Virology 426 (2012) 93–99

Contents lists available at SciVerse ScienceDirect

Virology

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

Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB_EcoP_G7C

Eugene Kulikov [a], Andrew M. Kropinski [b,c], Alla Golomidova [a], Erika Lingohr [b,c], Vadim Govorun [d], Marina Serebryakova [d], Nikolai Prokhorov [a], Maria Letarova [a], Anatolij Manykin [e], Alexandra Strotskaya [a], Andrey Letarov [a,*]

[a] Winogradsky Institute of Microbiology, Russian Academy of Science, pr. 60-letiya Oktyabrya 7, Building 2. 117 312, Moscow, Russia
[b] Public Health Agency of Canada, Laboratory for Foodborne Diseases, 110 Stone Road West, Guelph, ON N1G 3W4, Canada
[c] Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[d] Institute of Physico-Chemical Medicine, Federal Medico-Biological Agency, ul. M. Pirogovskaya 1a, 113435, Moscow, Russia
[e] Ivanovsky Institute of Virology, Russian Academy of Medical Sciences, ul. Gamaleya 16, 123098, Moscow, Russia

ARTICLE INFO

Article history:
Received 5 September 2011
Returned to author for revision
3 November 2011
Accepted 26 January 2012
Available online 15 February 2012

Keywords:
Escherichia coli
Podoviridae
N4
Lytic phage
Complete genome sequence
Sequence analysis
Proteomics
Tail fiber
N4-like coliphage
Enterobacteria bacteriophage

ABSTRACT

Lytic coliphage vB_EcoP_G7C and several other highly related isolates were obtained repeatedly from the samples of horse feces held in the same stable thus representing a component of the normal indigenous intestinal communities in this population of animals. The genome of G7C consists of 71,759 bp with terminal repeats of about 1160 bp, yielding approximately 73 kbp packed DNA size. Seventy-eight potential open reading frames, most of them unique to N4-like viruses, were identified and annotated. The overall layout of functional gene groups was close to that of the original N4 phage, with some important changes in late gene area including new tail fiber proteins containing hydrolytic domains. Structural proteome analysis confirmed all the predicted subunits of the viral particle. Unlike N4 itself, phage G7C did not exhibit a lysis-inhibited phenotype.

The ecology of intestinal coliphages in mammals appears to be significantly different between animal species (reviewed in (Letarov and Kulikov, 2009)). The communities of intestinal coliphages in horses are highly individual in different subjects, highly dynamic and, in contrast to humans (Reyes et al., 2010) and to some other mammal species (Letarov and Kulikov, 2009), is comprised mainly of virulent phages (Golomidova et al., 2007; Kulikov et al., 2007). Some bacteriophages in this system persist for months or years, while the others are only transitory components. Detailed molecular characterization of the viruses representing stable components of the intestinal ecosystem provides the necessary background for investigation of the mechanisms that stabilizes their long-term co-existence with the hosts, the prevalent mechanisms for *in situ* microevolution and of the molecular basis underlying other ecologically important phenomena (Letarov and Kulikov, 2009; Letarov et al., 2010). We present here an analysis of the genome and physiology of the horse-derived coliphage G7C, related to bacteriophage N4 isolated by Schito et al. from a sewer in Genoa, (Schito et al., 1966). Phage N4 possesses several very unusual features including the presence of the 3,500 amino acid rifampicin-resistant DNA-dependent virion RNA polymerase (vRNAP) which is injected together with the first 500 bp of the genomic DNA upon infection (Kazmierczak and Rothman-Denes, 2006). Once in the cytoplasm, the vRNAP initiates transcription from early promoters at the left terminal region of the genome. This causes the internalization of the rest of genomic DNA. The N4 vRNAP is the largest known protein delivered by a prokaryotic protein transport system.

