Virology 460-461 (2014) 138–146

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Supersize me: *Cronobacter sakazakii* phage GAP32

Reza Abbasifar [a], Mansel W. Griffiths [a], Parviz M. Sabour [b], Hans-Wolfgang Ackermann [c], Katrien Vandersteegen [d], Rob Lavigne [d], Jean-Paul Noben [e], Argentina Alanis Villa [a], Arash Abbasifar [a], John H.E. Nash [f], Andrew M. Kropinski [f,g,*]

[a] Canadian Research Institute for Food Safety, University of Guelph, Guelph, ON, Canada N1G 2W1
[b] Agriculture and Agri-Food Canada, Guelph Food Research Centre, Guelph, ON, Canada N1G 5C9
[c] Department of Microbiology-Infectiology and Immunology, Faculty of Medicine, Université Laval, Quebec, QC, Canada
[d] Laboratory of Gene Technology, Katholieke Universiteit Leuven, Leuven, Belgium
[e] Biomedical Research Institute and Transnational University Limburg, School of Life Sciences, Hasselt University, Diepenbeek, Belgium
[f] Public Health Agency of Canada, Laboratory for Foodborne Zoonoses, Guelph, ON, Canada N1G 3W4
[g] Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada N1G 2W1

ARTICLE INFO

Article history:
Received 9 December 2013
Returned to author for revisions
5 January 2014
Accepted 3 May 2014
Available online 2 June 2014

Keywords:
Cronobacter
Bacteriophage
vB_CsaM_GAP32
Myoviridae
Genome
Proteome
SAR lysin

ABSTRACT

Cronobacter sakazakii is a Gram-negative pathogen found in milk-based formulae that causes infant meningitis. Bacteriophages have been proposed to control bacterial pathogens; however, comprehensive knowledge about a phage is required to ensure its safety before clinical application. We have characterized *C. sakazakii* phage vB_CsaM_GAP32 (GAP32), which possesses the second largest sequenced phage genome (358,663 bp). A total of 571 genes including 545 protein coding sequences and 26 tRNAs were identified, thus more genes than in the smallest bacterium, *Mycoplasma genitalium* G37. BLASTP and HHpred searches, together with proteomic analyses reveal that only 23.9% of the putative proteins have defined functions. Some of the unique features of this phage include: a chromosome condensation protein, two copies of the large subunit terminase, a predicted signal-arrest-release lysin; and an RpoD-like protein, which is possibly involved in the switch from immediate early to delayed early transcription. Its closest relatives are all extremely large myoviruses, namely coliphage PBECO4 and *Klebsiella* phage vB_KleM-RaK2, with whom it shares approximately 44% homologous proteins. Since the homologs are not evenly distributed, we propose that these three phages belong to a new subfamily.

Introduction

Cronobacter sakazakii is an opportunistic pathogen found in both the environment and a variety of foods (Gurtler et al., 2005; Yan et al., 2012). The mortality of the infections can reach 80%, and antibiotics have no dramatic effect on clinical results (Norberg et al., 2012; Ray et al., 2007; Drudy et al., 2006; Lai, 2001).

Bacteriophages (phages), viruses that infect bacteria, have been applied as alternative control agents due to their high specificity and effectiveness in killing bacterial pathogens without affecting commensal microbiota. Phage biocontrol of many pathogens such as *Salmonella*, *Campylobacter* and *Listeria* in the food industry is recognized in the USA and Europe (Spricigo et al., 2013; Endersen et al., 2014) and could be particularly relevant for control of *Cronobacter* because of its intrinsic antibiotic resistance (Lai, 2001).

The objective of this study was to characterize at the molecular level, the newly isolated phage, vB_CsaM_GAP32 (GAP32). Genomic sequencing will assist in prediction of the likely reproductive behavior of the phage, and determine whether the phage has lysogenic capability, an essential and important consideration for a phage intended for use in prophylaxis and therapy.

Results and discussion

Isolation, host range and morphology

Cronobacter phage vB_CsaM_GAP32 (GAP32) was isolated by enrichment from a wastewater treatment plant in Guelph (Ontario, Canada). It produced small (1 mm), clear plaques on its host *C. sakazakii* strain HPB 3290, a cerebrospinal fluid isolate. It lysed 86% of the *C. sakazakii* strains tested; including both of the

* Corresponding author at: Public Health Agency of Canada, Laboratory for Foodborne Zoonoses, 110 Stone Road West, Guelph, ON, Canada N1G 3W4. Tel.: +1 519 826 2614; fax: +1 519 822 2280.
E-mail address: akropins@uoguelph.ca (A.M. Kropinski).

http://dx.doi.org/10.1016/j.virol.2014.05.003
0042-6822/

Cronobacter muytjensii, half of the *Cronobacter malonaticus*, plus one *Cronobacter dublinensis* strain (Table 1). It did not lyse *Cronobacter universalis*, *Enterobacter helveticus*, *Enterobacter cloacae*, *Escherichia coli* ATCC 11229, *Shigella flexneri*, *Klebsiella pneumoniae*, *Hafnia alvei*, *Salmonella enterica serovar* Typhimurium, *Serratia odorifera* or *Yersinia enterocolitica* strains. The host range and infection parameters of GAP32 suggests that this phage has great potential for controlling *C. sakazakii*.

Table 1
The host range of *Cronobacter* phage GAP32 on 21 *Cronobacter* strains determined by spot testing (16 h incubation at 30 °C). The results were recorded as 3 (very clear zone of complete lysis on the bacterial lawn), 2 (clear/turbid zone of lysis), 1 (turbid lysis) and 0 (no lysis).

Bacteria / Phage Isolate	*C. sakazakii* 2855	*C. sakazakii* 2870	*C. sakazakii* 2871	*C. sakazakii* 2876	*C. sakazakii* 3253	*C. malonaticus* 3263	*C. sakazakii* 3199	*C. muytjensii* 51329	*C. sakazakii* 3290	*C. dublinensis* 3169	*C. malonaticus* 3267	*C. muytjensii* 3270	*C. universalis* 3287	*C. sakazakii* 130/3	*C. sakazakii* 236/04	*C. sakazakii* 324/04	*C. sakazakii* 354/03	*C. sakazakii* 974/03	*C. sakazakii* 1084/04	*C. sakazakii* 1103/03	*C. malonaticus* 1154/04
32	3	2	0	3	0	3	2	2	3	1	3	3	0	3	2	3	1	3	3	3	0

