Published in final edited form as:

Curr Microbiol. 2015 June ; 70(6): 854–858. doi:10.1007/s00284-015-0799-5.

The *Caulobacter crescentus* transducing phage Cr30 is a unique member of the T4-like family of myophages

Bert Ely*, Whitney Gibbs¹, Simon Diez, and Kurt Ash

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208

Abstract

Bacteriophage Cr30 has proven useful for the transduction of *Caulobacter crescentus*. Nucleotide sequencing of Cr30 DNA revealed that the Cr30 genome consists of 155,997 bp of DNA that codes for 287 proteins and five tRNAs. In contrast to the 67% GC content of the host genome, the GC content of the Cr30 genome is only 38%. This lower GC content causes both the codon usage pattern and the amino acid composition of the Cr30 proteins to be quite different from those of the host bacteria. As a consequence, the Cr30 mRNAs probably are translated at a rate that is slower than the normal rate for host mRNAs. A phylogenetic comparison of the genome indicates that Cr30 is a member of the T4-like family that is most closely related to a new group of T-like phages exemplified by ϕ M12.

Keywords

Caulobacter; bacteriophage; genome sequence; T4-like

The availability of low cost next generation sequencing techniques has revived interest in the study of bacteriophages, and bacteriophage genomes are accumulating rapidly in the National Center for Biotechnology Information (NCBI) nucleotide sequence database. A large collection of bacteriophages that infect *Caulobacter crescentus* was isolated 40 years ago by Johnson, Wood, and Ely (17) and most of these phages resembled CbK, a large bacteriophage with an elongated head and a flexible tail (1). However, two of these phages, Cr30 and Cr35, were shown to be generalized transducing phage (13). The availability of a transducing phage focused attention on the development of a system of genetics for *C. crescentus*, and the large size of the Cr30 phage genome made it possible to construct a rudimentary genetic map of the *C. crescentus* genome using only transduction as a means of genetic exchange (5, 6). Therefore, Cr30 quickly became a key element for the genetic analysis of *C. crescentus* (12). The remainder of the phages isolated by Johnson, Wood, and Ely (17) were frozen for later use. However, a freezer meltdown occurred and Cr35 and most of the other phages in this collection were lost.

*Corresponding author: ely@sc.edu; fax 803-777-4002.

¹Current address: South Carolina College of Pharmacy, Medical University of South Carolina, Charleston, SC

Competing Interests

The authors declare that there are no competing interests.

In this study, we determined the nucleotide sequence of the CR30 genome and demonstrated that Cr30 is a T4-like phage. When compared to the T4 family of phage genomes, both the gene order and the amino acid sequences of the proteins they encode are highly conserved. However, the Cr30 genome most closely resembles the recently described genome of ϕ M12 (7). ϕ M12 is a unique member of the T4 superfamily that infects *Sinorhizobium meliloti* and is distinct from both major groups of the T4 phage superfamily, the cyanophages and the T4-like phages that infect enteric bacteria (pseudo T4 phages). For example, the ϕ M12 genome contains some genes that are more closely-related to those of the T4-like cyanophages and others that are more closely-related to those of the pseudo T4 phages that infect enteric bacteria. Consequently they proposed that ϕ M12 is the first member of a third group of T4-like bacteriophages to have the nucleotide sequence of its genome determined. This study of the Cr30 genome provides a second member of this third group of T4-like phages.

