bound RNA polymerase molecules have altered transcription patterns on T5 DNA [63]. The simplest notion is that A1 is a transcription factor that also interacts with a membrane protein. In support of this notion, the only T5 FST gene that encodes a predicted membrane protein is adjacent to *A1*. In this perspective, the various diverse phenotypes of *A1* mutants are derived from a failure to alter the promoter specificity during pre-early gene expression. Because of their location in the phiCbK-like phages within the DNA replication module, it seems unlikely that these A1 homologs would play a role in DNA degradation, and their function may be to alter phiCbK gene expression via interaction with the host RNA polymerase. This function is supported by an N-terminal similarity of these proteins to PDB family 1g2h_A (94.0% probability in phiCbK) as detected by HHpred; 1g2h is based on the DNA-binding domain of the TyrR transcription factor of *Haemophilus influenzae*.

Tyrosine recombinase proteins were identified in all six phiCbK-like phages, in phage phiCbK this protein is gp143. Site-specific tyrosine recombinases like gp143 are often associated with prophage integration. However, as there are no reports of phage phiCbK exhibiting temperate behavior, it is most likely that the gp143 recombinase instead plays some role in the resolution of replicative intermediates or recombinational multimers of the phage chromosome. Tyrosine recombinases of this type are well known to be involved in chromosomal segregation, as with *E. coli* XerC and XerD [64] or the resolution of recombinational dimers, as with the Cre protein of coliphage P1 [65]. At 208 aa, phiCbK gp143 is also considerably smaller than other well-studied tyrosine recombinases such as phage lambda Int (NP_040609, 356 aa) or P1 Cre (YP_006472, 343 aa). PhiCbK gp143 is comprised of little more than the recombinase catalytic core which begins with R52, the apparent equivalent of the catalytic R173 of Cre [66]. The equivalent of the N-terminal clamp domain of Cre appears to be absent in phiCbK gp143, suggesting it could require an accessory protein for this function.

Table 2 Virion-associated proteins of phage phiCbK

Putative function	Phage protein					
	phiCbK ^a	Magneto	Swift	Karma	Rogue	Colossus
Portal protein	gp42	gp40	gp40	gp43	gp46	gp38
Major capsid protein	gp68	gp66	gp66	gp69	gp70	gp81
Minor capsid protein	gp69	gp67	gp67	gp70	gp71	gp82
Major tail tube protein	gp92	gp89	gp88	gp93	gp91	gp115
Tape measure	gp95	gp92	gp91	gp96	gp94	gp118
Tail protein	gp97	gp94	gp93	gp98	gp96	gp120
Tail protein	gp99	gp96	gp95	gp100	gp98	gp122, gp123 ^b
Tail protein	gp101	gp98	gp97	gp102	gp100	gp125

^aPhiCbK proteins were detected by proteomic analysis of purified virions; the homologous proteins in the other phiCbK-like phages are listed.

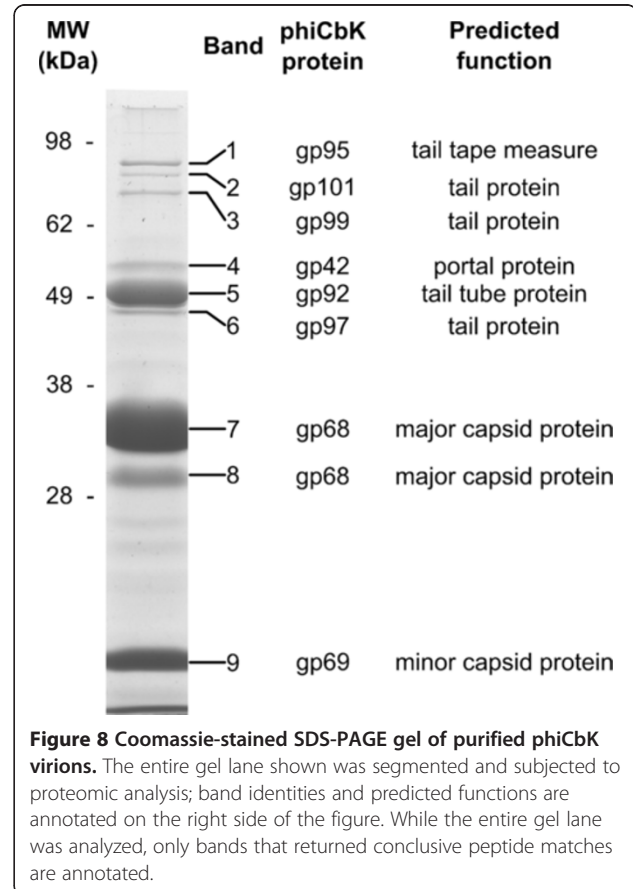
^bPhiCbK gp99 is duplicated in phage Colossus as gp122 and gp123.

Phage structural proteins

The structural proteome of phage phiCbK was analyzed by band excision from a SDS-PAGE gel followed by trypsin digestion and LC-MS/MS. As shown in Table 2, analysis of phiCbK directly identified eight proteins associated with the virion. The structural proteins detected in phiCbK are conserved in all six phiCbK-like phages (Table 2), indicating that despite differences in head length and genome composition, these phages are structurally very similar and likely share mechanisms of assembly and host adsorption.

Considerable study has been devoted to the structural proteins of phiCbK. Previous work by Leonard *et al.* [14] has indicated the phiCbK phage head is largely composed of two proteins: a major capsid protein of 36 kDa, and a minor capsid protein of 13.5 kDa, assembled in a 1:4 ratio. A third, minor protein of 33 kDa was also reported, with a stoichiometry suggesting a possible role as a head vertex protein. As shown in Figure 8, the major capsid protein (band 7), minor capsid protein (band 9) and putative vertex protein (band 8) exhibited molecular weights similar to those reported previously, and also correspond well to their calculated molecular weights. Interestingly, the bands corresponding to the major capsid protein (band 7) and the putative vertex protein (band 8) were both identified as gp68, suggesting that the band 8 protein is a processed form of the major capsid protein. In both bands 7 and 8, peptide hits were identified that spanned from A50 to R324 of the 339 aa gp68 protein, leaving the possibility open for a cleavage event to occur somewhere in the N-terminal 49 or C-terminal 15 amino acids of this protein. The Colossus gp81 major capsid protein is very similar (71% identity) to phiCbK gp68, except that it is truncated by 29 residues at its N terminus with respect to phiCbK gp68. In Figure 8, the gp92 tail tube protein (band 5), with a predicted MW of 64.8 kDa, exhibits an apparent molecular weight of ~52 kDa which is slightly lower