For many years, N4 was a database orphan; but recently five new members of this genus have been characterized: *Sulfitobacter* phage EE36Φ1, *Silicibacter* phage DSS3Φ2 (Zhao et al., 2009) and, *Pseudomonas* phages LIT1, LUZ7, and PEV2 (Ceyssens et al., 2010). The genomes of N4-related Enterobacter phage EcP1 (Genbank ID: HQ641380) (Yu et al., 2008) and *Erwinia* phage vB_EamP-S6 (Genbank ID: HQ728266) have also appeared recently.

* Corresponding author.
E-mail address: letarov@gmail.com (A. Letarov).

0042-6822/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2012.01.027

In this manuscript we characterized a novel bacteriophage vB_EcoP_G7C, infecting *E. coli* strain 4 s, isolated from the same sample of horse feces. Our results indicate that it is a new member of the "N4-like viruses" genus.

Results and Discussion

Isolation of phage G7C and its host

Escherichia coli strain 4 s and its phage G7C were isolated in 2006 from the same sample of the horse feces in the course of our previous work aimed at investigation of equine intestinal coliphages ecology (Golomidova et al., 2007). Strain 4 s was later identified as *E. coli* by Biotyper peptide mass-spectroscopy, biochemical profiling and 16S rDNA sequencing. Bacteriophage G7C was initially characterized by TEM (Fig. 1a) and random genome clones sequencing and was identified as a putative N4-like coliphage due to virion morphology and the presence of the genes encoding for the polypeptides homologous to N4 proteins including a fragment of the vRNAP gene. When G7C was isolated, no other N4-related phages had been reported.

We ultimately isolated five additional N4-related phage isolates, related but distinguishable from G7C by genomic DNA RFLP profiling. These isolates were obtained from the same animal that gave G7C, and from samples from another horse held in the same stable in 2010. These results indicate that G7C-related phages are able to persist as indigenous component of the horse intestinal microbial system for a long time (Clokie et al., 2011). The isolation from the same environment of highly related but different phages, some of which have altered host recognition regions (see below) indicate that they can undergo substantial molecular evolution *in situ*, in contrast to human intestinal phages that seems to remain quite conserved genetically over the time (Reyes et al., 2010).

Morphology of phage G7C and its plaques

Bacteriophage G7C forms clear plaques 1–2 mm in diameter on the lawn of *E. coli* 4 s with or without 1 – 1.5 mm turbid halo. The infectivity of phage is insensitive to chloroform treatment. By TEM study, the virions resembled that of N4, about 70 nm in diameter with a small non-contractile tubular tail, about 25 nm long. In G7C phage, the tail carries a ring-like structure with appendages folded onto the capsid wall (Fig. 1a) and also a collar-like structure carrying another set of appendages similar to that visible on N4 particle (Choi et al., 2008).

Host-range properties of phage G7C

G7C does not infect most laboratory *E. coli* strains, and the only strain supporting its growth was *E. coli* strain 4 s isolated from horse feces. Moreover, nearly 50 *E. coli* strains isolated from the same specimen of feces and from another animals held in the same stable did not support growth of N4. This narrow host range appears to be a typical trait for members of the N4 genus, because N4-like phages of *Pseudomonas* and roseobacteria were also able to reproduce only on the bacterial strain used for their original isolation (Ceyssens et al., 2010; Zhao et al., 2009).

Bacteriophage G7C growth

The adsorption curve of bacteriophage G7C on *E. coli* 4 s cells at 37 °C (Fig. 2a) indicates extremely rapid kinetics of phage binding to the cells, and the cells had to be diluted in some experiments down to 10^7 CFU ml^{-1} to resolve the initial stage of adsorption. The adsorption constant K calculated from these experiments was 2×10^{-8} ml^{-1} in 1–2 min period, and the remaining phage absorbed at much slower rate (Fig. 2a).

A one-step growth curve of phage G7C (Fig. 2b) indicates a latent period of 40–42 min with a burst size of 500–1000. This behavior differs strikingly from the lysis inhibition observed with N4-infected cells, in which full release of N4 progeny is achieved only days after cell death (Schito, 1974), although both phages encode highly homologous SAR-type lysin and holin proteins (Summer et al., 2007). Phage N4 progeny accumulate inside distended infected cells in the shape of polar paracrystalline bodies with a burst size as high as 3000 phages per cell. Light microscopy studies using Nomarski differential interference contrast did not reveal comparable intracellular bodies in infected 4 s *E. coli* cells (data not shown).