Results and Discussion

Genome characterization

The nucleotide sequence of the Cr30 genome was assembled into a single contig, and the ends of the contig were trimmed to eliminate redundancy. Although the size of the DNA isolated from Cr30 phage particles is approximately 180,000 bp, the Cr30 genome (Accession: NC_025422) consists of only 155,997 bp of DNA, suggesting that the Cr30 genome is packaged by a headful mechanism that incorporates sections of concatenated Cr30 genomes that are approximately 15% larger than the unit genome. This result is consistent with the observation that Cr30 transducing particles contain nearly 200 kb of DNA (6, 7, 11). To verify that the CR30 DNA sequence reads were assembled correctly, we analyzed the Cr30 genome sequence and identified two restriction enzymes, *Swa*I and *Spe*I, that cut the genome into large pieces. Therefore, agarose plugs containing intact Cr30 DNA were digested separately with each of these enzymes, and the resulting restriction fragments were separated by pulsed field gel electrophoresis. The *Spe*I digest produced the expected 81 Kb and 36 kb fragments, and the *Swa*I digest produced the expected 26 Kb fragment. The other fragments produced by these digests were either too small to be seen by pulsed field gel electrophoresis or they were of variable length due to the headful packaging.

The current annotation of the CR30 genome contains 287 protein coding genes and no transposases or mobile genetic elements. However, it does contain an integrase and one seg-like homing endonuclease. The average GC content of the Cr30 genome is 38% compared to the 67% GC content of the *C. crescentus* host genome. This large difference in genomic GC content and the corresponding difference in codon usage patterns might slow the process of translating phage mRNAs due to the fact that the phage mRNAs contain an abundance of codons that are rarely used by the host. In fact, a slower rate of translation would be consistent with the 90 minute eclipse period observed for Cr30 by Ely and Johnson (13). This eclipse time is about the same as the host doubling time under the same incubation conditions. In contrast, *Caulobacter* phage CbK has the same genomic GC content and codon usage pattern as its host (14) and has a 40 minute eclipse period (13). Also consistent with this hypothesis is the observation that the gene for the major capsid protein, which

would be one of the most abundant proteins produced by the phage, has a GC content of 46.5% which is much higher than the genome average. Thus codon usage in this highly expressed gene is skewed towards the codon usage pattern of the host. Similar observations of major capsid genes with GC contents that are closer to those of the host have been made for other phages whose genomes have a different codon usage bias than their hosts (8, 19).

One way to compensate for differences in codon usage patterns would be to include tRNA genes in the phage genome, and the Cr30 genome contains five tRNA genes. Three of these tRNA genes produce tRNAs that recognize codons (CCA, AGA, and GGA) that are used infrequently by the *C. crescentus* host. However, the other two tRNAs recognize either AUG or AAC codons which are used frequently by *C. crescentus*. Thus, the tRNAs produced by the Cr30 genome could improve the translation efficiency of only three of the 13 codons that are used at much higher frequencies by Cr30 than by the host bacterium and again the data are consistent with a slower rate of Cr30 mRNA translation. For example, the least used codon in the host genome is the leucine codon TTA (0.6 times per 1000 codons). In the Cr30 genome, TTA is used at a 25-fold higher frequency (15 per 1000), and it is used more frequently than the two leucine codons that are used most frequently by the NA1000 host bacteria. In fact as predicted by the codon usage analysis of Lightfield *et al.* (18), amino acids such as isoleucine, tyrosine, phenylalanine, asparagine and lysine that are coded by AT-rich codons are present at nearly twice the frequencies in Cr30 proteins than they are in NA1000 proteins, and the converse is also true. Amino acids such as glycine, alanine, and proline that are coded by GC-rich codons, are present at about half the frequencies in Cr30 proteins than they are in NA1000 proteins. Thus, both the codon usage pattern and the differences in the amino acid composition of the Cr30 proteins are likely to result in a slower rate of translation of Cr30 mRNAs.

This slow rate of translation may actually be beneficial to Cr30 since it is a lytic phage. If the phage were too efficient at replicating inside of its host, it might quickly eliminate all of the host bacteria in its environment leading to its own local extinction. In this context, Ardisson *et al.* (3) showed that the presence of a capsule protects cells in the stalked phase of the *C. crescentus* cell cycle against Cr30 infection and contributes to the relatively slow adsorption rate observed by Ely and Johnson (13). Since the mature stalked cells are the only cells capable of replication and division, their resistance to Cr30 infection would help ensure their survival while providing a constant source of immature swarmer cells that could serve as Cr30 hosts. A similar strategy is employed by the Caulobacter phage CbK which can only infect swarmer cells since it attaches via the flagellum and pili (16) that are present only in swarmer cells or predivisional cells that are about to divide. Thus, the stalked cells are immune to CbK infection but they continually produce a susceptible daughter cell.

