Arch Virol (2015) 160:2647–2650
DOI 10.1007/s00705-015-2548-2

ANNOTATED SEQUENCE RECORD

Complete genome sequence analysis and identification of putative metallo-beta-lactamase and SpoIIIE homologs in *Bacillus cereus* group phage BCP8-2, a new member of the proposed Bastille-like group

Paul Tetteh Asare[1] · Nadeeka Bandara[1] · Tae-Yong Jeong[1] · Sangryeol Ryu[2,3] · Jochen Klumpp[4] · Kwang-Pyo Kim[1]

Received: 17 January 2015 / Accepted: 20 July 2015 / Published online: 4 August 2015

Abstract *Bacillus cereus* group-specific bacteriophage BCP8-2 exhibits a broad lysis spectrum among food and human isolates (330/364) of *B. cereus* while not infecting *B. subtilis* (50) or *B. licheniformis* (12) strains. Its genome is 159,071 bp long with 220 open reading frames, including genes for putative methyltransferases, metallo-beta-lactamase, and a sporulation-related SpoIIIE homolog, as wells as 18 tRNAs. Comparative genome analysis showed that BCP8-2 is related to the recently proposed Bastille-like phages, but not with either SPO1-like or Twort-like phages of the subfamily *Spounavirinae*.

Bacteriophage BCP8-2 was isolated from Korean fermented food in an original scheme for alternative control of *B. cereus* in food [1]. It is a member of the family *Myoviridae* with a head (95 nm) and a long contractile tail (210 nm) and shows host specificity for members of the *B. cereus* group such as *B. cereus*, *B. thuringiensis*, *B. mycoides, B. weihenstephanensis* [1] and *B. anthracis* (unpublished data). Here, we determined an extensive host range of BCP8-2 among a large number (364) of *B. cereus* isolates obtained from either food or humans and analyzed its complete genome sequence and phylogenetic position.

Bacterial growth, high-titer phage preparation, host range study, and genomic DNA isolation of BCP8-2 were carried out according to a previous study [1]. Genome sequencing was carried out at Macrogen Inc., South Korea, using a 454 GS FLX Titanium sequencing system (Roche, Mannheim, Germany) (approximately 55x coverage with 344.78 nt of average read length). Coding sequences (CDSs) were predicted using Glimmer 3.02 [5] with a minimum open reading frame (ORF) length of 30. The proteins were functionally annotated using the BLASTP algorithm. tRNA genes were identified using the ARAGORN v1.2.36 program [8], and the codon usage of phage BCP8-2 was predicted using the Countcodon program (http://www.kazusa.or.jp/codon/). The codon usage of *Bacillus* spp. was retrieved from the Codon Usage Database [11]. Easyfig [17] and CoreGenes 3.0 [10] were employed to compare the phage genomes at the nucleotide and protein level, respectively. Maximum-likelihood trees were constructed for amino acid sequences of four markers (tail fiber protein, major capsid protein, DNA polymerase and terminase large subunit) using MEGA v6 as described previously [6].

An extensive host range study showed that BCP8-2 lysed 90.7 % (330/364 isolates) of *B. cereus* isolated from foods (91.8 %, 202/220) or human patients (89.9 %, 128/144). In contrast, none of the *B. subtilis* (50) and *B. licheniformis* (12) food isolates were sensitive to BCP8-2.

P. T. Asare and N. Bandara contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-015-2548-2) contains supplementary material, which is available to authorized users.

✉ Kwang-Pyo Kim
kpkim@jbnu.ac.kr

[1] Department of Food Science and Technology, College of Agriculture and Life Sciences, Chonbuk National University, Jeonju, Jeollabuk-do 561-756, Korea

[2] Department of Food and Animal Biotechnology, Center for Agricultural Biomaterials, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Korea

[3] Department of Agricultural Biotechnology, Center for Agricultural Biomaterials, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Korea

[4] Institute of Food, Nutrition and Health, ETH Zurich, Schmelzbergstrasse 7, 8092 Zurich, Switzerland

The complete genome of phage BCP8-2 (GenBank accession number KJ081346) is 159,071 bp long with an overall G+C content of 39.4 mol %, which is similar to other large-genome *B. cereus*-infecting *Myoviridae* phages [9]. By using Geneious software [7], a significantly larger read depth was observed in the region between 58,000 and 66,000 bp, which suggested the presence of a terminal redundancy region, as reported previously [18].

A total of 220 ORFs and 18 tRNAs were detected in the genome of BCP8-2 (Supplementary Fig. 1). Forty-five putative proteins exhibited significant homologies to functionally annotated proteins of bacteriophages in the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/), which were classified into a number of functional groups (Supplementary Table 1).