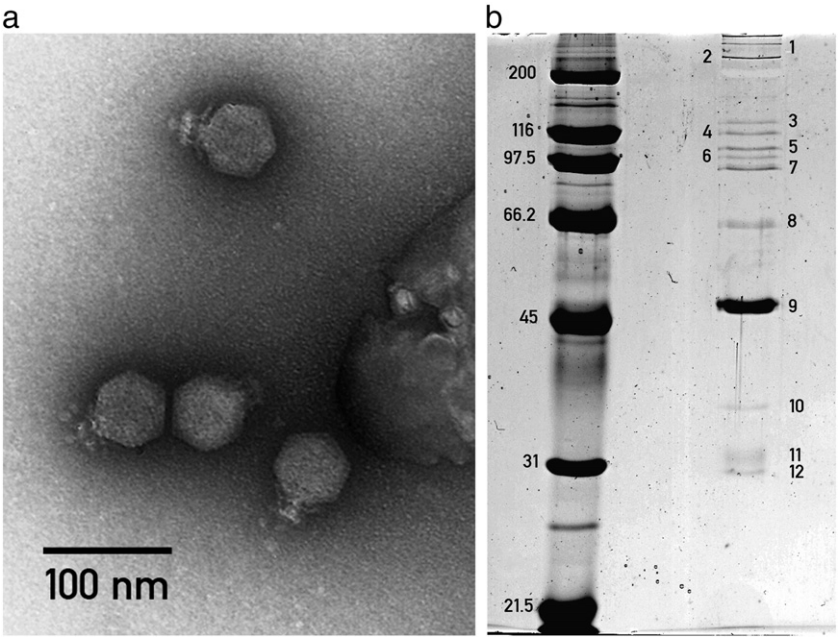

Fig. 1. a. Electron microphotography of G7C phage particles; magnification, x25000; staining, uranyl acetate. b. A silver-stained SDS-PAGE of G7C virus particles. The proteins identified in the bands (data combined from 2 experiments) were: 1 – non-identified HMW protein, 2- gp50 vRNAP, 3 - gp63.1 tail fiber, 4 - gp66 tail fiber, 5 - gp59 portal, 6 - gp56 major coat protein (possible dimer), gp7, gp63.1 tail fiber, 8 – gp51, 9- gp56 major coat protein, 10 - gp 54, 11 - gp17 capsid decorating protein, gp67 putative tail protein, 12 - cellular OmpA protein.