Phylogenetic comparison

Protein BLAST comparisons (2) of individual CR30 genes to the GenBank database indicated a high degree of similarity to the T4-like phages that infect the marine photosynthetic bacterial genera *Synechococcus* and *Prochlorococcus*. However, the best matches tended to be to genes of the recently characterized T4-like ϕ M12 phage that infects *S. melioli*, with most amino acid identities ranging from 35% to 45% (7). Like ϕ M12, Cr30

does not have the photosynthetic genes or phosphate metabolism genes that are often associated with the marine cyanophages (15, 29). However, when the amino acid sequences of the 33 T4-like core genes were compared across 14 representative T4-superfamily genomes, the concatenated Cr30 and ϕ M12 genes formed a monophyletic group that was more closely related to the cyanophage genomes than to the pseudo T4 group of phage genomes (Fig. 1). Brewer *et al.* (7) showed that the ϕ M12 gp20 protein grouped with those of a large number of uncultured phages that were separate from the major two groups of T4-superfamily phages and suggested that ϕ M12 may represent a third group of phages within the T4-superfamily. Thus, Cr30 may be considered a second member of this new group of phages even though the genomes of the two phages are more divergent than those of any two members of the cyanophage or the pseudoT4 groups. Since ϕ M12 and Cr30 both infect *Alphaproteobacteria*, this third group of T4-like phages may be comprised of those that infect alphaproteobacterial hosts. However, two other genomes of T4-like phage infect alphaproteobacterial hosts, *Pelagibacter* phage HTVC008M (31) and *Sphingomonas* phage PAU (25), have been sequenced and neither is closely-related to Cr30 or ϕ M12.

An alignment of the Cr30 genome with the T4 core genes of ϕ M12 shows that the order of the genes that code for the structural proteins and DNA replication machinery is highly conserved (Fig. 2). One exception is a cluster of three genes (gp41, *uvsX*, and gp43) located upstream of the *regA* and gp62 genes in the ϕ M12 genome as well as in the genomes of T4 and several other T4-like cyanophages that infect *Synechococcus* and *Prochlorococcus* (7). In the Cr30 genome, this cluster is located between the gp33 and the *rnh* genes, and in contrast to most of the other core genes, these three Cr30 genes code for proteins that are more similar to the corresponding gene products of other T4-like phage than they are to the ϕ M12 proteins. Thus, these three genes may have been acquired from another T4-like phage genome so that they have a different phylogenetic history from the rest of the Cr30 genome. The Cr30 genome also differs from the ϕ M12 genome in that it contains a gp2 gene where gp4 is located in the ϕ M12 genome and it contains the *nrdAB* genes that are present in most T4-like phages but are absent from the ϕ M12 genome. In addition the Cr30 genome does not contain the *nrdC* or *nrdJ* genes which are present in the ϕ M12 genome. Thus, there are substantial differences in the gene content of the Cr30 genome even though the gene order is highly conserved when compared to that of the ϕ M12 genome.

Methods

Cr30 phage lysates were prepared as described by Ely and Johnson (13). Cr30 DNA was isolated using a standard phenol/chloroform/isoamyl alcohol procedure. The Cr30 genome was sequenced with 454 Life Sciences sequencing technology that resulted in 1800X coverage with an average read length of 209 bases. The resulting sequence information was assembled de novo into a single contig using the DNASTAR's Lasergene Genomics Suite (DNASTAR, Inc, Madison, USA). The Cr30 genome was annotated using RAST (4) and checked with MiCheck (9). Manual corrections to the annotation were performed in Artemis (26). Cr30 tRNA genes were identified using tRNAscan-SE (27).