For host recognition, BCP8-2 ORF4 encodes a putative tail fiber protein, which shares sequence similarity (86 % identity) with a putative adsorption-associated tail protein in *Bacillus* phage Bc431v3 [6]. For injection of phage DNA into the host cytoplasm, ORF5 and ORF7 contain a cell wall-associated hydrolase domain and a glucosaminidase domain at the C-terminus, respectively. ORF11 features signal sequences at the N-terminal end and a 3D domain at the C-terminal end and is similar to the L-alanoyl-D-glutamate peptidase of Bc431v3 (88 %).

ORF155, ORF179, and ORF181 are predicted to encode a putative methyltransferase, and C-5 cytosine-specific DNA methylase I and II, respectively [14]. Phage BCP8-2 also contains at least 21 gene products possibly involved in nucleotide metabolism, replication, and transcription (Supplementary Table 1).

For DNA packaging and morphogenesis, ORF20 and ORF22 encode a major capsid protein and a prohead protease, respectively. The N-terminal region of ORF22 contains a peptidase U35 domain important for prohead assembly. ORF23 is located immediately upstream of the procapsid protease gene and could function as putative portal protein, forming the hole through which DNA enters the capsid [4]. A putative large terminase subunit is likely encoded by ORF35 and 37 in the BCP8-2 genome.

For host lysis, ORF38 contains a glycosyl hydrolase family 25 domain at the N-terminal end and an SH3 domain at the C-terminal end that might be responsible for breaking the glycosidic bond and attaching to the carbohydrate branch on the bacterial host cell wall, respectively. Using Protter v1.0 [12], we identified two transmembrane domains in the ORF166 product, which was annotated as the putative phage lysis protein, holin.

The presence of tRNAs in phage genomes has been argued to compensate for the difference in codon usage between the phage and the host, especially for codons that are poorly translated by the host machinery [3]. The BCP8-2 genome encodes 18 tRNA genes located between ORF26 and ORF28 (Supplementary Fig. 1), which includes tRNAs for seven codons that occur at a phage/host (*B. cereus*) ratio of $\geq$ 1.50, including Arg^{AGA} (phage to host codon usage ratio of 1.65), Ser^{AGC} (1.91), Cys^{TGC} (5.85), Tyr^{TAC} (2.80), Phe^{TTC} (1.98), His^{CAC} (2.56) and Leu^{CTA} (1.74) (Supplementary Table 2). It is intriguing to note that six of them (all but AGA for Arg) are poorly used by the host, suggesting that phage-encoded tRNA would help to achieve better translation of the phage genes. Similarly, Santos et al. previously reported the importance of phage-encoded genes and tRNAs in the broad host capacity of *Salmonella* phage PVP-SE1 [15].

Codon usage among other *B. cereus* stains (*B. cereus* [gbbct]: 6363 and *B. cereus* E33L [gbbct]: 5641; data not shown) and *B. cereus* group species (*B. thuringiensis* shown in Supplementary Table 1 and *B. anthracis* str. Ames [gbbct]: 5311 [data not shown]) were very similar each other (Supplementary Table 2) and thus their ratios were close each other.

Codon usage of non-*B. cereus* group spp. (*B. subtilis* and *B. licheniformis*) is notably different from that of *B. cereus* group spp. for a number of amino acids. For example, TTA (or CAA) and TTG (or CAG) are equally frequent for Leu (or for Gln) in non-*B. cereus* spp., whereas TTA (CAA) is largely preferred to TTG (CAG) in *B. cereus* group spp. (Supplementary Table 2). When codon usage of *B. subtilis* was compared with that of BCP8-2, seven codons (six of them are same as for *B. cereus* but Pro^{CCA} is included in *B. subtilis* instead of Ser^{AGC} in *B. cereus*) were found to be $\geq$ 1.5 of the ratio. Similar results are found for *B. licheniformis* (Supplementary Table 2). Taken together, these data suggest that tRNAs in the BCP8-2 genome may help to overcome codon bias in the host but not be related to host specificity.

Another hypothesis is that phage tRNA may help late lytic growth by compensating for the degradation or inadequacy of the host tRNAs [13]. We tried to validate this hypothesis in BCP8-2 (Supplementary Table 2). When codons of BCP8-2 in total and those of late genes (mainly encoding structural genes; ORF1 - 25 and 210 - 220) were analyzed, the frequencies of eight codons corresponding to BCP8-2 tRNA were higher in late genes (e.g., AGC and TGC, were more than two-fold higher). On the other hand, those of 10 codons were lower in late genes (e.g., GGA, ATG, ACA, TTA, TCA, CAA and CTA, which were more than two-fold lower). Similar results were observed for representative late genes (MCP [major capsid protein] and TSH [tail sheath protein]) in BCP8-2. These data suggest that the presence of tRNAs in BCP8-2 might not have evolved to help late lytic growth of the phage.