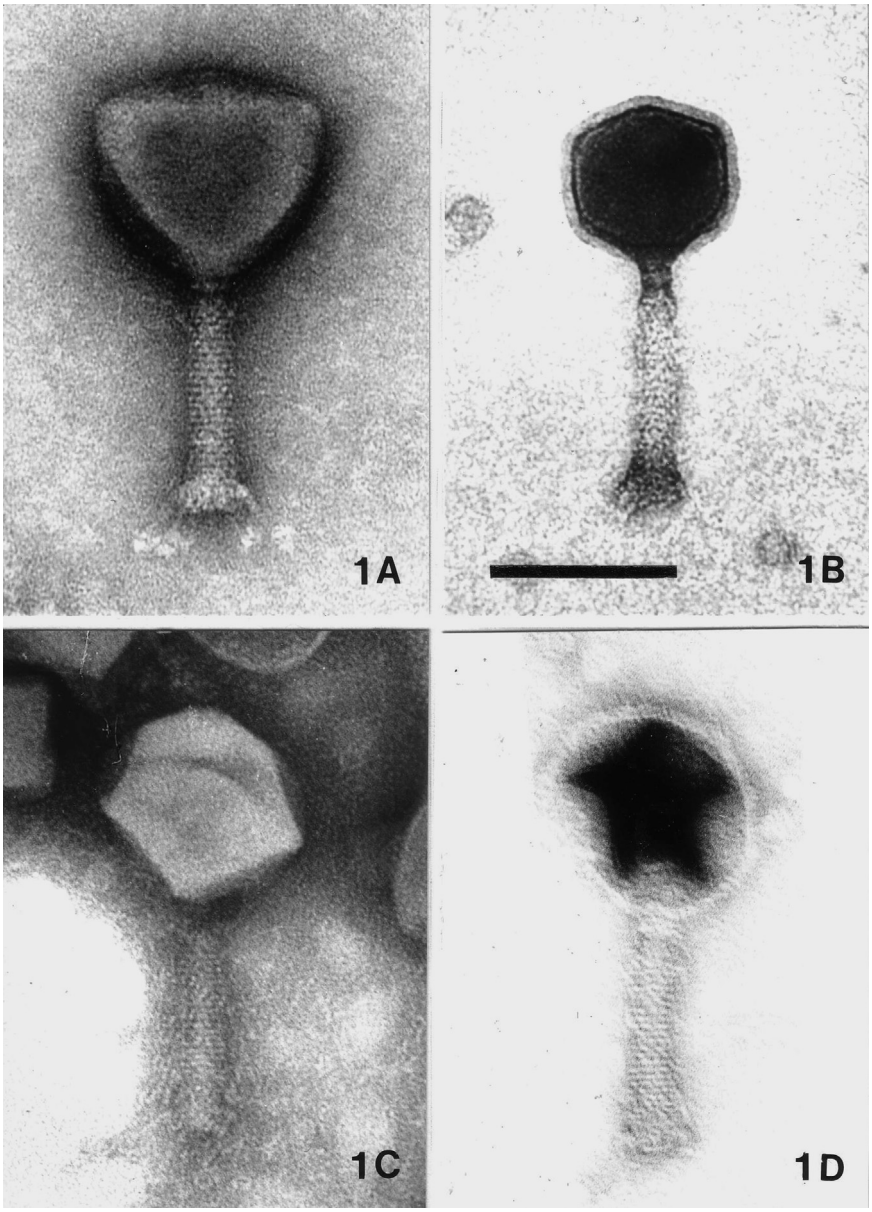

Fig. 1. Electron micrographs of GAP32 negatively stained with phosphotungstate ((A) and (C)) and uranyl acetate ((C) and (D)). Positively stained heads are characteristically shrunken (B). (D) Shows a rare photograph of an empty head that was partially filled by the stain. The center of the star corresponds to an apex.

GAP32 belongs morphologically to the family *Myoviridae* (Fig. 1), displaying a large head (113 nm) and a comparatively short (118 × 23 nm) contractile tail. The head is an icosahedron as evidenced by the presence of six-sided or five-sided capsids. Some empty heads show pentagonal stars. The neck is 10 nm long and carries a collar of 20 × 3 nm. The normal tail sheath displays 24–25 striations. The contracted sheath, of 40 × 30 nm, which was only rarely observed, remains attached to the base plate, showing a protruding core 5 nm wide. The base plate generally resembles a thick globule of 30–40 × 12 nm, with indistinct, about 15 nm-long spikes, and terminal spherules.

Genome sequencing and analysis

Phage GAP32 has a giant double-stranded DNA genome of 358,663 bp with a G+C content of 35.5%, making it the second largest sequenced phage genome after *Bacillus megaterium* phage G (497,513 bp; GenBank accession number JN638751). We identified 571 genes including 545 protein coding sequences (CDSs) and 26 tRNA-encoding genes (Supplementary data; Table S2). This is more genes than are present in by the smallest known bacterium, *Mycoplasma genitalium* G37, which only encodes 470 proteins (Fraser et al., 1995).

Within the phage realm, GAP32 is most closely related to *Escherichia* O157-specific phage PBECO4 (348,113 bp; 550 CDSs) (Kim et al., 2013) with which it shares 243 protein homologs, equivalent to 44.6% of the predicted CDSs showing recognizable homology. Another similar virus is *Klebsiella* phage vB_KleM-RaK2 (345,809 bp; 534 CDSs) (Simoliunas et al., 2013, 2012) with which it shares 236 homologs (43.3% homology). Fig. 2 shows the genome map of phage GAP32 in comparison to phages PBECO4 and RaK2 created using CGview with TBLASTX comparisons between the related viruses (Stothard and Wishart, 2005).

On the basis of sequence similarity search, analysis for transmembrane domains, proteomic and HHpred analyses, the proteins specified by GAP32 were classified into 193 hypothetical proteins, 30 hypothetical membrane proteins, 180 conserved hypothetical proteins and 10 conserved hypothetical membrane proteins (Supplementary data; Table S2). We could only confidently predict the function of 130 of this phage's proteins. None of the gene products were related to toxins, integration proteins (integrase, transposase), plasmid association (ParAB homologs) or antimicrobial resistance, the presence of which would exclude it from consideration as a therapeutic agent. In the following sections we shall be dealing with some of these proteins.