Agarose plugs containing Cr30 DNA for PFGE was generated using the protocol of Dingwall *et al.* (10). Restriction digests using the *SpeI* and *SwaI* enzymes were performed

for 4 h according to the manufacturer's directions. Pulsed field gel electrophoresis was performed in a 1% agarose gel (1.5 g pulsed field gel agarose and 150 mL 1X SBA (35 mM Boric Acid, 10 mM NaOH, pH=8.5)).

Acknowledgments

This work was funded in part by National Science Foundation grant EF-0826792 and NIH grants R25GM066526 and R25GM076277 to BE.

References

1. Agabian-Keshishian N, Shapiro L. Stalked bacteria: Properties of deoxyribonucleic acid bacteriophage ϕ CBK. *J Virol.* 1970; 5:795–800. [PubMed: 4193836]
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–410.10.1093/nar/25.17.3389 [PubMed: 2231712]
3. Ardisson S, Fumeaux C, Bergé M, Beaussart A, Théraulaz L, Radhakrishnan SK, Dufrêne YF, Viollier PH. Cell cycle constraints on capsulation and bacteriophage susceptibility. *eLife.* 2014; 3:e03587.10.7554/eLife.03587.001
4. Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics.* 2008; 9:75.10.1186/1471-2164-9-75 [PubMed: 18261238]
5. Barrett JT, Croft RH, Ferber DM, Gerardot CJ, Schoenlein PV, Ely B. Genetic mapping with Tn5-derived auxotrophs of *Caulobacter crescentus*. *J Bacteriol.* 1982a; 151:888–898. [PubMed: 6284715]
6. Barrett JT, Rhodes CS, Ferber DM, Jenkins B, Kuhl SA, Ely B. Construction of a genetic map for *Caulobacter crescentus*. *J Bacteriol.* 1982b; 149:869–876.
7. Brewer TE, Stroupe ME, Jones K. The genome, proteome and phylogenetic analysis of phage Φ M12, the founder of a new group of T4-super family phages. *Virology.* 2014; 450–451:84–97.10.1016/j.virol.2013.11.027
8. Carbone A. Codon bias is a major factor explaining phage evolution in translationally biased hosts. *J Mol Evol.* 2008; 66:210–223.10.1007/s00239-008-9068-6 [PubMed: 18286220]
9. Cruveiller S, Le Saux J, Vallenet D, Lajus A, Bocs S, Médigue C. MICheck: A web tool for fast check of syntactic annotations of bacterial genomes. *Nucl Ac Res.* 2005; 33(Web Server issue):W471–9.10.1093/nar/gki498
10. Dingwall A, Shapiro L, Ely B. Analysis of bacterial genome organization and replication using pulsed-field gel electrophoresis. *Methods.* 1990; 1:160–168.10.1016/S1046-2023(05)80131-8
11. Ely B. The *Caulobacter crescentus* genetic map. *Genetic Maps.* 1990; 5:2.100–2.103.
12. Ely, B. Genetics of *Caulobacter crescentus*. *Bacterial Genetics Systems.* In: Miller, JH., editor. *Methods in Enzymology.* Vol. 204. San Diego: Academic Press, Inc; 1991. p. 372-84.
13. Ely B, Johnson RC. Generalized transduction in *Caulobacter crescentus*. *Genetics.* 1977; 87:391–99. [PubMed: 17248770]
14. Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, et al. The *Caulobacter crescentus* bacteriophage ϕ CbK: genomics of a canonical phage. *BMC Genomics.* 2012; 13:542. [PubMed: 23050599]
15. Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle CA, Weinbauer MG, Sandaa RA, Breitbart M. Development of *phoH* as a novel signature gene for assessing marine phage diversity. *Appl Environ Microbiol.* 2011; 77(21):7730–39.10.1128/AEM.05531-11 [PubMed: 21926220]
16. Guerrero-Ferreira RC, Viollier PH, Ely B, Poindexter JS, Georgieva M, Jensen GJ, Wright ER. A novel mechanism for bacteriophage adsorption to the motile bacterium *Caulobacter crescentus*. *Proc Nat Acad Sci USA.* 2011; 108:9963–9968.10.1073/pnas.1012388108 [PubMed: 21613567]
17. Johnson RC, Wood NB, Ely B. Isolation and characterization of bacteriophages for *Caulobacter crescentus*. *J gen Virol.* 1977; 37:323–35.10.1099/0022-1317-37-2-323