than the previously reported 58 kDa [14,67]. The tail tape measure protein, phiCbK gp95, was identified with the virions and appears as band 1 in Figure 8. The lengths of phiCbK gp95 (1,964 aa) and its Colossus homolog, gp118 (2,155 aa) predict tail lengths of 295 nm for phiCbK and 323 nm for Colossus assuming the alpha-helical structure found in other phage tail tape measure proteins [68]. This is good agreement with the



measured lengths of the tails of these phages (Table 1). In many dsDNA phages with long tails (i.e., *Myoviridae* and *Siphoviridae*), the gene encoding the tape measure protein is preceded by a gene that encodes two tape measure chaperone proteins, one of which is an extended form of the other as specified by a programmed -1 translational frameshift [69]. In all six phages, a pre-tape measure chaperone protein (phiCbK gp93) and its extended version (phiCbK gp94) could be identified, with the frameshift site located in the slippery sequence 5' - AAAAAAC - 3'. In Colossus, this slippery site is 5' - GGGAAAC - 3'.

Band 4 in Figure 8 was identified as phiCbK gp42, a protein that contains a DUF4055 domain (IPR025129) but otherwise has no homologs detectable by BLASTp with experimentally confirmed functional annotations in the NCBI database. HHpred searches detected a strong relationship to PDB family 2jes_A (99.9% probability in phiCbK), which is based on the phage SPP1 portal protein. This annotation as the portal protein is supported by the positions of several gp42 homologs (e.g., XP15 (YP_239276, 27.0% identity), BcepGomr (YP_001210225, 24.5% identity) and SE2 (YP_005098128, 22.9% identity)), which are all located immediately downstream of their predicted *terL*, suggesting their function as portal proteins. The 12-member portal ring formed by the portal protein is the nucleus of phage capsid assembly and is also the docking site of the DNA packaging complex [70]. Because of their required intimate interactions with the TerL protein, portal protein genes are often genetically coupled to the large terminase gene. A recent positional analysis of phage genomes found that the location of the large terminase gene immediately upstream of the portal-encoding gene to be common [71]. In the phiCbK-like phages, this arrangement appears to be disrupted as the portal gene is located 18–20 kb downstream from *terL*.

Phage phiCbK preferentially adsorbs to the swarmer cell type [13], as stalked cells lack the both the flagellar apparatus and pili [72] required for adsorption. Phage phiCbK is known to adsorb to its host cell via a two-step process [18]. First, a filament extending from the apical vertex of the phage head associates with the rotating bacterial flagellum; this attachment is not strictly required for phage adsorption, but flagellar defects reduce adsorption efficiency by ~3-fold. Flagellar rotation brings the phage into close proximity to the cell pole, where the phage tail tip adsorbs to the cell at the site of the pilus portal and completes the infection process. Because of their similar genomic and proteomic compositions, all six phiCbK-like phages presented here are likely to use the same host adsorption mechanism. The phiCbK head filament measures ~200 nm in length and is highly flexible, but nothing is otherwise known about its structural conformation or mechanism of association with the

flagellum (E. Wright, personal communication). In addition to the head filament, phage phiCbK has been observed to possess a single ~50 nm tail fiber that extends from the distal end of the phage tail [15] and is presumably involved in adsorption of the tail tip to the cell surface. The identity of the head filament and tail fiber proteins in phiCbK is of significant interest as this mechanism of phage attachment appears to be unique in phage biology.

Aside from the major tail tube subunit and the tape measure protein, three putative tail proteins were found to be associated with the phiCbK virion: gp101 (1,412 aa), gp99 (1,158 aa) and gp97 (541 aa) seen in Figure 8 as bands 2, 3 and 6, respectively. The genes encoding these proteins are in close proximity to other tail structural protein genes, the tail tube gene 92 and tail tape measure gene 95. Gp101 exhibits weak similarity to a putative tail fiber found in *Burkholderia* phage BcepNazgul (NP_918975, $E = 4.8 \times 10^{-11}$, 7.4% identity). While gp101 does not possess the conserved fibritin domain of the BcepNazgul protein, it does contain a pair of apparent ligand-binding domains (IPR000421, IPR008979), one of which is C-terminal, which may be involved in receptor binding. The band 3 protein gp99 contains an immunoglobulin-like (Ig) domain (IPR007110) and a phage tail-like domain (PF13550), and is weakly related ($E = 2 \times 10^{-22}$, 15.9% identity) to the *Rhodobacter capsulatus* gene transfer agent orf 15 (ABK27263), which is believed to be responsible for host specificity [73]. The N-proximal position of the gp99 Ig-like domain is not likely to play a role in receptor binding, as phage tail fibers typically have their receptor-binding domains in their C-termini. The phage tail-like domain is located in the central portion of gp99, although the function of this domain is unknown. The coliphage lambda tail fiber protein J also contains this Pfam tail fiber domain, but this region of the protein is outside of the C-terminal receptor-recognition domain [74] and thus may contribute to tail fiber structure rather than binding specificity. Finally, phiCbK gp97 contains three phage-associated conserved domains (IPR011928, IPR018964, IPR019228) of unknown function. A brief survey of other well-characterized phage tail fibers shows that such fibers yield approximately one nm of length per 12–20 amino acid residues (data not shown). If phiCbK gp101 formed tail fibers of this nature, the resulting fiber length should be on the order of 70–120 nm, while gp99 would form fibers of approximately 58–95 nm, and gp97 would form fibers of 27–48 nm. All of these lengths are in reasonable agreement with the size of the previously observed ~50 nm tail fiber. The identity of this fiber cannot be conclusively assigned at this time; the size and conserved domains of gp101 and gp99 make them more probable candidates for this role.

