Arch Virol (2014) 159:3137–3140
DOI 10.1007/s00705-014-2140-1

ANNOTATED SEQUENCE RECORD

Complete genome sequence of enterobacteria phage 4MG, a new member of the subgroup "PVP-SE1-like phage" of the "rV5-like viruses"

Minsik Kim · Sunggi Heu · Sangryeol Ryu

Received: 20 February 2014 / Accepted: 30 May 2014 / Published online: 18 June 2014

Abstract A novel virulent enterobacteria phage, 4MG, which was isolated from soil near a sewer, belongs to the family *Myoviridae*, as it possesses an isometric head and a long contractile tail. The complete genome of 4MG consists of a double-stranded DNA with a length of 148,567 bp, a G + C content of 46.3 %, 271 open reading frames (ORFs), and 21 tRNAs. Bioinformatic analysis revealed that 4MG highly resembles "rV5-like viruses" but can be separated, together with *Salmonella* phage PVP-SE1 and *Cronobacter sakazakii* phage vB_CsaM_GAP31, as part of the subgroup "PVP-SE1-like phage".

In 2009, Lavigne and colleagues classified 102 *Myoviridae* phages using Coregenes/CoreExtractor proteome comparison techniques [1]. Many related phages were grouped together to form specific subfamilies or genera, but several phages with no apparent relatives were also identified as singleton phages of the family *Myoviridae*. Until the complete genome sequence of *Salmonella* phage PVP-SE1 supported the suggestion of a new genus denoted "rV5-like viruses" [2, 3], *Escherichia coli* O157:H7-sepcific phage rV5 was considered to be such a singleton *Myoviridae* phage group. Recently, *E. coli* K-12-specific phage vB_EcoM_FV3 (hereafter, FV3), *Cronobacter sakazakii*-specific phage vB_CsaM_GAP31 (hereafter, GAP31), avian pathogenic *E. coli* strain-specific phage phAPEC8, and pathogenic *E. coli* strain-specific phage phi92 were also assigned as members of the "rV5-like viruses" [4–7]. Although these myoviral rV5-like phages share similar morphology and genomic contents, it is noticeable that rV5 and FV3 are closer to each other than the other related phages PVP-SE1 and GAP31 and phAPEC8 and phi92 [3, 4, 6].

Using *E. coli* MG1655 as the host bacterium, we isolated phage 4MG from soil near a domestic sewage drain in Yongin-si, Gyeonggi-do, South Korea, according to methods described elsewhere [8]. In a conventional double-layer agar overlay assay with LB medium, 4MG formed relatively small (~1 mm in diameter) clear plaques with diffuse edges after overnight incubation at 37 °C. The determination of host range through an overlay assay revealed that 4MG infects many of the common laboratory *E. coli* K-12 strains with various EOPs (efficiency of plating), but not the B strains (Table S1 in the Online Resource). Phage 4MG from a single plaque was propagated with MG1655, and the phages were then precipitated with 10 % (wt/vol) polyethylene glycol 6000 in 1 M NaCl and subjected to CsCl density gradient ultracentrifugation at 78,500 × g and 4 °C for 2 h [8]. The opalescent phage band was collected and dialyzed against dialysis buffer (50 mM Tris-HCl [pH 8.0], 10 mM NaCl, and 10 mM $MgSO_4$). The purified and concentrated phage stock was

Electronic supplementary material The online version of this article (doi:10.1007/s00705-014-2140-1) contains supplementary material, which is available to authorized users.

M. Kim · S. Ryu (✉)
Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, South Korea
e-mail: sangryu@snu.ac.kr

S. Heu
Microbial Safety Division, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-707, South Korea

S. Ryu
Department of Food and Animal Biotechnology, Research Institute for Agriculture and Life Sciences, and Center for Food and Bioconvergence, Seoul National University, Seoul 151-921, South Korea

used to analyze the morphology of phage 4MG by transmission electron microscopy (LEO 912AB TEM; Carl Zeiss, Jena, Germany) at 80 kV as described previously [8]. As shown in Fig. 1a, 4MG is morphologically assigned to the family *Myoviridae* because it possesses an isometric head with an apical diameter of 94 nm and a long contractile tail of 121 by 18 nm (n = 8) with five to six thin tail fibers.