18. Lightfield J, Fram NR, Ely B. Across bacterial phyla distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PLoS ONE*. 2011; 6(3):e17677. [PubMed: 21423704]
19. Lucks JB, Nelson DR, Kudla GR, Plotkin JB. Genome landscapes and bacteriophage codon use. *PLoS Computational Biology*. 2008; 4(2):e1000001.10.1371/journal.pcbi.1000001 [PubMed: 18463708]
20. Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, Wheatley PJ, et al. The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine *Synechococcus*. *J Bacteriol*. 2005; 187:3188–320. [PubMed: 15838046]
21. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, et al. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol*. 2003a; 185:5220–5233. [PubMed: 12923095]
22. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. Bacteriophage T4 genome. *Microbiol Mol Biol Rev*. 2003b; 67:86–156. [PubMed: 12626685]
23. Nolan JM, Petrov V, Bertrand C, Krisch HM, Karam JD. Genetic diversity among five T4-like bacteriophages. *Virol J*. 2006; 3:30. [PubMed: 16716236]
24. Petrov VM, Nolan JM, Bertrand C, Levy D, Desplats C, Krisch HM, Karam JD. Plasticity of the gene functions for DNA replication in the T4-like phages. *J Mol Biol*. 2006; 361:46–68. [PubMed: 16828113]
25. Pope, WH.; Hua, J.; Hatfull, GF.; Hendrix, RW. Sequence of the Genome of Sphingomonas Phage PAU. National Center for Biotechnology Information; 2012. www.ncbi.nlm.nih.gov
26. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000; 16:10.944–45.10.1093/bioinformatics/16.10.944 [PubMed: 10812472]
27. 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:W686–89.10.1093/nar/gki366 [PubMed: 15980563]
28. Sullivan MB, Coleman M, Weigele P, Rohwer F, Chisholm SW. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol*. 2005; 3:e144. [PubMed: 15828858]
29. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology*. 2010; 12(11):3035–56.10.1111/j.1462-2920.2010.02280.x [PubMed: 20662890]
30. Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, et al. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol*. 2007; 9:1675–1695. [PubMed: 17564603]
31. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB, Giovannoni SJ. Abundant SAR11 viruses in the ocean. *Nature*. 2013; 494(7437):357–60.10.1038/nature11921 [PubMed: 23407494]

Fig. 1.

A phylogenetic tree based on T4-like phage core gene amino acid sequences. Phages P-HM2, P-SSM2, and P-SSM4 were isolated using *Prochlorococcus* sp. as the host; while phages S-PM2, S-ShM2 and syn9 were isolated using *Synechococcus* sp. as the host (20, 28–30). Cr30 infects *Caulobacter crescentus* (17). PhiM12 infects *Sinorhizobium meliloti* (7). Phage KVP40 infects *Vibrio parahaemolyticus* (21). Phage Aeh1 infects *Aeromonas hydrophila* (23). Phages T4, RB43 and RB49 infect *Escherichia coli* (22, 23). Phage 25 infects *Aeromonas salmonicida* (24).

Fig. 2.
An alignment of the ϕ M12 and Cr30 core genes. The numbers correspond to T4 Gp numbers. The approximate sizes of the spaces between shared genes are shown in parentheses. The general classes of gene function are color-coded as defined in the legend.