Phage phiCbK and its relatives also encode a very large conserved protein which, while not detected as associated with the virion in this analysis, remains the most probable candidate for the head filament protein. In phiCbK, gp76 is 2,799 amino acids in length and contains a concanavalin A-like lectin domain (IPR008985, IPR013320) in its central region. Gene 76 is located in the phage structural module, and is located closer to genes involved in head morphogenesis than to those associated with the phage tail (Figure 2). While this protein is conserved within all six of the phiCbK-like phages, it does not have any similarity to other known proteins detectable by BLASTp. The protein is also highly enriched in glycine (21% of amino acid residues, with multiple poly-Gly runs), alanine (12%), and proline (7%), with glycine repeat domains (PFAM12810) in the N-terminal and central domains of the protein. The high glycine content of gp76 would be expected to result in a highly flexible structure, and analysis by PsiPred [75] suggests the protein predominantly assumes random coil or beta strand conformation. Thus, gp76 fulfills the major criteria expected of the phiCbK head filament in terms of its size, flexibility, domain architecture and genomic location. Analysis of phiCbK 76 at the DNA level suggests that this gene has undergone expansion by tandem duplications in its N-terminal half. The high molecular weight bands above band 1 in Figure 8 did not return any peptide matches following MS analysis, and digestion of entire freeze-thawed virions followed by LC-MS also did not yield any peptide matches additional to those already described (data not shown). The head filament protein may not have been directly detected in the proteomic analysis due to the fact that it is relatively fragile and could have been lost during the purification procedure (E. Wright, personal communication).

Phage lysis proteins

In many phage types, the lysis genes are clustered in a "lysis cassette". For phages of Gram-negative hosts, the typical cassette consists of genes for the holin, the endolysin and the spanin subunits, responsible for permeabilization or destruction of the cytoplasmic membrane, the peptidoglycan, and the outer membrane, respectively. Of these three classes of genes, the endolysin and the spanins can be identified with the most confidence, because endolysins have a variety of easily identified catalytic motifs, and spanins have unique primary and secondary structure features. Using these considerations, the lysis cassettes were identified in all six phiCbK-like phages. The phiCbK cassette consists of genes 104 through 107 (Figure 9A), encoding the predicted endolysin, holin and spanin proteins, respectively; this order is conserved in phages Karma, Magneto, Swift and Rogue. In Colossus, the predicted holin gene has been transposed to

a position following the spanin genes, resulting in the gene order endolysin-spanin-holin, encoded by genes 134 to 137, respectively (Figure 9A). No such inversion of gene order has been detected before in phages of the same type, although among phages of the T7 phage-type the endolysin gene is in some cases transposed to an early gene transcriptional unit.

The endolysin proteins

All six phiCbK-like genomes have two proteins with catalytic motifs that are common in endolysins: the tail tape measure protein (gp95 in phiCbK), with a CHAP endopeptidase domain (IPR007921); and phiCbK gp104, which has a glycoside hydrolase domain (IPR002196). The presence of a CHAP domain in the tape measure protein is not unusual and is consistent with proposed mechanisms in which this protein is involved in penetrating the peptidoglycan after being ejected from the phage tail [76]. PhiCbK gp104 (gp134 in Colossus) is clearly the endolysin, since it has a strongly conserved glycoside hydrolase domain and has extensive homology to many phage proteins annotated as endolysins. However, only two coliphage homologs, P1 Lyz and R²¹ (both with an E value $\sim 10^{-7}$) have been subjected to biochemical and structural analysis. Both are SAR endolysins carrying export signals at their N-termini [77,78], a feature not shared by the phiCbK-like endolysins. Both R²¹ and P1 Lyz belong to the canonical T4 lysozyme family, enzymes that have a characteristic N-terminal Glu-8X-Cys/Asp-5X-Thr catalytic triad. A comparison of the N-terminal domains of these endolysins (Figure 9B) suggests that the phiCbK-like proteins represent novel variants of the canonical lysozyme catalytic triad. While all have retained the Cys/Asp and Thr residues with the standard spacing, in phiCbK and its close relatives Magneto, Karma, Swift and Rogue, four nearby acidic residues are available to play the role of the catalytic Glu common to all T4-related enzymes. The simplest notion that E15 is the catalytic Glu, with the spacing between the catalytic Glu and Cys residues increased to 16 residues by the insertion of an intervening Gly-Pro rich loop. In Colossus gp134, the normal spacing would be retained but the catalytic Asp/Cys would be replaced by another Glu residue. In P1 Lyz, the presence of the Cys residue in the catalytic triad is critical to the regulation of the enzyme because it is occupied in a disulfide bond until the enzyme is activated [79]. However, phiCbK gp104 lacks a SAR domain and is thus a soluble endolysin, dependent on the holin for release to the periplasm. It does have two other Cys residues, Cys124 and Cys 213, but neither is conserved in Colossus, making disulfide bond regulation unlikely. Thus phiCbK gp104 may have evolved from a P1 Lyz-like SAR endolysin, consistent with our previous proposal that SAR endolysins might