Nucleotide metabolism, DNA synthesis and recombination

Like many of the large myoviruses, GAP32 codes for several ribonucleotide reductases (RNR), which reduce ribonucleotides to deoxyribonucleotides. GAP32 specifies Class I (NrdA [gp298]-NrdB [gp300]) and Class III (NrdD [gp138]) RNRs (Dwivedi et al., 2013; Lundin et al., 2009). Its genome also contains two CMP/dCMP deaminases (gp109, 134), the former containing a PHA02588 (2.09e−34) resembling similar proteins in RaK2, PBECO4, and *Yersinia*

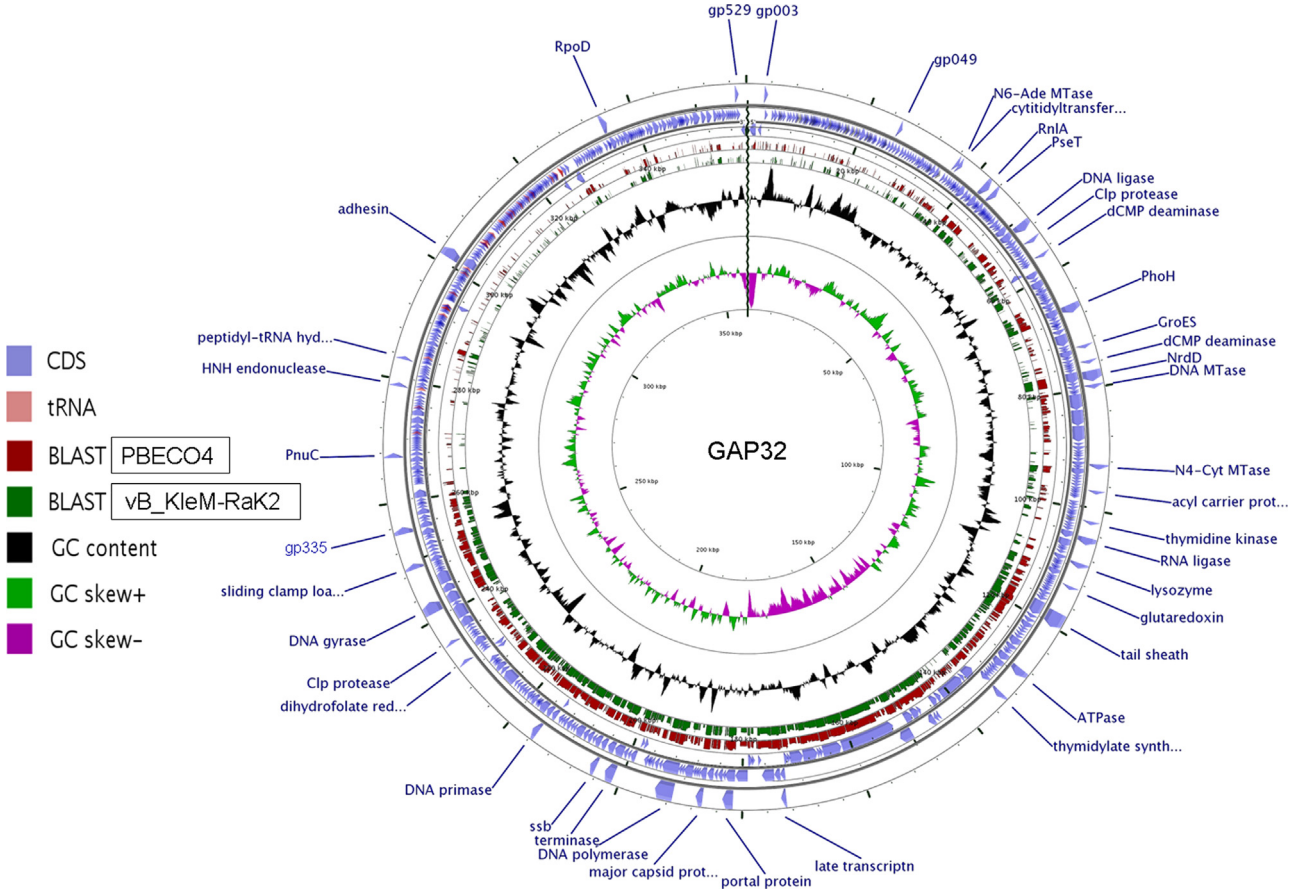

Fig. 2. Genome map of phage vB_CsaM_GAP32 in comparison to coliphage PBECO4 (red) and *Klebsiella* phage vB_KleM-RaK2 (green) created using CGview employing TBLASTX analyses (Stothard and Wishart, 2005). Working from the outside the concentric rings correspond to the named genes, genes on the plus strand, genes on the negative strand, BLAST homologs to PBECO4, BLAST homologs to RaK2, GC content; and, GC skew.

phage phiR1-RT (YP_007236048), while the latter lacks this motif. Deoxycytidylic deaminase provides a source of dUMP for the *de novo* synthesis of thymidylic acid, from dUMP, catalyzed by thymidylate synthetase (gp217) and involving tetrahydrofolate which functions as a methyl donor (product of action of dihydrofolate reductase [gp308]). Thymidine kinase (gp170) phosphorylates UdR or TdR to give dUMP or TMP. Other nucleotide converting enzymes include a dNMP kinase (gp197).

One of the unique features of this virus is that it encodes a nicotinamide mononucleotide transporter (gp354, PnuC; pfam04973 [1.43e−16]; COG3201 [1.20e−13]) with homologs to proteins in *Salmonella* phages PVP-SE1 (YP_004893838) and SSE-121 (AFU63697) and among the *Pasteurellaceae*. Nicotinate mononucleotide adenylyltransferase (gp112) catalyzes the conversion of β-nicotinate D-ribonucleotide to NAD. Thus GAP32, like the cyanophages, appears to synthesize homologs that complement host proteins (Millard et al., 2004; Sandaa et al., 2008) by recovering nicotinamide mononucleotide from the medium. It is worth noting that *Cronobacter* spp. also synthesize PnuC homologs.

This phage would appear to specify several potential enzymes involved in DNA degradation (gp260, 275, 295, 370), and the usual cast of enzymes involved in DNA synthesis (DNA polymerase [gp323], helicases and primase [gp276, 285, 286], sliding clamp proteins [gp327, 328], DNA gyrase [gp319] and ligase [gp99]) and recombination (gp239, 240, 241 and 272) proteins.

Lysis

GAP32 specifies three proteins, which possess bacteriophage T4-like lysozyme domains (gp180, 233, 234). Since we have mass spectroscopic evidence that gp234 is a structural protein, the endolysin candidates for this phage include gp180 and gp233. Both of these do possess C-terminal bacteriophage T4-like lysozyme (cd00735) and baseplate hub subunit and tail lysozyme (PHA02596) motifs. Considering its location within the genome, we have tentatively assigned the endolysin function to gp180. *In silico* analyses of this 35 kDa protein reveal an N-terminal transmembrane domain and that the C-terminus would be extracytoplasmic. The first 16 amino acids are predicted to be in the cytoplasm and contain 4 strongly basic amino acids with a predicted pI of 10. In this regard, this protein resembles P1 Lyz (YP_006484) and φKMV gp45 lysin (NP_877484), which contain signal-arrest-release motifs (SAR) (Xu et al., 2004; Briers et al., 2011). Like the coliphage P1 lysin, GAP32 gp180 contains a glycine–glutamic acid pair at residues 123 and 124 at which cleavage may occur. The lysozyme domain is fully located downstream of these residues. While this virus specifies large numbers of hypothetical and conserved hypothetical proteins with transmembrane domains, none have homology to known holins. The presence of a N-terminal TMD and a SpII cleavage signal in gp160/gp161, respectively, suggest these proteins may encode the Rz/Rz1 spannin pair (Summer et al., 2007).