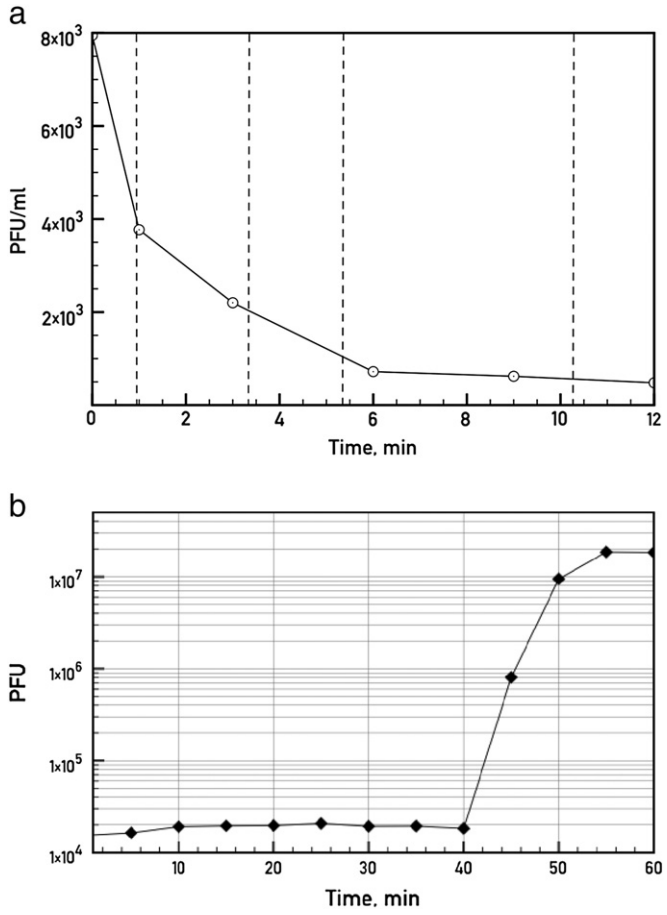

Fig. 2. a. The adsorption curve of bacteriophage G7C to 4 s E. coli cells. The concentration of bacterial cells was $3x10^7$ CFU/ml. The vertical dotted lines separate sequential half-adsorption periods. b. Single burst growth curve of bacteriophage G7C on 4 s E. coli cells.

The N4-like phages of *Pseudomonas* also do not show any lysis inhibition, producing 100–150 phage per cell in 25–30 minutes after infection (Ceyssens et al., 2010), while roseobacteria phages begin to exit from the cell only 3–4 hours after infection in larger (1000–1500) numbers (Zhao et al., 2009), thus resembling the original N4 phage.

G7C genome

The genomic DNA of G7C was submitted to pyrosequencing, resulting in ~50-fold genome coverage. The unique sequence of the G7C genome was 71,759 bp long. An overall G + C content is 42.0%. This is close to the mol% G + C content and length of N4 - 41.3% and 70,153 bp, respectively. The length was determined by the sequencing outrun (Fig. S1). using the gel-purified restriction fragments, containing the genomic termini as templates. The determined length of the TRs in bacteriophage G7C genome appears to be ~1,157 bp, thus yielding the size of virion encapsidated DNA about 72,920 bp. The left end of the G7C genome has a good nucleotide identity with the same end of N4 genome, while the right genomic boundary is fuzzy indicating that the terminal cut may occur in different positions in distinct packaging events.

Prediction of coding regions

78 ORFs were identified, ranging in length from 114 bp (ORF32.1) to 11,247 bp (ORF50, virion RNA polymerase), . Among predicted ORFs, 17 ORFs (22%) code for products with little or no homology to proteins from the NCBI databases. The N4 phage genome was used as a template for naming the ORFs identified in G7C. G7C ORFs with attributable functions are shown in Table 1. Except for four HNH – homing endonucleases, ubiquitous in phage and bacterial genomes, no non-phage ORFs were detected in phage G7C. The G7C genome contains 4.7% of non-coding sequence; there is little or no space between most genes. A 2 kb genomic stretch devoid of any ORFs is present in both N4 and G7C downstream of gene 47 (or gene 47.2 in G7C). Phage N4 contains 4 predicted tRNA genes in this region but no tRNAs were predicted in G7C using ARAGORN (Laslett and Canback, 2004).

Overall genome organization

Like N4, G7C early and middle genes have one orientation, while late genes are all transcribed in the opposite direction (Fig. 3). The sequence corresponding to the extreme left early promoter PE1 (Glucksmann-Kuis et al., 1996) (5'-cacgttgctccgcaacttatgg-3') was identified at the left end of G7C genome. Promoter P3 (5'-tgggctgttccacagcttacg-3') was found immediately upstream of the ORF 1.2. Homologues of many N4 early genes are absent in G7C and are replaced by non-orthologous sequences, which makes it difficult to determine the exact boundaries between early and middle clusters.

We were unable to unequivocally predict middle and late promoters of bacteriophage G7C presumably due to lack of information on their consensus sequences.