BCP8-2 ORF94 exhibited significant homology to a metallo-beta-lactamase superfamily protein, with 57 % (94 % query coverage) identity to its counterpart in *B.*

thuringiensis [3]. In addition, ORF115 encodes a putative FtsK/SpoIIIE-family protein with 30 % identity (94 % query cover) to a DNA translocase stage III sporulation protein (SpoIIIE) of *B. cereus.* Putative sporulation-related genes similar to BCP8-2 ORF115 were reported previously in other *Bacillus* phages (Bcp1 and Bc431v3) [6, 16]. However, no phage-encoded gene product has been experimentally proven to be directly related to bacterial sporulation.

CoreGenes analysis showed that BCP8-2 shares 30.00–31.82 % of its proteome with members of ICTV-classified Twort-like phages (Twort, A511, P100, G1, and K) and 20.45 % with SPO1-like phage SPO1. In addition, 4.09 % and 18.64 % correspondence was observed for ICTV-unassigned *Spounavirinae* phages phiEF24C and LP65, respectively. In contrast, BCP8-2 shares a significantly high percentage (37.73-80.00 %) of its proteome with other phages that have not been classified by ICTV but were recently proposed to belong to the Bastille-like group phages in *Spounavirinae* [2] and Bc431v3, whose gene products exhibited high similarity to those of BCP8-2 (Fig. 1A).

Amino-acid-sequence-based phylogenic studies using four single gene products (major capsid protein, tail sheath protein, terminase, and DNA polymerase) also clustered BCP8-2 with phages of the Bastille-like group (only major capsid protein data are shown in Fig. 1B). In addition, when Easyfig [17] was used to compare the whole genome sequence of BCP8-2 with those of other phages, it showed a significantly high level of shared synteny with the proposed Bastille-like phages (Supplementary Fig. 2).

These data collectively confirm the presence of a third distinct group of phages within the subfamily *Spounavirinae* and also show that BCP8-2 is a member of the newly proposed Bastille-like group phages in *Spounavirinae*.