Transcription

Seven of the 17 predicted RpoD-like promoters (Supplementary data) lay upstream of small CDSs encoding hypothetical or conserved hypothetical protein, which might be involved in host takeover. Using MEME analysis and visual inspection, we discovered a regulatory motif. A WebLogo (Crooks et al., 2004) for this motif (Fig. 3; Supplementary data; Table S3), contains a classic extended −10 domain (Mitchell et al., 2003) and a −35 region, which diverges from the classic (TTGACA) motif. We believe that these may represent delayed early promoters, which may require gp490, and possibly gp488 for transcription. Gp490 is an acidic 38 kDa protein that contains a TIGR02937 (1.93e−05) RNA polymerase−70 like domain, plus PF04542 (*E*-value: 7.13e−06) and PF04539 (2.78e−04) domains, which are, respectively, RNA polymerase sigma −70 region 2 and 3. The latter domain is involved in both −10 promoter recognition and core RNA polymerase binding. Its closest homologs are from phages PBECO4 and RaK2, with the next closest bacterial hit being a protein (YP_002462259) in the green nonsulfur bacterium *Chloroflexus aggregans* DSM 9485.

As many large myoviruses, GAP32 also specifies a T4 gp55-like late transcription factor (gp238), which suggests another gene switch occurs, from the delayed early expression to late transcription. The consensus sequence for the late promoters of the *Viunalikevirus* genus is CTAAATA (Adriaenssens et al., 2012), while that for the T4 phages is TATAAATA (Nechaev and Geiduschek, 2008; Geiduschek and Kassavetis, 2010). Again, using MEME on upstream sequences associated with structural proteins we were able to identify a motif ATAAATA directly upstream of several genes (*221*, *222*, *231*, 237, (*249*) and *254*), which are involved in the synthesis of structural components. Its verification as the late promoter will require experimental data.

In addition, we were able to identify 72 putative rho-independent terminators (Supplementary data; Table S3).

Protein and lipid synthesis

A number of interesting proteins that play roles in translation were identified in GAP32. The first of these is a histidyltRNA guanylyltransferase, which is involved in the synthesis of pppGp-tRNAHis. In eukaryotes the addition of the guanine nucleotide is essential for aminoacylation by His-tRNA synthetase (Jackman and Phizicky, 2008). This protein has rarely been observed in phage genomes; being restricted to *Bacillus cereus* phage vB_BceM-Bc431v3 (YP_007677042), *Bacillus megaterium* phage G (AEO93637), *Caulobacter* phage CcrColossus (YP_006988534), *Pseudomonas* phage 201φ2-1 (YP_001957040) and, *Sinorhizobium* phage ΦM12 (AGR47633). It is not found in either PBECO4 or RaK2.

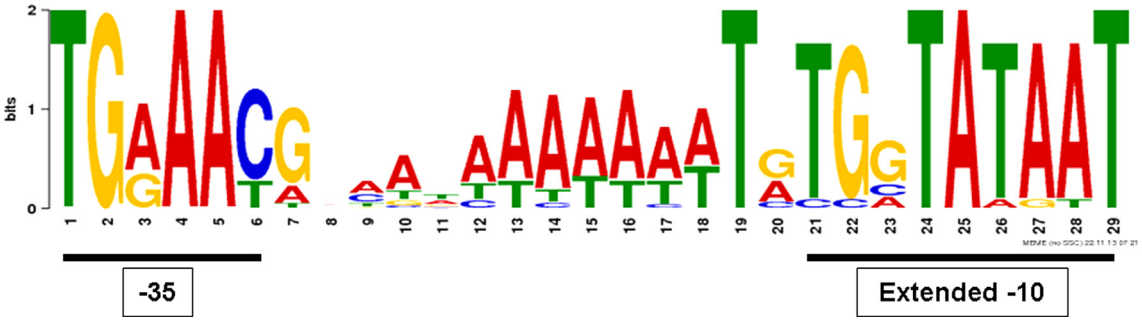

Fig. 3. Sequence logo of the putative delayed early promoters of phage GAP32 calculated from the data presented in Supplementary file 2 (Table S2); and, created using WebLogo (http://weblogo.berkeley.edu/) (Crooks et al., 2004).

Gp131 is a 47 kDa protein containing TIGR00234 (2.10e−105), COG0162 (1.49e−132) and PRK05912 (8.20e−167) tyrosyl-tRNA synthetase motifs. There are no phage homologs, and the closest proteins are synthetases from *Vibrio proteolyticus* (WP_021 706075) and *Listeria monocytogenes* (YP_002349927). Interestingly, GAP32 does encode a tyrosyl-tRNA.

Peptidyl-tRNA hydrolases play an essential role in protein synthesis by releasing tRNAs bound to nascent peptides (Das and Varshney, 2006). These proteins are common to *Mycobacterium* myoviruses, and GAP32 gp382 homologs can also be found in *Sinorhizobium* phage ΦM12 and RaK2 along with several *Caulobacter* phages.

The 18.9 kDa GAP32 protein gp313 contains PRK00028 (1.00e−33), COG0290 (7.56e−30) and TIGR00168 (1.42e−23) motifs all defining translation initiation factor IF-3. Among phages, homologs are only to be found in RaK2 and PBECO4 with the closest bacterial homologs present in the *Clostridiales*. This protein binds to 30S ribosomes and is responsible for assembly fidelity of the initiation complex (ribosome, mRNA, initiator tRNA).

The 84 amino acid gp159 contains PTZ00171 (7.72e−13) and PRK00982 (1.72e−12) acyl carrier protein motifs. This protein plays a significant role in fatty acid biosynthesis, requiring 4′-phosphopantetheine as a covalently attached cofactor. Acyl carrier protein homologs have been identified in several *Vibrio* phages including *Vibrio helene* phage 12B3 (YP_007877402) and *Prochlorococcus* phage P-HM1 (YP_004322452) but are absent from PBECO4 and RaK2. Interestingly, gene *69* encodes a protein that is structurally related to phosphopantetheine adenylyltransferase, which is involved in the biosynthetic pathway of Coenzyme A, specifically in the formation of 3′-dephospho-CoA. Homologs of this nucleotidyltransferase are found in PBECO4 and Rak2. Further research will be required into fatty acid or lipoprotein synthesis in these phages.