At least three potential rho-independent transcription terminators discovered using MFOLD (Zuker, 2003) are located between ORF47.2 and ORF50. A large number of predicted stem-and-loop structures flanking important genes were also found, most of them being potential rho-independent terminators of transcription that could influence transcriptional activity in the affected regions (Fig. 3). These structures are found in all three transcriptional modules of the G7C genome, but they apparently do not prevent transcription *in vivo* – for example, gene 63.1 whose protein product is found in viral particle, is flanked by a pair of 16 bp nearly identical palindromic DNA repeats (sense strand: 5'-ccccgccgaagcggggtt(a/t)(t/c)tt-3'), complete with poly-T tail. Exactly the same sequences for the terminators was found using WebGeSTer (http://pallab.serc.iisc.ernet.in/gester/) in *Brucella melitensis*, the causative agent for Malta fever (Paulsen et al., 2002). The presence of such structures in G7C genome suggests a mechanism for transcription antitermination potentially existing in this phage. Such exact DNA repeats could also serve as tags for site-specific recombination, potentially mobilizing the flanked gene.

Functional identification of predicted ORFs

Homologs of all N4 essential genes except for the adsorption apparatus are found in G7C (Table 1). However, many of N4 genes, especially those located in the left and central parts of the genome, are missing in G7C (including genes 3–5, 7, 8, 10, 13, 18–23, 28, 48, 49, 64, 65), being replaced by 23 new ORFs having no reliable similarity to N4 or other phage proteins. ORF27.1 codes for a small (87 aa) protein, which has significant sequence similarity to the immunity to superinfection protein of T4-related coliphage JS98 as well as to some other phage immunity membrane proteins.

DNA metabolism and DNA packaging genes

The G7C genome encodes a number of enzymes involved in DNA metabolism; most of which are homologous to the corresponding N4 proteins (Table 1). However, the G7C genome also lacks a N4 gp8 homolog, that is a 8.2 kDa protein directly involved in host DNA replication shutdown upon N4 infection in *E. coli* selectively binding a delta subunit of *E. coli* DNA polymerase III and inhibiting its function and clamp loading. Gp8 is not essential for phage N4 growth under the laboratory conditions but contributes significantly to the phage yield