The nucleic acid of 4MG was manually extracted from a phage stock using proteinase K/sodium dodecyl sulfate [9] and subjected to pyrosequencing using a 454 Genome Sequencer FLX Titanium by Macrogen (Seoul, South Korea). A single contig assembled with quality-filtered reads was analyzed using the tools GeneMarkS [10], Glimmer 3.02 [11], and FgenesB (Softberry, Inc., NY, USA) for prediction of open reading frames (ORFs), BlastP [12], InterProScan [13], and NCBI Conserved Domain Database [14] for annotation of the predicted ORFs, and tRNAscan-SE [15] for searching tRNAs. The complete genome of 4MG is composed of a double-stranded 148,567-bp-long DNA with a G + C content of 46.3 %. Approximately 73 % of the predicted ORFs (198 of 271 ORFs) showed no homologies with functional proteins and were thus annotated as 'hypothetical

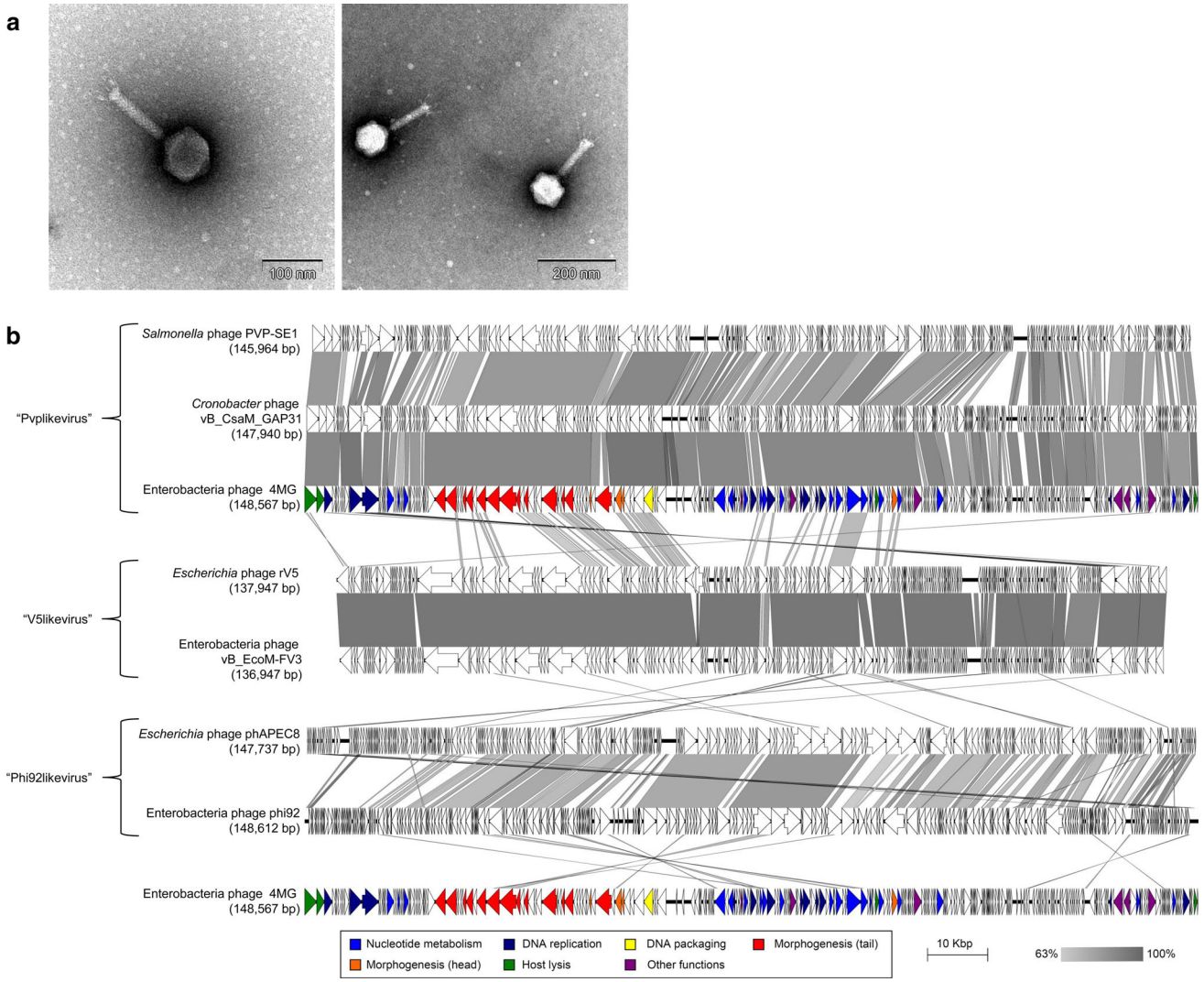

Fig. 1 Transmission electron micrograph of enterobacteria phage 4MG and DNA-level alignment of the genomes of rV5-like viruses. (a) Negatively stained (2 % uranyl acetate, pH 4.0) 4MG particles were observed by TEM (31,500× and 20,000× for the left and right panels, respectively). The scale bar at the bottom right corner indicates 100 and 200 nm for the left and right panel, respectively. (b) The whole genomes of 4MG (GenBank accession number KF550303), PVP-SE1 (GU070616.1), GAP31 (JN882284.1), rV5 (DQ832317.1), FV3 (JQ031132.1), phi92 (FR775895.2), and phAPEC8 (JX561091.1) were compared at the DNA level using Easyfig. The gray regions indicate high sequence similarity between the genomes. The predicted functions of 4MG proteins are indicated by different colors (a color figure is available online). Note that the genome of GAP31 was rearranged to allow an easier comparison with the others