Packaging

Most phages encode two proteins TerS and TerL, which make up the DNA packaging (terminase) complex. In the case of GAP32, two proteins Gp268 and Gp269 are both defined, on the basis of homology and HHpred analyses, as large subunit terminases (TerL). Interestingly, the mass of Gp268 is more in keeping with the size of TerS. This protein contains a "17[PHA02533], large terminase protein" domain (*E*-value 1.49e−12) and is homologous to *Synechococcus* and *Prochlorococcus* phage proteins. The larger protein also contains this domain (*E*-value 1.03e−76) with a "Terminase_6[pfam03237], Terminase-like family" domain (*E*-value 9.19e−16); and shows similar homologs. The only known phage encoding two large terminase subunits is ΦM12, a transducing phage infecting *Sinorhizobium* (Brewer et al., 2014).

Gp1 is a 385 amino acid residue protein which a Conserved Domain Database scan reveals to contain a "ATS1[COG5184], Alpha-tubulin suppressor and related RCC1 domain-containing proteins" domain (*E*-value 2.96e−11), while InterProScan 4 reveals a "IPR009091 Regulator of chromosome condensation 1/beta-lactamase-inhibitor protein II" (*E*-value 9.6e−26). Since GAP32 gp1 is a structural protein, we postulate that it might be involved in phage chromosome condensation within the capsid, in a manner analogous to the proteins present in *Pseudomonas aeruginosa* phage φKZ (Wu et al., 2012).

Structural proteome analysis

A number of genes specifying proteins involved in morphogenesis were identified on the basis of BLASTP homologs or HHpred analyses, but as a supplement we analyzed the structural proteome of GAP32 by SDS-PAGE, trypsin digestion and mass spectrometry (ESI-MS/MS). The denaturing-polyacrylamide gel (Fig. 4) was sliced into 48 pieces and peptides from 49 predicted GAP32 proteins were experimentally identified (Table 2; Supplementary data, Table S4). The number of unique identified peptides and the corresponding protein sequence coverage ranges from 1 to 19 and 1.5% to 80.3%, respectively. No clear cases of proteolytic cleavage/maturation could be observed from the MS data. A reliable proline spectrum justified the single-peptide identification of gp234 and gp237.

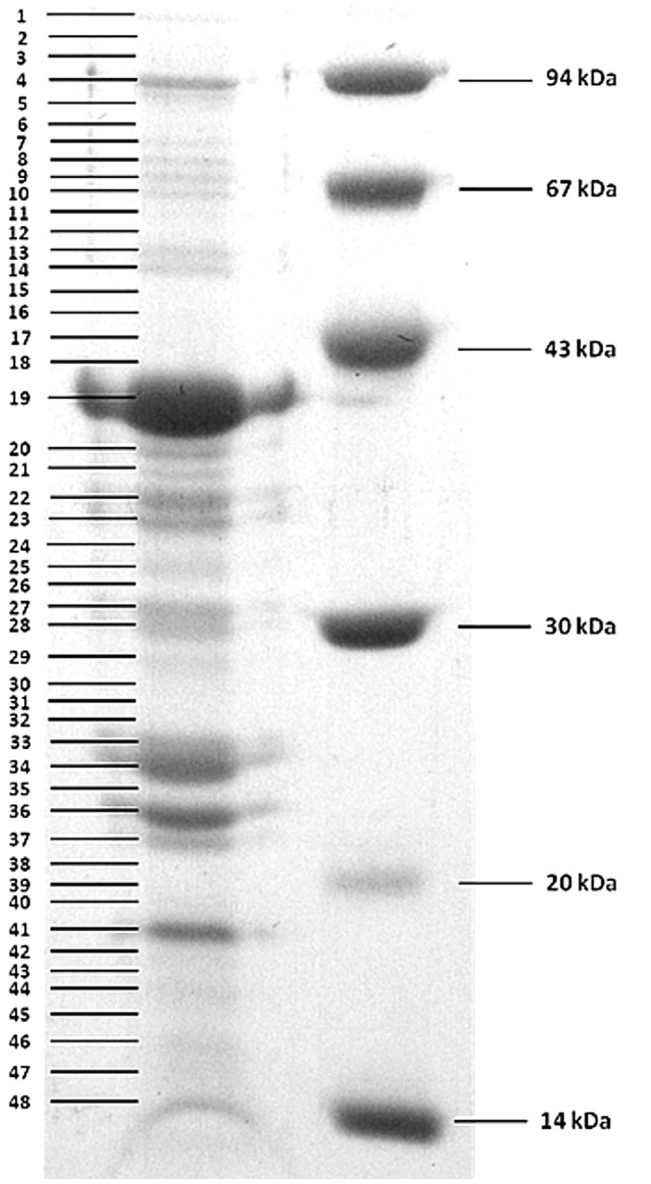

Fig. 4. Coomassie G-250 stained SDS-PAGE gel of the structural proteome of phage GAP32. A low molecular weight marker is shown at the right. The numbers on the right correspond to the analyzed gel pieces of which the identified proteins are given in Table 2.

For each identified gene product (gp) of phage GAP32 the band number(s) (according to Fig. 4), the molecular weight in kDa, the number of unique identified peptides (>99% protein identification probability with manual validation), the sequence coverage in percent, the putative function according to HHpred and the corresponding *e*-value are given.

The identification of other GAP32 proteins throughout the SDS-PAGE gel (Fig. 4) illustrates their abundance in the phage particle as well, for example the major tail protein gp111 and the tail sheath protein gp198. As such, the molecular weights of several proteins do not always correspond to the position of a

Table 2
Characteristics of the ESI-MS/MS identified proteins of phage GAP32.

Gp	Molecular mass (kDa)	Number of identified peptides	Sequence coverage (%)
1	41.5	9	31.9
25	30.8	6	27.7
32	19.0	4	30.6
111	30.2	9	53.1
118	16.7	2	20.9
144	99.6	2	3.3
145	99.7	5	11.4
198	95.6	19	45.0
199	25.4	4	20.5
200	29.1	2	12.8
204	21.0	2	18.8
205	20.8	2	13.6
206	52.7	3	8.9
218	68.4	12	36.7
219	24.2	5	38.7
220	154.1	9	10.8
221	30.1	2	10.5
222	41.4	9	38.8
223	16.7	2	20.5
224	21.8	2	13.4
225	40.6	6	23.0
227	51.3	3	10.8
228	57.3	2	4.6
230	381.2	5	2.3
231	131.4	5	6.2
232	14.8	2	13.2
234	97.0	1	1.5
244	78.4	10	23.0
245	65.2	12	29.5
248	23.4	4	23.8
250	42.1	14	65.0
254	56.9	11	19.7
264	22.4	11	80.3
265	19.1	9	73.1
284	36.6	7	34.7
293	23.8	3	20.0
297	20.9	7	57.3
311	19.9	6	54.1
315	20.7	2	15.8
316	61.8	6	14.2
317	95.9	12	23.1
321	88.3	6	11.5
330	25.8	3	20.3
339	20.9	2	13.6
475	41.2	2	6.1
515	10.3	2	33.3
530	23.4	5	3.7
531	19.2	2	20.0

specific protein band on the gel. The major capsid protein (gp250) could be assigned by its location in the SDS-PAGE gel, as the most prominent band.