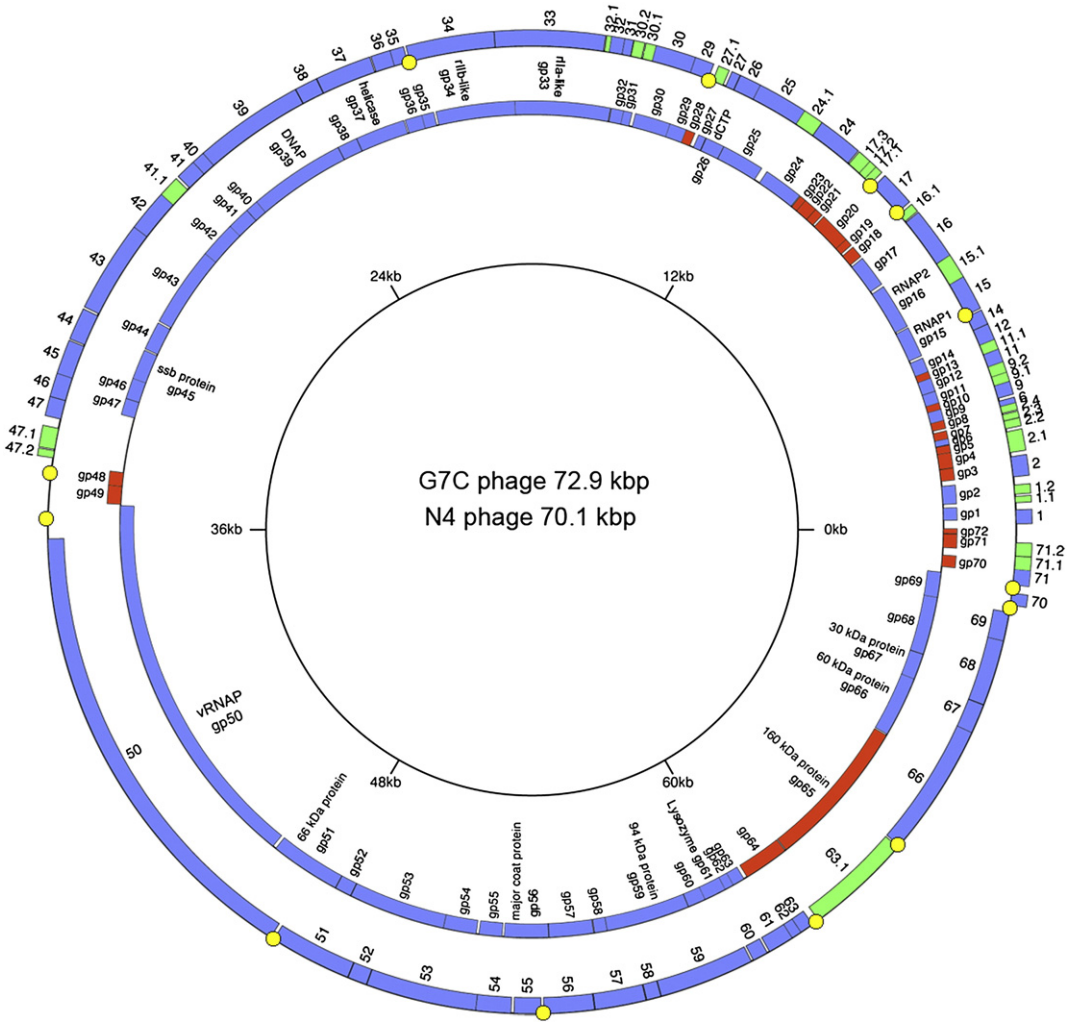

Fig. 3. A comparative circularized genomic map of G7C phage (outer circle) and N4 phage (inner circle). ORFs conserved between both phages are colored in blue, new ORFs present in G7C are green, and N4 ORFs deleted in G7C are red. The map is constructed using GenomeVx online tool and CoreGenes data. The yellow dots indicate the position of stem-loop structures with potential transcription terminator activity in the genome of G7C.

(Yano and Rothman-Denes, 2011). One of the potential explanations for effective growth of G7C without gp8 might includes the non-orthologous functional replacement of this gene by a product of another G7C ORF. It is interesting that gp8 is not conserved in any of the currently known N4-like phages (Ceyssens et al., 2010; Zhao et al., 2009).

Another component of the set of DNA metabolism enzymes encoded by G7C is comprised by four homing endonucleases (ORFs 15.1, 24.1, 41.1, 47.1), about 500 bp long each, existing in one copy per genome, as it is usual for homing endonucleases (Burt and Koufopanou, 2004). Roseobacteria N4-related phages DSS3ΦF2 and EE36ΦF1 both encode a putative HNH homing endonuclease homologous to gp22 in N4 (but missing in G7C genome), and EE36ΦF1 encodes two more endonucleases (ORFs 8 and 51).

The large terminase subunit in G7C (gp68) is highly divergent from other phage sequences and was identified only using HHpred (Soding et al., 2005). Its closest relative in the protein structural database is T4 phage gp17. We were not able to identify the small terminase subunit in G7C genome. The most probable candidate for that role is gp69, which has the appropriate size and genomic location and is not present in the virion.

The most remarkable protein that make N4 bacteriophage family unique among all other phages is gp50, the virion RNA polymerase (Falco et al., 1980). The direct benefits of such a huge protein transporter machinery is not clear, but the level of conservation of this feature across this phage genus allows us to say that giant virion-encapsidated RNA polymerase presumably plays an important role in the propagation strategy of N4-like phages. The sequence of bacteriophage G7C gp50 is highly similar to its N4 homolog (67.7% amino acid identity), but there is a region close to the N-terminus that is not homologous. It makes possible to design PCR primers matching conserved regions of this gene, but amplifying the divergent sequence thus allowing both identification and differentiation of N4-related isolates. This strategy was used in this work for identification of new G7C-related phages mentioned above. All the amino acid motifs and particular catalytic residues, described by Kazmierczak and co-authors (Kazmierczak et al., 2002) for N4 VRNAP are present in gp50 of G7C, as well as in the VRNAPs of all other N4-like phages. This suggests that the mechanism of action for this enzyme is conserved and typical for all members of N4 genus.