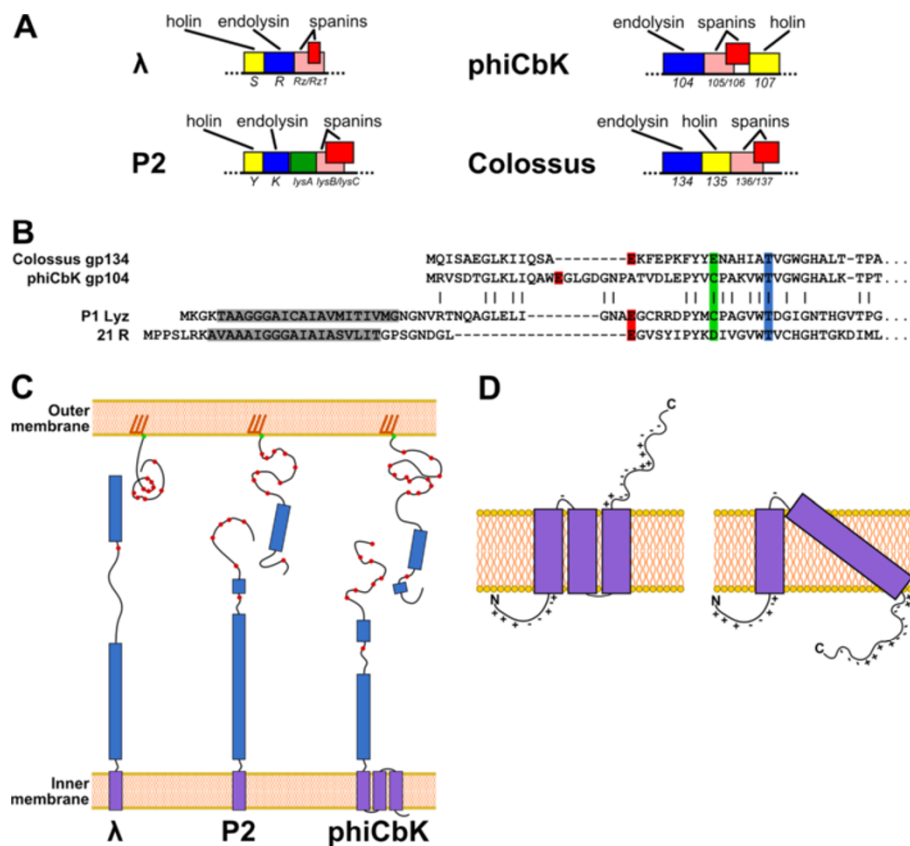


Figure 9 Schematic representations of the lysis genes and proteins of *C. crescentus* phages. Lysis components of phages phiCbK and Colossus are compared to those of other phages. **A:** Overall organization of the lysis cassettes of phages lambda, P2, phiCbK and Colossus; genes are represented by colored boxes, with gene functions labeled above each module and gene names below. Color indicates conserved function, not sequence similarity. **B:** Protein sequence alignment of the N-termini of the phages P1, 21, phiCbK and Colossus endolysins; the positions of equivalent E-D/C-T catalytic residues are highlighted in red, green and blue, transmembrane SAR domains are highlighted in grey. Identical residues in phiCbK gp104 and P1 Lyz are indicated by vertical bars, illustrating their relationship. **C:** The mature spanin proteins of phages lambda, P2 and phiCbK. Transmembrane domains are purple, predicted alpha-helical domains are blue, unstructured domains are black curves, and the positions of proline residues are represented by red dots. The spanin IM components are anchored to the inner membrane by a transmembrane helix, OM components are tethered to the outer membrane by a lipoylated (red bars) Cys residue (green dot). The P2 complex contains alpha-helical IM and OM components with proline-rich domains, and the phiCbK spanin complex more closely resembles that of P2. **D:** Possible membrane topologies of the phiCbK-like holin proteins. Left, three TMDs with N-in, C-out topology; right, two TMDs — including a 36-residue TMD2 — with N-in, C-in topology. Charge distributions of the protein and TMD length strongly favor the three TMD model.

constitute the most primitive form of regulated phage lysis, predating the appearance of holins in phage genomes [80].

The spanin subunits

By far the most common spanin genes are those that encode two-component spanins [81]. Each two component spanin locus encodes an integral inner membrane protein (the IM spanin), and an outer membrane lipoprotein (the OM spanin); these are the functional equivalents of the lambda Rz and Rz1 proteins, respectively. Genetic, biochemical and ultrastructural analyses have shown that the two proteins interact by virtue of their C-terminal domains to form the spanin complex, thus spanning the entire periplasm [82]. During lysis,

after the endolysin has degraded the peptidoglycan, the Rz-Rz1 complexes are thought to oligomerize laterally to form coiled-coil bundles, which somehow leads to OM disruption [83]. There are three known two-component spanin gene architectures: embedded, in which the gene encoding the OM spanin component is entirely contained within the IM component gene in the +1 reading frame (Figure 9A, lambda Rz/Rz1); overlapped, where the OM gene also starts within the gene encoding the IM component in the +1 frame but extends beyond it (Figure 9A, P2 lysB/lysC); and separated, where the coding regions do not overlap at all [81]. The phiCbK-like spanins (gp105/gp106 in phiCbK) are remarkably similar in their architecture to LysB and LysC found in the classical coliphage P2. As shown in Figure 9A, both gene

pairs belong to the overlapped spanin class. Although the Colossus spanin proteins are significantly diverged from those of the other phiCbK-like phages (29.7% and 36.0% similarity to phiCbK in the inner and outer membrane spanins, respectively), and these bear no sequence similarity to the spanins of phage P2, the periplasmic domains of all of these IM spanins both have similar predicted alpha helical character with proline-rich C-terminal domains, unlike lambda Rz (Figure 9C). One striking feature of the phiCbK-like IM spanins is the presence of a large N-terminal hydrophobic domain preceding the anchoring TMD, which strongly suggests the presence at least one and possibly two additional TMDs. The N-terminal TMD of the IM subunit of lambda Rz can be substituted by heterologous TMDs and is thus thought to be purely a membrane tether [82]. The extra membrane component in the phiCbK-like IM spanins would be unnecessary for simple membrane tethering and thus may reflect an additional function in lysis, perhaps in coordinating the function of the spanin complex with the function of the holin, which is an IM protein in all phages. Similarly, the periplasmic domains of the phiCbK-like and P2 OM spanins are remarkably parallel in primary structural features, with an N-terminal domain that is very enriched in proline residues and a C-terminal segment predicted to be alpha-helical but containing a single distal proline (Figure 9C). Little genetic analysis has been done with any spanin so it is unclear which of these features are important for function. However, missense changes in the first two proline residues in LysC have been shown to cause a premature lysis phenotype in cells with compromised peptidoglycan [84]. To date, these *lysC* alleles are unique in being the only alleles outside of holin and antiholin genes that have early lysis phenotypes, perhaps another hint that P2-like spanins may interact functionally with their cognate holins.

The holin proteins

Unless there is sequence similarity with one of the very few holins that have been experimentally verified, it is difficult to identify holins with the same degree of confidence available for endolysins and spanins. However, most of the experimentally tested holins have multiple TMDs and are encoded by genes clustered within a lysis cassette; phiCbK gene *107* (gene *135* in Colossus) fits these criteria (Figure 9A) and is thus predicted to be the holin gene. Experimentally confirmed holins have been identified in three different membrane topologies: class I (3 TMDs with N-out, C-in); class II (2 TMDs with N and C in); and class III (one TMD with N-in, C-out) [85]. The first TMD of phiCbK gp107 is unambiguous, between Leu37 and Phe60, and the charge distribution flanking this TMD would put Asp66 at the periplasmic

interface, so TMD1 is N-in, C-out. Beyond this, however, phiCbK-like holin topology becomes ambiguous. Commonly used algorithms like TMHMM and TMPred predict these holins to have only two TMDs, the second of which would be over 30 residues in length (Figure 9D). An alternative perspective is that this region encodes three TMDs, taking advantage of the fact that Lys residues, like Lys100, can “snorkel” and thus be mapped up to three residues inside the bilayer. This would give the phiCbK-like holins a novel topology, with three TMDs and an N-in, C-out orientation (Figure 9D). In support of this notion, the C-terminal domain beyond the ambiguous hydrophobic region is strongly acidic, unlike class I and class II holins, which both feature positively charged C-terminal cytoplasmic domains and consistent with the positive-inside, negative-outside theorem that dominates prokaryotic TMD topology [86]. It should be noted, however, that absent experimental confirmation, the identity of *107* as the holin is speculative. Especially in view of the extra TMDs present in the IM spanin subunit of phiCbK, it is worth noting that genes implicated in antiholin function have been identified in other phages, with gene products having a different variety of topologies.