The mass spectrometric identification of the 49 GAP32 proteins was supported by a HHpred search in which a higher *e*-value was taken into account as well (Table 2). Several identified GAP32 proteins appeared to show similarity to structural proteins of *Escherichia* phage T4, which was already clear from the genome annotation. In addition, homologies with the major (gp111) and minor (gp219) tail protein and the tail length tape measure protein (gp244) are found among the structural proteins of *Escherichia* phages P2 and Mu, *Helicobacter* phage HP1, and *Prochlorococcus* phage P-SSM2 were observed. Finally, four proteins (gp1, gp25, gp118, gp206) without a predicted structural function were found.

Conclusions

Giant enterobacterial phages are rare, but not exceptional as suggested by occasional reports or samples of similar phages which H.-W.A. received over the years for identification (Table 3). *Cronobacter* phage vB_CsaM_GAP32 possesses the second largest phage genome yet analyzed. Interestingly, one might have expected that such a giant phage required special isolation approaches (Serwer et al., 2004, 2007), but this was not the case, nor was it in the isolation of other giant phages (Table 3).

The genome and proteome analysis allowed for a comprehensive overview of the encoded gene products. Although the number of unknown proteins remain high, no know toxicity or lysogeny associated genes were found, suggesting this phage could be a candidate for phage therapy applications. Also, the proteogenomics-based identification of the structural proteins provides a first step in the structural analysis of this phage and experimentally verifies the function of 10% of the protein coding sequences.

Based on previously defined taxonomy parameters (Lavigne et al., 2008, 2009), the lack of DNA sequence identity, the existence of species-specific proteins, and the lack of an even distribution of homologs across the genomes we propose that GAP32, PBECO4 and RaK2 be classified to a new subfamily.

Though this phage lacks many proteins considered to be core proteins of the *Tevenvirinae* (Petrov et al., 2010), the annotation of this phage revealed that 20.1% of the proteins were homologous to those of coliphage T4, which raises the question as to whether this phage should be considered part of a T4 superfamily.

Material and methods

Bacterial strains

Twenty-one *Cronobacter* strains (including 14 *C. sakazakii* strains) were used for isolation, propagation, and determination of host range of phage GAP32 (Table S1). To grow *Cronobacter* strains and to isolate and propagate the phages, Tryptic Soy Broth (TSB), Tryptic Soy Agar (TSA), and Tryptose Top Agarose (TSB+0.5% agarose) (Fisher Scientific, Ottawa, ON) were used.

Bacteriophage isolation

Lytic bacteriophages against *C. sakazakii* were isolated from samples of untreated sewage (wastewater treatment plant, Guelph, ON, Canada), using the method described by Van Twest and Kropinski (2009). To prepare stocks, the phage was propagated by the top agar overlay method described by Adams (1959), allowing for confluent lysis. After incubation at 30 °C overnight, 3 ml of SM buffer was added to each plate and kept at 4 °C overnight. The top layer and liquid from all the plates was recovered using sterile glass rods. The tubes were centrifuged at 6000 × *g* for 20 min at 4 °C, and the supernatant was transferred to another tube and treated with chloroform. For each phage stock, the titre was determined by preparing 10-fold serial dilutions and tested using the overlay method described previously. The phage lysate was refrigerated at 4 °C.

Host range study

A 6–8 h culture of each *Cronobacter* strain (in TSB) was inoculated on TSA plates with sterile cotton swabs and allowed to dry. Three drops of 20 μl of each phage sample were spotted onto the plate and allowed to dry. Plates were examined for lysis after 16 h incubation at 30 °C, for the presence of plaques.

Table 3
Isolation and sightings of phages with similar morphology to GAP32.

Phage	Host	Head diameter (nm)	Tail length (nm)	Calibration	References and comments
121	*Proteus vulgaris*	150*	150*		Nacesco et al. (1969)
Type 121	*E. coli* strain 70.1	116	116 × 16	Catalase	Ackermann and Nguyen (1983). More similar phages were enriched by *E. coli* O112:B11 and O128:B12
CW2, CW3	*E. coli*	122*	115 × 20*	T4-tails	Isolated by E. Kutter, Olympia, WA, USA, 2010
F78E	*E. coli*	116*	114 × 18*	T4 tails	Isolated by J. Azeredo, Braga, Portugal, 2007
Nephage 1	*Salmonella* sp.?	121*	115*	T4 tails	Isolated by D.R. Bhatta, not purified, observed in a mixture of phages, 2005
PBECO4	*E. coli* O157:H7	132	125		Kim et al. (2013)
vB_KleM-RaK2	*Klebsiella* sp.	123	128 × 21.5	T4 tails, catalase	Simoliunas et al. (2012, 2013)
GAP32	*Cronobacter sakazakii*	117	113 × 18	T4 tails	This work

* Phosphotungstate only.

Electron microscopy

A sterile high-titer phage lysate was centrifuged at 25 000 × *g* for 60 min, using a Beckman (Palo Alto, CA) J2-21 centrifuge and a JA19.1 fixed-angle rotor at 4 °C. The pellet was washed twice in ammonium acetate (0.1 M, pH 7.0) under the same conditions. Phages were stained with uranyl acetate (2%, pH 4.5) or phosphotungstate (2%, pH 7.0) and examined under a Philips EM 300 electron microscope. Specimens were mounted on Athene type copper grids. Accelerating voltage was 60 kV. Nominal magnification was × 29,700 and magnification was controlled by means of T4 phage tails. Phage heads were measured between opposite apices. A total of 40 phage particles were measured.

Sequencing

The DNA of selected phages was extracted and purified by Midi Lambda DNA purification kit (Qiagen Midi25; Mississauga, ON, Canada), and the genomic sequence was determined using pyrosequencing (454 technology; McGill University and Génome Québec Innovation Centre; Montreal, QC, Canada). The data, representing 57-fold coverage, was assembled using Newbler v2.5.3. (Roche, Branford, CT) by the sequencing center. No attempt was made to determine the physical genome ends, and, since at the time of completion the sequence was unique, the genome was annotated as received. Its sequence is available from GenBank through accession number JN882285.