Other structural proteins

The structural genes loci are well conserved between N4 and G7C phages (except for the adsorption module – see below). Proteomic analysis of the G7C structural proteins was undertaken to aid the genome annotation. G7C phage was purified by CsCl density gradient centrifugation, and the virion proteins were analyzed by 1D denaturing gel electrophoresis and identified by MALDI-TOF analysis of tryptic digestion products (Fig. 1b). Nine predicted phage proteins were identified (gps17, 50, 51, 54, 56, 59, 63.1, 66,

67 – see Table 1 for their functions). All structural virion proteins with known functions were reliably identified, including virion RNA polymerase and both potential tail fiber proteins. Gp17, a capsid decorating protein with detectable Ig-like domain, is also present in G7C genome. The high degree of conservation of gp17 sequences between G7C and N4 is interesting, since such decorating proteins containing Ig-like domains often show a high degree of sequence divergence even in closely related phages ((Fraser et al., 2007) and our unpublished observations).

The functions of three proteins found in G7C structural proteome (gps 51, 54 and 67) are unknown. Presumably these proteins are involved in tail construction, but among them only gp67 is conserved among all the known N4 related phages. Another unusual finding was made – one of the well-pronounced protein bands was identified as *E. coli* OmpA protein, an abundant outer membrane protein in *E. coli* (Henning et al., 1978). The fact that this protein is co-purified with the phage in CsCl gradient suggests that it could act as a co-receptor for G7C. This hypothesis is consistent with the high rate of the phage G7C adsorption to cells since OmpA is very highly represented on the *E. coli* surface - up to 10^5 copies per cell (Morona et al., 1985).

Adsorption module

The most drastic difference between G7C and N4 is an organization of its adsorption module (Fig. 3). It consists of ORFs 63.1 and 66, both encoding fiber-like proteins, flanked by the conserved ORFs 63 and 67, sharing good homology with the corresponding N4 genes, while ORFs 64 and 65 are missing in G7C. A tail sheath protein gp65 (160 kDa) in N4 is essential for adsorption to its host receptor, Nfr (McPartland and Rothman-Denes, 2009). It is not clear which protein functionally replaces gp65 in G7C.

ORF63.1, 2556 bp long, possesses N-terminal domain homologous to several phage tailspike proteins with the best match for N-terminal 150 amino acid residues of Det7 phage tailspike. Det7 is a myovirus which gained, in course of its evolution, an adhesin from P22-like podovirus (Walter et al., 2008). It is also a nearly exactly match of the first 120 amino acid residues in ORF210 of CBA120, T-even related phage that carries podovirus-style tailspikes (Genbank ID: AEM91896).

In the central part of gp63.1 the conserved domain of SGNH hydrolase (potential lipase or esterase) was identified. We hypothesized that this domain might be involved in receptor modification facilitating the transition between reversible and irreversible phage attachment to the host cell. To test this hypothesis, we cloned gene 63.1 into pET-23a expression vector (Novagen) and expressed the recombinant gp63.1 protein under T7 promotor. The expression level was reaching about 40% of the total cell protein (data not shown). The protein was soluble that indicates that it does not require any phage-encoded chaperon for its proper folding.

We demonstrated that gp 63.1 binds effectively to the surface of *E. coli* 4 s cells (the host strain used for G7C isolation), but not to any *E. coli* strain resistant to this phage. Gp63.1 binds specifically to 4 s cell surface within 1 min, but leaves it within 15 min. The binding-release cycle can be repeated if new aliquot of the cells is added (Fig. 4). These results indicate that gp63.1 protein specifically recognizes and irreversibly modifies some structures of the host cell surface, probably hydrolyzing a part of cell exopolysaccharide. G7C gp63.1 shares no homology with N4 proteins. It is not clear at present which protein domain mediates its attachment to the phage particle.