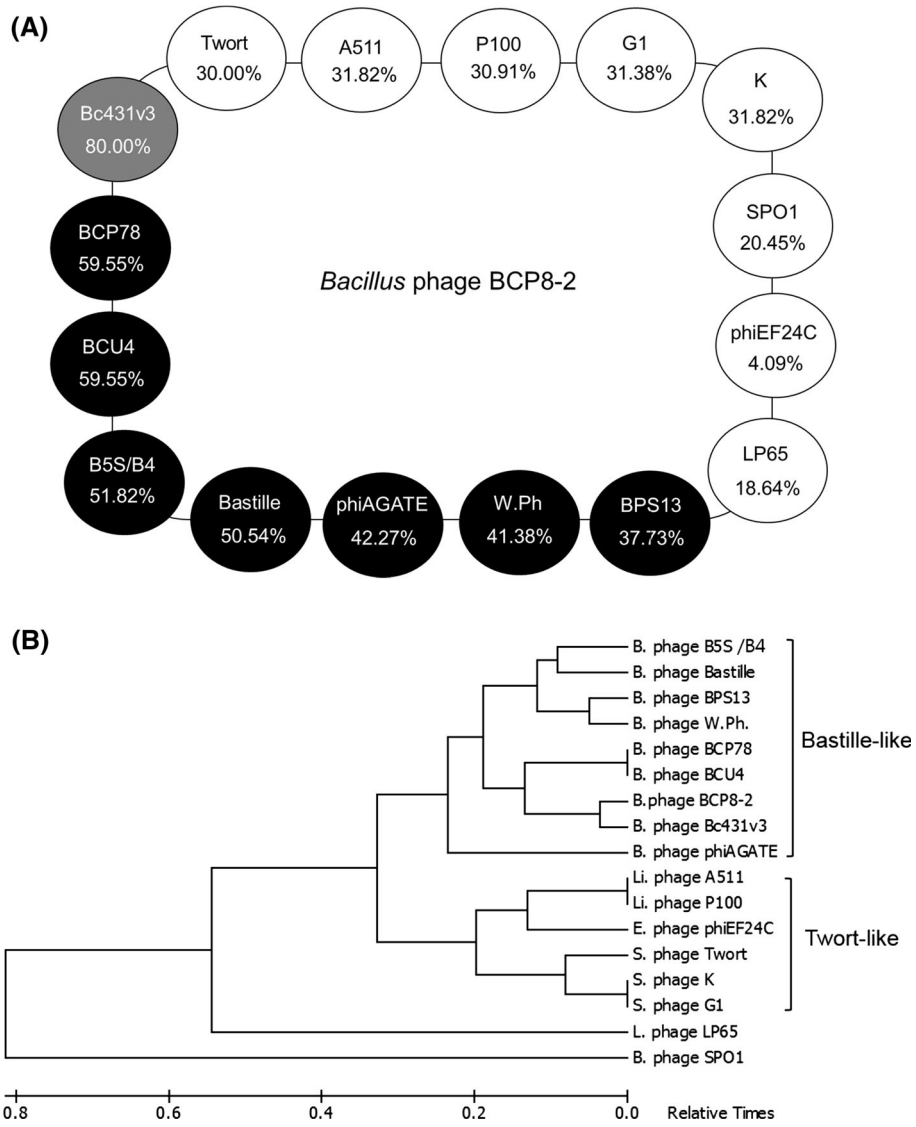

Fig. 1 (A) CoreGenes analysis of BCP8-2 against ICTV-classified *Spounavirinae* phages, indicated by white circles (Twort-like phages [Twort, A511, P100, G1, K), SPO1-like phage [SPO1], and unassigned phages [phiEF24C and LP65]), the previously proposed Bastille-like phages, indicated by black circles (BPS13, W.Ph, phiAGATE, Bastille, B5S/B4, BCU4, BCP78), and Bc431v3, indicated by a gray circle, whose gene products exhibited high similarities to those of BCP8-2. (B) Comparative phylogenetic analysis of major capsid proteins using the maximum-likelihood (ML) method

Acknowledgments This work was supported by the High Value-Added Food Technology Development Program (Project No. 112108-2 and 313037-3), Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) of Republic of Korea.

References

1. Bandara N, Jo J, Ryu S, Kim K-P (2012) Bacteriophages BCP1-1 and BCP8-2 require divalent cations for efficient control of *Bacillus cereus* in fermented foods. Food Microbiol 31:9–16
2. Barylski J, Nowicki G, Goździcka-Józefiak A (2014) The discovery of phiAGATE, a novel phage infecting *Bacillus pumilus*, leads to new insights into the phylogeny of the subfamily *Spounavirinae*. PLoS ONE 9:e86632
3. Bebrone C (2007) Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. Biochem Pharmacol 74:1686–1701
4. Casjens SR, Gilcrease EB (2009) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Methods Mol Biol 502:91–111
5. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679
6. El-Arabi TF, Griffiths MW, She Y-M, Villegas A, Lingohr EJ, Kropinski AM (2013) Genome sequence and analysis of a broad-host range lytic bacteriophage that infects the *Bacillus cereus* group. Virol J 10:48
7. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649
8. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16
9. Lee J-H, Shin H, Ryu S (2014) Characterization and comparative genomic analysis of bacteriophages infecting members of the *Bacillus cereus* group. Arch Virol 159:871–884
10. Mahadevan P, King JF, Seto D (2009) CGUG: in silico proteome and genome parsing tool for the determination of "core" and unique genes in the analysis of genomes up to *ca.* 1.9 Mb. BMC Res Notes 2:168
11. Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28:292
12. Omasits U, Ahrens CH, Müller S, Wollscheid B (2013) Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30:884–886
13. Pope WH, Anders KR, Baird M et al (2014) Cluster M mycobacteriophages Bongo, PegLeg, and Rey with unusually large repertoires of tRNA isotypes. J Virol 88:2461–2480
14. Ryazanova AY, Abrosimova L, Oretskaya T, Kubareva E (2012) Diverse domains of (cytosine-5)-DNA methyltransferases: structural and functional characterization. In: Dricu A (ed) Methylation—from DNA, RNA and histones to disease and treatment. InTech, Croatia, pp 29–69
15. Santos SB, Kropinski AM, Ceyssens P-J, Ackermann H-W, Villegas A, Lavigne R, Krylov VN, Carvalho CM, Ferreira EC, Azeredo J (2011) Genomic and proteomic characterization of the broad-host-range *Salmonella* phage PVP-SE1: creation of a new phage genus. J Virol 85:11265–11273
16. Schuch R, Pelzek AJ, Fazzini MM, Nelson DC, Fischetti VA (2014) Complete genome sequence of *Bacillus cereus* sensu lato bacteriophage Bcp1. Genome Announc 2:e00334-00314
17. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010
18. Yee LM, Matsumoto T, Yano K, Matsuoka S, Sadaie Y, Yoshikawa H, Asai K (2011) The genome of *Bacillus subtilis* phage SP10: a comparative analysis with phage. Biosci Biotechnol Biochem 75:944–995