Bioinformatic analysis

The genome was annotated by Rapid Annotations using Subsystems Technology (myRAST) with gene calls verified using Kodon (Applied Maths, Austin, TX). Transfer RNAs were predicted using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy, 1997). For each protein,,,, the number of amino acids, molecular weight and the isoelectric point was calculated by Batch MW and pI Finder (http://greengene.uml.edu/programs/FindMW.html). Homologs were identified using BatchBLAST (http://greengene.uml.edu/programs/ NCBI_Blast.html), and the BLASTP feature of Geneious Pro R6 (Biomatters; Auckland, New Zealand) against the Viruses (taxid:10239) organism database. To predict protein motifs Transmembrane Hidden Markov Model (TMHMM) (http://www.cbs.dtu.dk/services/TMHMM-2.0/) (Kall et al., 2004; Sonnhammer et al., 1998), Phobius (http://phobius.sbc.su.se) (Kall et al., 2004), and Pfam (http://pfam.sanger.ac.uk/) (Finn et al., 2010) were used. To detect potential bacterial toxins the proteome of phage GAP32 was screened against a proprietary database of 78 bacterial toxins (Carter et al., 2012) using the BLAST feature of BioEdit (Tippmann, 2004). HHpred was used to detect remote protein homology and conduct the structure prediction (Hildebrand et al., 2009; Soding et al., 2005). Total proteome comparisons were made using CoreGenes 3.5 (Turner et al., 2013).

A genome diagram was constructed using CGView (Stothard and Wishart, 2005) employing TBLASTX with *Escherichia* phage PBECO4 (Kim et al., 2013) (Accession number: KC295538) and *Klebsiella* phage vB_KleM-RaK2 (Simoliunas et al., 2013, 2012) (Accession number: JQ513383). CoreGenes (Turner et al., 2013) was used to examine the phages for conserved proteins.

Potential promoters were discovered using extractUpStreamDNA at http://lfz.corefacility.ca/extractUpStreamDNA/ to extract 100bp of DNA sequence upstream from every gene. This was submitted to MEME (Multiple Em for Motif Elicitation) at http://meme.nbcr.net/meme/cgi-bin/meme.cgi (Bailey et al., 2009). In addition, the sequence was visually screened using Kodon for sequences closely related to the consensus host promoter TTGACA(N17)TATAAT. Rho-independent terminators were found using ARNold at http://rna.igmors.u-psud.fr/toolbox/arnold/ (Gautheret and Lambert, 2001; Macke et al., 2001), with the predictions verified using MFOLD QuikFold employing RNA energy rules 3.0 (Zuker and Zuker, 2003).

Proteomic analysis

Extraction of phage proteins from a high-titer GAP32 stock solution ($> 10^{11}$ pfu/ml) was performed by methanol/chloroform extraction (1:1:0.75, v/v/v). The resulting protein pellet was dissolved in loading buffer (1% SDS, 6% sucrose, 100 mM dithiothreitol, 10 mM Tris pH 6.8, 0.0625% bromophenol blue (Moak and Molineux, 2004), loaded onto a 12% SDS-PAGE gel and subjected to gel electrophoresis. Subsequent staining with Simply Blue™ Safe Stain (Invitrogen Ltd, Paisley, UK) revealed the GAP32 structural proteome. Gel fragments were cut out, subjected to trypsinization (Shevchenko et al., 1996) and analyzed using tandem electrospray ionization-mass spectrometry (ESI-MS/MS) on a LCQ Classic (ThermoFinnigan; Thermo Scientific, France) equipped with a nano-LC column switching system as described by Lavigne et al. (2006). The analysis of the mass spectrometric RAW data was carried out using Proteome Discoverer software v.1.3 (Thermo Scientific) with build-in Sequest v.1.3.0339 and interfaced with an in-house Mascot v.2.4 server (Matrix Science). MS/MS spectra were searched against a database containing all host proteins and all 'stop-to-stop' protein sequences in all six frames of phage GAP32. Peptide scoring for identification was

based on following search criteria: enzyme trypsin, maximum missed cleavages 2, precursor mass tolerance of 3 Da and a fragment mass tolerance 1 Da. Carbamidomethylation of cysteine and oxidation of methionine, histidine and tryptophan were set as fixed and dynamic modifications, respectively. Result files from both search engines were evaluated in Scaffold v.3.6.1 (Proteome Software) and reported ORFs had a minimal peptide and protein identification probability of 95% and 98%, respectively.

Acknowledgments

The authors thank Franco Pagotto, and Roger Stephan for providing bacterial strains. This work was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (NSERC IRCRJ 127376-07).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2014.05.003.