protein'. Among them, 19 ORFs are unique for 4MG (Table S2). The other 27 % ORFs encode putative functional proteins that were assigned to the following clusters (Fig. 1b): nucleotide metabolism (e.g., phosphoribosyl pyrophosphate synthetase [ORF093], thymidylate synthase [ORF130], and ribonucleotide reductases [ORF133, ORF134]), DNA replication (e.g., helicase [ORF003], DNA polymerase [ORF013], and ATP-dependent DNA ligase [ORF108]), DNA packaging (terminase large subunit [ORF074]), morphogenesis (tail fiber proteins [ORF039, ORF044, ORF049, ORF067], baseplate component [ORF046], tail tape measure protein [ORF055], and major capsid protein [ORF069]), and host lysis (lysozyme [ORF139] and cell wall hydrolase SleB [ORF270]) (Table S2). The absence of a lytic/lysogenic control module (e.g., phage repressor, antirepressor, and integrase) suggests that this clear plaque-forming phage 4MG is a virulent phage. Of the 21 tRNAs predicted in the 4MG genome, 15 tRNAs (e.g., tRNA-Arg, -Lys, -Pro, and -Thr) were found to be more frequently present in the 4MG phage than in its *E. coli* host, suggesting that these tRNA genes facilitate the translation of phage mRNAs.

A whole-genome comparison at the DNA level using Easyfig [16] suggested that 4MG highly resembles rV5-like phages (Fig. 1b). The similar virion dimensions and genome sizes of rV5-like phages also support the close relationship between these phages and 4MG (Table S3). In particular, the genomic contents and organization of 4MG were more similar to those of the "PVP-SE1-like phages" (i.e., PVP-SE1 and GAP31) than those of the other rV5-like phages (i.e., rV5, FV3, phAPEC8, and phi92). A comparison at the proteome level using CoreGenes3.0 [17] revealed that 201/244 (82.38 %) and 241/269 (89.59 %) proteins encoded by PVP-SE1 and GAP31, respectively, are homologous to those encoded by 4MG. *Salmonella* Typhimurium and its lipopolysaccharide-deficient derivatives [18] as well as other 16 *Salmonella* ssp. were found to be completely insusceptible to 4MG, but *C. sakazakii* ATCC29544 and BAA-894 were susceptible (Table S1), indicating that 4MG is more closely related to *C. sakazakii* phage GAP31 than to *Salmonella* phage PVP-SE1. In contrast, comparisons with rV5 or Phi92 revealed only 98/233 (42.06 %) or 68/250 (27.2 %) shared proteins, respectively. However, a common feature of the rV5-like viruses, i.e., the high percentage of small ORFs (less than 100 residues) throughout the genome [2], was