PhiCbK-like GcrA homolog

C. crescentus exhibits a peculiar dimorphic lifestyle, in which the stalked cell form divides asymmetrically to produce a flagellated “swarmer” cell. The swarmer cells are not capable of DNA replication nor division until they undergo the transition to the stalked morphotype (recently reviewed in [4]). Transformation of cell shape, chromosomal replication and cell division are tightly regulated in *C. crescentus* by a cascade of master regulatory proteins including DnaA [87], CtrA [3,88] and GcrA [89], and by the methylation state of the chromosome as mediated by CcrM [90]. Under normal conditions, GcrA is present in stalked cells undergoing DNA replication and division, and is not detectable in swarmer cells where CtrA is dominant [89]. CtrA binds to the GcrA transcriptional start site, negatively regulating its expression during the swarmer phase [89]. When expressed during early stages of DNA replication and cell division, GcrA up-regulates transcription of 89 *C. crescentus* genes, including several components of the cell's DNA replication machinery such as a DNA gyrase, topoisomerase IV and a DNA pol III epsilon subunit [89].

The tropism of phage phiCbK for the swarmer cell type means that phiCbK and its relatives predominantly infect cells in which DNA replication is halted and GcrA is downregulated. It is significant, therefore, that all of the phiCbK-like phages with the exception of Colossus encode a protein with homology to the *C. crescentus* GcrA cell cycle regulator; in phage phiCbK this protein

is gp222. The GcrA homologs encoded by phages phiCbK, Magneto, Swift and Karma are identical and the Rogue-encoded homolog is slightly diverged, a reflection of the overall relationship of these phages. Forty-one amino acid residues are conserved across all phage-encoded proteins and the *C. crescentus* CB15 GcrA homolog, and an additional 18 residues are similar. Furthermore, all of the phage-encoded homologs are detectable as members of the GcrA protein family (pfam07750, IPR011681). If the phage-encoded GcrA homolog possesses the same functionality as its *C. crescentus* counterpart, the expression of phiCbK gp222 in the infected swarmer cell may function to up-regulate expression of the *C. crescentus* DNA replication apparatus, components of which may be required for the successful replication of the phiCbK chromosome. GcrA is rapidly degraded in swarmer cells [87], therefore it also might be expected that the phage-encoded GcrA homologs are able to resist this degradation.

Possible translational modulators in phage Colossus

While phage Colossus lacks the GcrA-like proteins found in the other phiCbK-like phages, it possesses a number of additional genes, not found in the other phages, that may serve to modulate the host cell to the phage's advantage. For example, Colossus gp247 appears to be a truncated form of ribosomal elongation factor Ts (EF-Ts), with a predicted EF-Ts domain (IPR014039). EF-Ts catalyzes the cycling of GDP to GTP in elongation factor Tu (EF-Tu) by binding to the EF-Tu•GDP complex and accelerating both the release of GDP and the association of GTP. Expression of a phage-encoded EF-Ts-like protein during infection could theoretically benefit the phage by increasing the available charged EF-Tu pool and accelerating translation. While Colossus gp247 is only 98 amino acids long, approximately one-third the length of known EF-Ts proteins, it contains the N-proximal EF-Ts subdomain N with its conserved F81 residue (F32 in gp247) that plays a major role in GDP nucleotide release from EF-Tu [91]. However, gp247 seems to lack the C-proximal dimerization domain and other residues that are required for the interaction of EF-Ts with EF-Tu in other proteins, thus it is not clear if gp247 possesses full EF-Ts-like functionality.

Colossus also contains a possible ribosomal RNA methyltransferase protein, gp358. This protein possesses a RsmD-like rRNA methyltransferase domain (IPR004398) and HHpred searches detect similarity to a number of methyltransferase families with roughly equal probability, including type I restriction-like DNA methyltransferases (e.g., 2okc_A, 2ar0_A, both 99.8% probability) and rRNA methyltransferases (e.g., 1uwv_A, 99.7% probability, 3g89_A, 99.6% probability). Methylated rRNA bases tend to be clustered around active sites of the mature ribosome [92], although their function is not entirely clear as

absence of methylation at these sites typically yields a mild phenotype [93,94]. Modification of the bacterial ribosome by a phage-encoded rRNA methyltransferase may provide a fitness advantage to the phage, although the function of gp358 remains to be determined.

Conclusions

The phiCbK-like phages described here constitute a novel phage type and are distinguished by large genomes of 200 – 300 kb and prolate siphophage morphology. They all possess a T7-like DNA polymerase, suggesting that they employ a T7-like DNA replication strategy; this is the first case to our knowledge of a T7-like DNA polymerase found outside of phages of the T7 superfamily. The structural proteome of phiCbK identified the major phage structural components and several minor tail proteins of currently unknown function, although at least one of these is likely to be the tail fiber protein. The head filament protein was not conclusively established in the phiCbK virion but a strong candidate protein for this role, gp76 in phiCbK, was identified by bioinformatic analysis. These phages were found to contain several proteins that may allow the phage to more efficiently utilize cellular resources, including homologs of the GcrA cell cycle regulator and proteins that may interact with the cellular translational apparatus. The phages were sequenced by 454 pyrosequencing to coverages of 38 to 291-fold depth. These coverages were sufficient to identify coverage breakpoints that represented the boundaries of non-permuted terminal repeats in these genomes. The terminal ligation method developed here confirmed these boundaries in phage phiCbK. These methods will allow for greater throughput in the identification of phage genomic termini in the future.