References

Ackermann, H.W., Nguyen, T.M., 1983. Sewage coliphages studied by electron microscopy. Appl. Environ. Microbiol. 45, 1049–1059.
Adams, M.D., 1959. Bacteriophages. Interscience Publishers, Inc., New York.
Adriaenssens, E.M., Ackermann, H.W., Anany, H., Blasdel, B., Connerton, I.F., Goulding, D., Griffiths, M.W., Hooton, S.P., Kutter, E.M., Kropinski, A.M., Lee, J. H., Maes, M., Pickard, D., Ryu, S., Sepehrizadeh, Z., Shahrbabak, S.S., Toribio, A.L., Lavigne, R., 2012. A suggested new bacteriophage genus: *Viunalikevirus*. Arch. Virol. 157, 2035–2046.
Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208.
Brewer, T.E., Stroupe, M.E., Jones, K.M., 2014. The genome, proteome and phylogenetic analysis of *Sinorhizobium meliloti* phage ΦM12, the founder of a new group of T4-superfamily phages. Virology 450–451, 84–97.
Briers, Y., Peeters, L.M., Volckaert, G., Lavigne, R., 2011. The lysis cassette of bacteriophage ϕKMV encodes a signal-arrest-release endolysin and a pinholin. Bacteriophage 1, 25–30.
Carter, C.D., Parks, A., Abuladze, T., Li, M., Woolston, J., Magnone, J., Senecal, A., Kropinski, A.M., Sulakvelidze, A., 2012. Bacteriophage cocktail significantly reduces *Escherichia coli* O157: H7 contamination of lettuce and beef, but does not protect against recontamination. Bacteriophage 2, 178–185.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.
Das, G., Varshney, U., 2006. Peptidyl-tRNA hydrolase and its critical role in protein biosynthesis. Microbiology 152, 2191–2195.
Drudy, D., Mullane, N.R., Quinn, T., Wall, P.G., Fanning, S., 2006. *Enterobacter sakazakii*: an emerging pathogen in powdered infant formula. Clin. Infect. Dis. 42, 996–1002.
Dwivedi, B., Xue, B., Lundin, D., Edwards, R.A., Breitbart, M., 2013. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. BMC Evol. Biol. 13, -33.
Endersen, L., O'Mahony, J., Hill, C., Ross, R.P., McAuliffe, O., Coffey, A., 2014. Phage therapy in the food industry. Annu. Rev. Food Sci. Technol..
Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., 2010. The Pfam protein families database. Nucleic Acids Res. 38, D211–D222.
Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D. M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison III, C.A., Venter, J.C., Fraser, C. M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., Venter, J.C., 1995. The minimal gene complement of *Mycoplasma genitalium*. Science 270, 397–403.
Gautheret, D., Lambert, A., 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J. Mol. Biol. 313, 1003–1011.
Geiduschek, E.P., Kassavetis, G.A., 2010. Transcription of the T4 late genes. Virol. J. 7 (288), 288.
Gurtler, J.B., Kornacki, J.L., Beuchat, L.R., 2005. *Enterobacter sakazakii*: a coliform of increased concern to infant health. Int. J. Food Microbiol. 104, 1–34.
Hildebrand, A., Remmert, M., Biegert, A., Soding, J., 2009. Fast and accurate automatic structure prediction with HHpred. Proteins 77 (9), 128–132 (128-32).
Jackman, J.E., Phizicky, E.M., 2008. Identification of critical residues for G-1 addition and substrate recognition by tRNA(His) guanylyltransferase. Biochemistry 47, 4817–4825.
Kall, L., Krogh, A., Sonnhammer, E.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.
Kim, M.S., Hong, S.S., Park, K., Myung, H., 2013. Genomic analysis of bacteriophage PBECO4 infecting *Escherichia coli* O157:H7. Arch. Virol. 158, 2399–2403.
Lai, K.K., 2001. *Enterobacter sakazakii* infections among neonates, infants, children, and adults. Case reports and a review of the literature. Medicine (Baltimore) 80, 113–122.
Lavigne, R., Darius, P., Summer, E.J., Seto, D., Mahadevan, P., Nilsson, A.S., Ackermann, H.-W., Kropinski, A.M., 2009. Classification of *Myoviridae* bacteriophages using protein sequence similarity. BMC Microbiol. 9, 224.
Lavigne, R., Noben, J.P., Hertveldt, K., Ceyssens, P.J., Briers, Y., Dumont, D., Roucourt, B., Krylov, V.N., Mesyanzhinov, V.V., Robben, J., Volckaert, G., 2006. The structural proteome of *Pseudomonas aeruginosa* bacteriophage ϕKMV. Microbiology 152, 529–534.
Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W., Kropinski, A.M., 2008. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. Res. Microbiol. 159, 406–414.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Lundin, D., Torrents, E., Poole, A.M., Sjoberg, B.M., 2009. RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. BMC Genomics 10, 589.
Macke, T.J., Ecker, D.J., Gutell, R.R., Gautheret, D., Case, D.A., Sampath, R., 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res. 29, 4724–4735.
Millard, A., Clokie, M.R.J., Shub, D.A., Mann, N.H., 2004. Genetic organization of the psbAD region in phages infecting marine *Synechococcus strains*. Proc. Nat. Acad. Sci. U.S.A. 101, 27.
Mitchell, J.E., Zheng, D., Busby, S.J., Minchin, S.D., 2003. Identification and analysis of 'extended −10' promoters in *Escherichia coli*. Nucleic Acids Res. 31, 4689–4695.
Moak, M., Molineux, I.J., 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol. Microbiol. 51, 1169–1183.
Nacesco, N., Constantinesco, S.P., Petrovici, A., 1969. Aspects electrono-optiques du phage convertisseur *Proteus vulgaris* 121. Arch. Roum. Pathol. Exp. Microbiol. 28, 838–848.
Nechaev, S., Geiduschek, E.P., 2008. Dissection of the bacteriophage T4 late promoter complex. J. Mol. Biol. 379, 402–413.
Norberg, S., Stanton, C., Ross, R.P., Hill, C., Fitzgerald, G.F., Cotter, P.D., 2012. *Cronobacter* spp. in powdered infant formula. J. Food Prot. 75, 607–620.
Petrov, V.M., Ratnayaka, S., Nolan, J.M., Miller, E.S., Karam, J.D., 2010. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. Virol. J. 7, 292.
Ray, P., Das, A., Gautam, V., Jain, N., Narang, A., Sharma, M., 2007. *Enterobacter sakazakii* in infants: novel phenomenon in India. Indian J. Med. Microbiol. 25, 408–410.
Sandaa, R.A., Clokie, M., Mann, N.H., 2008. Photosynthetic genes in viral populations with a large genomic size range from Norwegian coastal waters. FEMS Microbiol. Ecol. 63, 2–11.
Serwer, P., Hayes, S.J., Thomas, J.A., Hardies, S.C., 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. Virol. J. 4, 21.
Serwer, P., Hayes, S.J., Zaman, S., Lieman, K., Rolando, M., Hardies, S.C., 2004. Improved isolation of undersampled bacteriophages: finding of distant terminase genes. Virology 329, 412–424.
Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Nat. Acad. Sci. U.S.A. 93, 14440–14445.
Simoliunas, E., Kaliniene, L., Truncaite, L., Klausa, V., Zajanckauskaite, A., Meskys, R., 2012. Genome of *Klebsiella* sp.-infecting bacteriophage vB_KleM_RaK2. J. Virol. 86, 5406–5412.
Simoliunas, E., Kaliniene, L., Truncaite, L., Zajanckauskaite, A., Staniulis, J., Kaupinis, A., Ger, M., Valius, M., Meskys, R., 2013. *Klebsiella* phage vB_KleM-RaK2—a giant singleton virus of the family *Myoviridae*. PLoS One 8, e60717.
Soding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248.
Sonnhammer, E.L., von, H.G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol., 175–182 (ISMB 6).
Spricigo, D.A., Bardina, C., Cortes, P., Llagostera, M., 2013. Use of a bacteriophage cocktail to control *Salmonella* in food and the food industry. Int. J. Food Microbiol. 165, 169–174.
Stothard, P., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21, 537–539.
Summer, E.J., Berry, J., Tran, T.A., Niu, L., Struck, D.K., Young, R., 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. J. Mol. Biol. 373, 1098–1112.
Tippmann, H.F., 2004. Analysis for free: comparing programs for sequence analysis. Brief. Bioinform. 5, 82–87.

Turner, D., Reynolds, D., Seto, D., Mahadevan, P., 2013. CoreGenes3.5: a webserver for the determination of core genes from sets of viral and small bacterial genomes. BMC Res. Notes 6 (140), 140–146, http://dx.doi.org/10.1186/1756-0500-6-140.

Van Twest, R., Kropinski, A.M., 2009. Bacteriophage enrichment from water and soil. Methods Mol. Biol. 501, 15–21.

Wu, W., Thomas, J.A., Cheng, N., Black, L.W., Steven, A.C., 2012. Bubblegrams reveal the inner body of bacteriophage ϕKZ. Science 335, 182.

Xu, M., Struck, D.K., Deaton, J., Wang, I.N., Young, R., 2004. A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proc. Nat. Acad. Sci. U.S.A. 101, 6415–6420.

Yan, Q.Q., Condell, O., Power, K., Butler, F., Tall, B.D., Fanning, S., 2012. *Cronobacter* species (formerly known as *Enterobacter sakazakii*) in powdered infant formula: a review of our current understanding of the biology of this bacterium. J. Appl. Microbiol. 113, 1–15.

Zuker, M., Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.