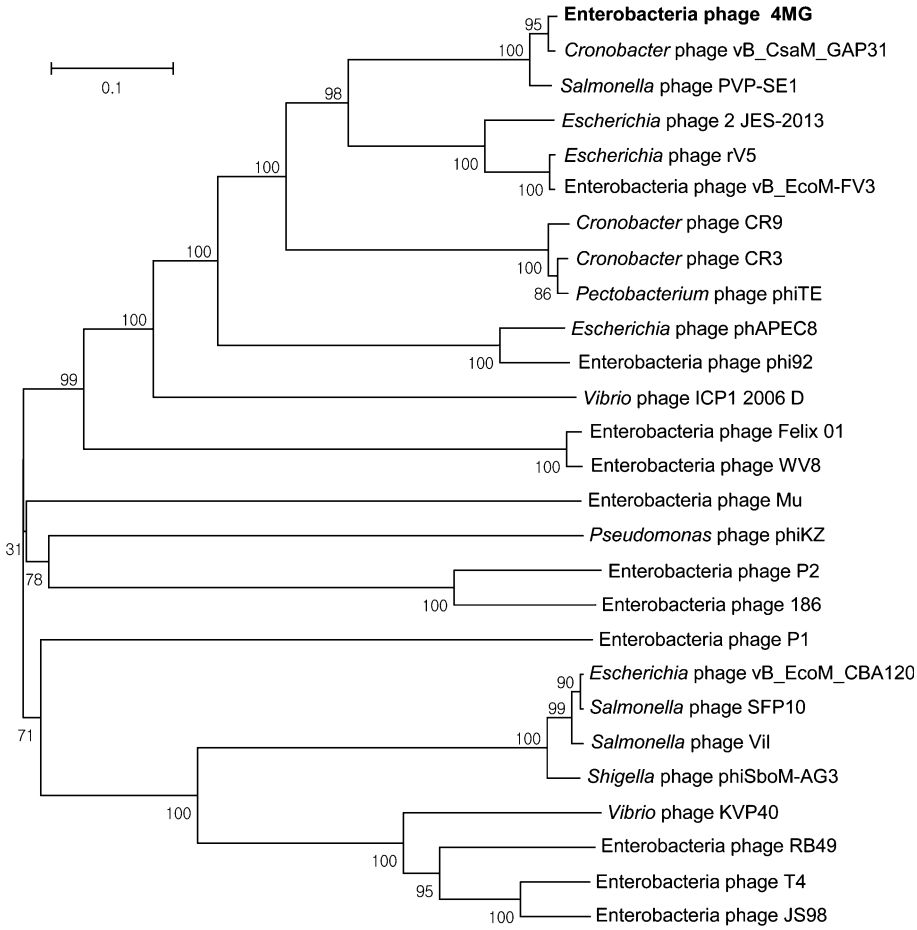

Fig. 2 Neighbor-joining phylogenetic tree analysis based on the alignment of the amino acid sequences of the major capsid proteins from various *Myoviridae* family phages from GenBank. The bootstrap probabilities are indicated at the nodes

observed in all of these phages. Although several introns and homing endonucleases are often found in the genomes of members of the family *Myoviridae*, including the rV5-like phages, 4MG encodes only one, free-standing HNH endonuclease (ORF119), which is homologous to that found in GAP31 (ORF155), PVP-SE1 (ORF122), rV5 (ORF91), and FV3 (ORF85).

Phylogenetic analysis of the conserved major capsid protein (MCP) further demonstrated the evolutionary relationships between these phages. The amino acid sequences of the MCP from 4MG and other *Myoviridae* family phages from GenBank were aligned and subsequently subjected used for construction of a phylogenetic tree using MEGA5 [19]. Phage 4MG formed a separate branch with GAP31 and PVP-SE1 in the resulting phylogenetic tree, whereas rV5 and FV3, phAPEC8, and phi92 generated markedly distinct clades (Fig. 2). Altogether, the present study suggests that the enterobacteria phage 4MG belongs to the subgroup "PVP-SE1-like phage" of the "rV5-like viruses".

Nucleotide sequence accession number

The complete genome sequence of 4MG is available at GenBank under the accession number KF550303.

Acknowledgments This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ009842)", Rural Development Administration, Republic of Korea.

References

1. Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, Ackermann HW, Kropinski AM (2009) Classification of *Myoviridae* bacteriophages using protein sequence similarity. BMC Microbiol 9:224
2. Santos SB, Kropinski AM, Ceyssens PJ, Ackermann HW, 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
3. Kropinski AM, Waddell T, Meng J, Franklin K, Ackermann HW, Ahmed R, Mazzocco A, Yates J 3rd, Lingohr EJ, Johnson RP (2013) The host-range, genomics and proteomics of *Escherichia coli* O157:H7 bacteriophage rV5. Virol J 10:76
4. Abbasifar R, Kropinski AM, Sabour PM, Ackermann HW, Alanis Villa A, Abbasifar A, Griffiths MW (2012) Genome sequence of *Cronobacter sakazakii* myovirus vB_CsaM_GAP31. J Virol 86:13830–13831
5. Schwarzer D, Buettner FF, Browning C, Nazarov S, Rabsch W, Bethe A, Oberbeck A, Bowman VD, Stummeyer K, Mühlenhoff M, Leiman PG, Gerardy-Schahn R (2012) A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. J Virol 86:10384–10398
6. Truncaite L, Šimoliūnas E, Zajančkauskaite A, Kaliniene L, Mankevičiūte R, Staniulis J, Klausa V, Meškys R (2012) Bacteriophage vB_EcoM_FV3: a new member of "rV5-like viruses". Arch Virol 157:2431–2435
7. Tsonos J, Adriaenssens EM, Klumpp J, Hernalsteens JP, Lavigne R, De Greve H (2012) Complete genome sequence of the novel *Escherichia coli* phage phAPEC8. J Virol 86:13117–13118
8. Kim M, Ryu S (2011) Characterization of a T5-like coliphage, SPC35, and differential development of resistance to SPC35 in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. Appl Environ Microbiol 77:2042–2050
9. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
10. Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618
11. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
13. Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847–848
14. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2007) CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res 35:D237–D240
15. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964
16. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010
17. 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
18. Kim M, Ryu S (2012) Spontaneous and transient defence against bacteriophage by phase-variable glucosylation of O-antigen in *Salmonella enterica* serovar Typhimurium. Mol Microbiol 86:411–425
19. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739