Additional file

Additional file 1: Tables S1–S6. Complete genome annotation tables with evidence.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JJG, JDB and RY conceived the study. JDB and JJG sequenced and assembled the phage genomes, and JJG, JDB and RY finalized genome annotations. JJG conducted comparative genomic analyses, generated maps and figures, and conceived of and analyzed terminal labeling experiments. JJG and RY wrote the initial draft of the manuscript. WKR and DHR conducted proteomic analysis and analyzed the data. DAE prepared phage for the proteomic analysis and conducted SDS-PAGE experiments. LL conducted the terminal labeling experiments and helped close and finalize the phage genomes. MM and SM conducted experiments to close and finalize the phage genomes, and assisted with annotation. DH, AK, JK, RM and AMT conducted initial genomic annotations and analyses. All authors have read and approved the final manuscript.

Acknowledgements

The authors would like to thank students in the Texas A&M University Bacteriophage Genomics class of 2010 for the initial isolation of phages: Rachel Gubbels, Alex Klattenhoff, Kristen Thomas and Joshua Tsau for phages Colossus, Rogue and Swift; Justin Kaspar, Bryan Johnson, David Migl and Misty Reynolds for Karma and Storm; Jared Fradette, Andrew Hawrylak, Ashlyn Thorn and TJ Tidwell for Magneto. The authors would like to thank Roger Hendrix of the University of Pittsburgh for the direct phage concentration protocol, and Elizabeth Wright of Emory University for her insights into the structure of phiCbK and its head filament. This work was supported by NSF Award No. 0949351 and by support to the Center for Phage Technology from the Texas A&M University IUMRI program. Proteomic studies were supported by the NSF Major Research Instrumentation Program Grant DBI-0821700.

Note added in proof

After submission of this manuscript, we learned that Panis *et al.* also sequenced the genome of phiCbK [95], deposited under accession number JX163858.

Author details

¹Center for Phage Technology, 2128 TAMU, Texas A&M University, College Station, Texas, TX 77843, USA. ²Dept. of Biochemistry & Biophysics, 2128 TAMU, Texas A&M University, College Station, Texas, TX 77843, USA.

³Department of Chemistry, Texas A&M University, College Station, Texas, TX 77843, USA.

Received: 14 June 2012 Accepted: 1 October 2012

Published: 10 October 2012

References

1. Stove JL, Stanier RY: Cellular Differentiation in Stalked Bacteria. *Nature* 1962, **196**(4860):1189–1192.
2. Shapiro L, Agabian-Keshishian N, Bendis I: Bacterial differentiation. *Science* 1971, **173**(4000):884–892.
3. McAdams HH, Shapiro L: A bacterial cell-cycle regulatory network operating in time and space. *Science* 2003, **301**(5641):1874–1877.
4. Curtis PD, Brun YV: Getting in the loop: regulation of development in *Caulobacter crescentus*. *Microbiol Mol Biol Rev* 2010, **74**(1):13–41.
5. Goley ED, Iniesta AA, Shapiro L: Cell cycle regulation in *Caulobacter*: location, location, location. *J Cell Sci* 2007, **120**(Pt 20):3501–3507.
6. Schmidt JM, Stanier RY: Isolation and Characterization of Bacteriophages Active against Stalked Bacteria. *J Gen Microbiol* 1965, **39**:95–107.
7. Ely B, Johnson RC: Generalized Transduction in *Caulobacter crescentus*. *Genetics* 1977, **87**(3):391–399.
8. Bender RA: Improved generalized transducing bacteriophage for *Caulobacter crescentus*. *J Bacteriol* 1981, **148**(2):734–735.
9. Shapiro L, Agabian-Keshishian N: Specific Assay for Differentiation in the Stalked Bacterium *Caulobacter crescentus*. *Proc Natl Acad Sci USA* 1970, **67**(1):200–203.
10. Agabian-Keshishian N, Shapiro L: Bacterial differentiation and phage infection. *Virology* 1971, **44**(1):46–53.
11. Lagenaur C, Agabian N: *Caulobacter crescentus* pili: structure and stage-specific expression. *J Bacteriol* 1977, **131**(1):340–346.
12. Agabian-Keshishian N, Shapiro L: Stalked bacteria: properties of deoxyribonucleic acid bacteriophage phiCbK. *J Virol* 1970, **5**(6):795–800.
13. Lagenaur C, Farmer S, Agabian N: Adsorption properties of stage-specific *Caulobacter* phage phiCbK. *Virology* 1977, **77**(1):401–407.
14. Leonard KR, Kleinschmidt AK, Agabian-Keshishian N, Shapiro L, Maizel JV Jr: Structural studies on the capsid of *Caulobacter crescentus* bacteriophage phiCbK. *J Mol Biol* 1972, **71**(2):201–216.
15. Leonard KR, Kleinschmidt AK, Lake JA: *Caulobacter crescentus* bacteriophage phiCbK: structure and in vitro self-assembly of the tail. *J Mol Biol* 1973, **81**(3):349–365.
16. Lake JA, Leonard KR: Structure and protein distribution for the capsid of *Caulobacter crescentus* bacteriophage phiCbK. *J Mol Biol* 1974, **86**(3):499–518.
17. Lake JA, Leonard KR: Bacteriophage structure: determination of head-tail symmetry mismatch for *Caulobacter crescentus* phage phiCbK. *Science* 1974, **183**(4126):744–747.
18. Guerrero-Ferreira RC, Viollier PH, Ely B, Poindexter JS, Georgieva M, Jensen GJ, Wright ER: Alternative mechanism for bacteriophage adsorption to the motile bacterium *Caulobacter crescentus*. *Proc Natl Acad Sci USA* 2011, **108**(24):9963–9968.
19. Adams MH: *Bacteriophages*. New York: Interscience Publishers; 1959.
20. Summer EJ: Preparation of a phage DNA fragment library for whole genome shotgun sequencing. *Methods Mol Biol* 2009, **502**:27–46.
21. Lukashin AV, Borodovsky M: GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 1998, **26**(4):1107–1115.
22. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL: Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999, **27**(23):4636–4641.
23. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: Artemis: sequence visualization and annotation. *Bioinformatics* 2000, **16**(10):944–945.
24. Schattner P, Brooks AN, Lowe TM: The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 2005, **33**:W686–W689.
25. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, et al: Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Res* 2011, **39**:D141–D145.
26. Kingsford CL, Ayanbule K, Salzberg SL: Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 2007, **8**(2):R22.
27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: BLAST+: architecture and applications. *BMC Bioinforma* 2009, **10**:421.
28. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al: InterPro: the integrative protein signature database. *Nucleic Acids Res* 2009, **37**:D211–D215.
29. Soding J, Biegert A, Lupas AN: The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 2005, **33**:W244–W248.
30. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A: Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 2003, **12**(8):1652–1662.
31. Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000, **16**(6):276–277.
32. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al: The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008, **9**:75.
33. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: Circos: an information aesthetic for comparative genomics. *Genome Res* 2009, **19**(9):1639–1645.
34. Gill JJ, Summer EJ, Russell WK, Cologna SM, Carlile TM, Fuller AC, Kitsopoulos K, Mebane LM, Parkinson BN, Sullivan D, et al: Genomes and characterization of phages Bcep22 and BcepL02, founders of a novel phage type in *Burkholderia cenocepacia*. *J Bacteriol* 2011, **193**(19):5300–5313.
35. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, **227**(5259):680–685.
36. Valentine RC, Shapiro BM, Stadtman ER: Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. *Biochemistry* 1968, **7**(6):2143–2152.
37. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**(17):3389–3402.
38. Lohr JE, Chen F, Hill RT: Genomic analysis of bacteriophage PhiL001: insights into its interaction with a sponge-associated alpha-proteobacterium. *Appl Environ Microbiol* 2005, **71**(3):1598–1609.
39. Lang AS, Beatty JT: Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci USA* 2000, **97**(2):859–864.
40. Lawrence JG, Hatfull GF, Hendrix RW: Imbroglis of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol* 2002, **184**(17):4891–4905.
41. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF: Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci USA* 1999, **96**(5):2192–2197.
42. Casjens SR: Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae. *Res Microbiol* 2008, **159**(5):340–348.

43. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, et al: **Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size.** *J Mol Biol* 2010, **397**(1):119–143.
44. Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, Ackermann HW, Kropinski AM: **Classification of Myoviridae bacteriophages using protein sequence similarity.** *BMC Microbiol* 2009, **9**:224.
45. Lavigne R, Seto D, Mahadevan P, Ackermann HW, Kropinski AM: **Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools.** *Res Microbiol* 2008, **159**(5):406–414.
46. Sayers JR: **Bacteriophage T5.** In *The Bacteriophages, 2nd Edition*. Edited by Calendar R, Abedon ST. New York, NY: Oxford University Press; 2006:268–276.
47. Casjens SR, Gilcrease EB: **Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions.** *Methods Mol Biol* 2009, **502**:91–111.
48. Gill JJ, Young R: **Therapeutic Applications of Phage Biology: History, Practice and Recommendations.** In *Emerging Trends in Antibacterial Discovery: Answering the Call to Arms*. Edited by Miller AA, Miller PF. Norfolk, UK: Caister Academic Press; 2011:367–410.
49. Koboldt DC, Larson DE, Chen K, Ding L, Wilson RK: **Massively parallel sequencing approaches for characterization of structural variation.** *Methods Mol Biol* 2012, **838**:369–384.
50. Summer EJ, Gonzalez CF, Borner M, Carlile T, Embry A, Kucherka AM, Lee J, Mebane L, Morrison WC, Mark L, et al: **Divergence and mosaicism among virulent soil phages of the Burkholderia cepacia complex.** *J Bacteriol* 2006, **188**(1):255–268.
51. Elleuche S, Poggeler S: **Inteins, valuable genetic elements in molecular biology and biotechnology.** *Appl Microbiol Biotechnol* 2010, **87**(2):479–489.
52. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Pietrokovski S, Perler FB: **Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile.** *J Biol Chem* 2010, **285**(4):2515–2526.
53. Casjens SR, Gilcrease EB, Winn-Stapley DA, Schickmaier P, Schmieger H, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW: **The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy.** *J Bacteriol* 2005, **187**(3):1091–1104.
54. Willi K, Meyer J: **DNA analysis of temperate bacteriophage Aa(phi)23 isolated from actinobacillus actinomycetemcomitans.** *Mol Gen Genet* 1998, **258**(4):323–325.
55. Schwudke D, Ergin A, Michael K, Volkmar S, Appel B, Knabner D, Konietzny A, Strauch E: **Broad-host-range Yersinia phage PY100: genome sequence, proteome analysis of virions, and DNA packaging strategy.** *J Bacteriol* 2008, **190**(1):332–342.
56. Stewart CR, Casjens SR, Cresawn SG, Houtz JM, Smith AL, Ford ME, Peebles CL, Hatfull GF, Hendrix RW, Huang WM, et al: **The genome of Bacillus subtilis bacteriophage SP01.** *J Mol Biol* 2009, **388**(1):48–70.
57. Lee SJ, Richardson CC: **Choreography of bacteriophage T7 DNA replication.** *Curr Opin Chem Biol* 2011, **15**(5):580–586.
58. Tabor S, Huber HE, Richardson CC: **Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7.** *J Biol Chem* 1987, **262**(33):16212–16223.
59. Bedford E, Tabor S, Richardson CC: **The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on Escherichia coli DNA polymerase I.** *Proc Natl Acad Sci USA* 1997, **94**(2):479–484.
60. McCorquodale DJ, Lanni YT: **Patterns of protein synthesis in Escherichia coli infected by amber mutants in the first-step-transfer DNA of T5.** *J Mol Biol* 1970, **48**(1):133–143.
61. McCorquodale DJ, Warner HR: **Bacteriophage T5 and Related Phages.** In *The Bacteriophages*. Edited by Calendar R. New York, NY: Plenum Press; 1988:439–475.
62. Duckworth DH, Dunn GB: **Membrane protein biosynthesis in T5 bacteriophage-infected Escherichia coli.** *Arch Biochem Biophys* 1976, **172**(2):319–328.
63. McCorquodale DJ, Chen CW, Joseph MK, Woychik R: **Modification of RNA polymerase from Escherichia coli by pre-early gene products of bacteriophage T5.** *J Virol* 1981, **40**(3):958–962.
64. Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, Sherratt DJ: **Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12.** *Cell* 1993, **75**(2):351–361.
65. Austin S, Ziese M, Sternberg N: **A novel role for site-specific recombination in maintenance of bacterial replicons.** *Cell* 1981, **25**(3):729–736.
66. Guo F, Gopaul DN, van Duyn GD: **Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse.** *Nature* 1997, **389**(6646):40–46.
67. Papadopoulos S, Smith PR: **The structure of the tail of the bacteriophage phi CbK.** *J Ultrastruct Res* 1982, **80**(1):62–70.
68. Hendrix RW: **Tail length determination in double-stranded DNA bacteriophages.** *Curr Top Microbiol Immunol* 1988, **136**:21–29.
69. Xu J, Hendrix RW, Duda RL: **Conserved translational frameshift in dsDNA bacteriophage tail assembly genes.** *Mol Cell* 2004, **16**(1):11–21.
70. Rao VB, Feiss M: **The bacteriophage DNA packaging motor.** *Annu Rev Genet* 2008, **42**:647–681.
71. Li J, Halgamuge SK, Tang SL: **Genome classification by gene distribution: an overlapping subspace clustering approach.** *BMC Evol Biol* 2008, **8**:116.
72. Sommer JM, Newton A: **Sequential regulation of developmental events during polar morphogenesis in Caulobacter crescentus: assembly of pili on swarmer cells requires cell separation.** *J Bacteriol* 1988, **170**(1):409–415.
73. Chen F, Spano A, Goodman BE, Blasier KR, Sabat A, Jeffery E, Norris A, Shabanowitz J, Hunt DF, Lebedev N: **Proteomic analysis and identification of the structural and regulatory proteins of the Rhodobacter capsulatus gene transfer agent.** *J Proteome Res* 2009, **8**(2):967–973.
74. Wang J, Hofnung M, Charbit A: **The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12.** *J Bacteriol* 2000, **182**(2):508–512.
75. Buchan DW, Ward SM, Loble AE, Nugent TC, Bryson K, Jones DT: **Protein annotation and modelling servers at University College London.** *Nucleic Acids Res* 2010, **38**:W563–W568.
76. Piuri M, Hatfull GF: **A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells.** *Mol Microbiol* 2006, **62**(6):1569–1585.
77. Xu M, Struck DK, Deaton J, Wang IN, Young R: **A signal-arrest-release sequence mediates export and control of the phage P1 endolysin.** *Proc Natl Acad Sci USA* 2004, **101**(17):6415–6420.
78. Sun Q, Kutay GF, Arockiasamy A, Xu M, Young R, Sacchettini JC: **Regulation of a muralytic enzyme by dynamic membrane topology.** *Nat Struct Mol Biol* 2009, **16**(11):1192–1194.
79. Xu M, Arulandu A, Struck DK, Swanson S, Sacchettini JC, Young R: **Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme.** *Science* 2005, **307**(5706):113–117.
80. Park T, Struck DK, Dankenbring CA, Young R: **The pinholin of lambdaoid phage 21: control of lysis by membrane depolarization.** *J Bacteriol* 2007, **189**(24):9135–9139.
81. Summer EJ, Berry J, Tran TA, Niu L, Struck DK, Young R: **Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts.** *J Mol Biol* 2007, **373**(5):1098–1112.
82. Berry J, Summer EJ, Struck DK, Young R: **The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes.** *Mol Microbiol* 2008, **70**(2):341–351.
83. Berry J, Savva C, Holzenburg A, Young R: **The lambda spanin components Rz and Rz1 undergo tertiary and quaternary rearrangements upon complex formation.** *Protein Sci* 2010, **19**(10):1967–1977.
84. Markov D, Christie GE, Sauer B, Calendar R, Park T, Young R, Severinov K: **P2 growth restriction on an rpoC mutant is suppressed by alleles of the Rz1 homolog lysC.** *J Bacteriol* 2004, **186**(14):4628–4637.
85. Wang IN, Young R: **Phage Lysis.** In *The Bacteriophages, 2nd ed.* Edited by Calendar R, Abedon ST. New York, NY: Oxford University Press; 2006:104–125.
86. Von Heijne G: **Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule.** *J Mol Biol* 1992, **225**(2):487–494.
87. Collier J, Murray SR, Shapiro L: **DnaA couples DNA replication and the expression of two cell cycle master regulators.** *EMBO J* 2006, **25**(2):346–356.
88. Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L: **Global analysis**

- of the genetic network controlling a bacterial cell cycle. *Science* 2000, **290**(5499):2144–2148.
89. Holtzendorff J, Hung D, Brende P, Reisenauer A, Viollier PH, McAdams HH, Shapiro L: **Oscillating global regulators control the genetic circuit driving a bacterial cell cycle.** *Science* 2004, **304**(5673):983–987.
90. Collier J, McAdams HH, Shapiro L: **A DNA methylation ratchet governs progression through a bacterial cell cycle.** *Proc Natl Acad Sci USA* 2007, **104**(43):17111–17116.
91. Kawashima T, Berthet-Colominas C, Wulff M, Cusack S, Leberman R: **The structure of the *Escherichia coli* EF-Tu.EF-Ts complex at 2.5 Å resolution.** *Nature* 1996, **379**(6565):511–518.
92. Mueller F, Brimacombe R: **A new model for the three-dimensional folding of *Escherichia coli* 16 S ribosomal RNA. I. Fitting the RNA to a 3D electron microscopic map at 20 Å.** *J Mol Biol* 1997, **271**(4):524–544.
93. Jemiolo DK, Taurence JS, Giese S: **Mutations in 16S rRNA in *Escherichia coli* at methyl-modified sites: G966, C967, and G1207.** *Nucleic Acids Res* 1991, **19**(15):4259–4265.
94. Lesnyak DV, Osipiuk J, Skarina T, Sergiev PV, Bogdanov AA, Edwards A, Savchenko A, Joachimiak A, Dontsova OA: **Methyltransferase that modifies guanine 966 of the 16 S rRNA: functional identification and tertiary structure.** *J Biol Chem* 2007, **282**(8):5880–5887.
95. Panis G, Lambert C, Viollier PH: **Complete genome sequence of *Caulobacter crescentus* bacteriophage ϕ CbK.** *J Virol* 2012, **86**(18):10234–10235.

doi:10.1186/1471-2164-13-542

Cite this article as: Gill et al.: The *Caulobacter crescentus* phage ϕ CbK: genomics of a canonical phage. *BMC Genomics* 2012 13:542.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