The G7C gp66 protein shares sequence similarity with its N4 homolog only in its N-terminal region, presumably responsible for the attachment of the tail fiber to the phage particle. The rest of the protein is mosaic with patches of homology to various phage tail spike proteins and some stretches of weak homology to bacterial proteins. It also contains a putative hydrolytic domain in its C-terminal part (putative polygalactouranase, as identified by BLAST search for conserved protein domains). In contrast to gp 63.1, the expression of

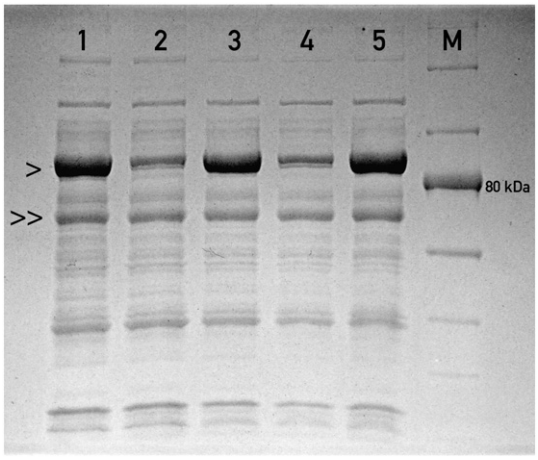

Fig. 4. Reversible binding of gp63.1 to 4 s *E. coli* host cells. Lanes: 1 - lysate of induced BL21(DE3) E. coli cells containing about 40% of recombinant gp63.1 protein (band marked by > symbol); 2 - same lysate 1 min after addition of 10^9 4 s *E. coli*; 3- same at 15 min; 4 - same mixture 1 min after addition of fresh 10^9 4 s *E. coli* aliquot; 5- same 15 min later; M - molecular weight marker (Sibenzyme). Only supernatants were loaded on the gel. Bovine serum albumin (BSA) was added to as a negative adsorption control, (band marked by >> symbol).

gp66 in a similar system was poor (about 10% of total cell protein), but yielding a soluble protein. Induced culture expressing gp66 stops to grow, indicating the toxicity of gp66 to the cells No direct interaction of the recombinant gp66 with the surface of host cells was detected (data not shown). While gp66 has an N-terminus much like N4 gp66 tail fiber, it could potentially mediate interactions of gp63.1 with the phage particle. A central part of gp66 (amino acid range 220–470) has 37% identity with CBA120 (Genbank ID: AEM91899.1) tailspike. This region might be an interface for gp63.1 interaction with gp66 via its N-terminal CBA120/Det7-related region.

A very similar organization of the adsorption module is observed in the recently published N4-related phage EcP1 infecting *Enterobacter* species (Yu et al., 2008), which has two potential tail fiber genes partially homologous to abovementioned G7C proteins. However, in the tail fibers of EcP1 no putative enzymatic domains could be detected. In *Pseudomonas* phages PEV2/LIT1 and LUZ7, the whole tail gene locus is completely replaced by a new cluster of genes, embedded in DNA replication module in reverse orientation. These genes are most probably acquired from a prophage integrated in *Pseudomonas* host genome, and the cluster is rather conservative across all three *Pseudomonas* N4-like phages. A modular replacement is also present in coliphage ALT isolated by us. This virus is otherwise closely related to G7C, in which the whole g63.1-g66 module is swapped for a sequence block containing two ORFs featuring only very distant and patchy homology to corresponding G7C proteins. It looks like in N4 phage family the adsorption and structural module changes completely as a result of the adaptation to the novel hosts, in contrast to some other phage families, like T-even related phages, where extensive modular swapping of protein domains of tail fibers is the prevalent strategy of the host adaptation (Comeau et al., 2007; Letarov et al., 2005).

Cell lysis genes

G7C genome codes for an endolysin, ORF61, with a very high level of identity to the classical N4 bacteriophage 208aa ORF61 protein which has an N-acetyl-muramidase activity *in vitro* (Stojković and Rothman-Denes, 2007). However, the sequence of gp61 N-terminal secretion–arrest-release (SAR) domain , tagging it for the transport to the periplasmic space, and mediating its arrest in inactive membrane-bound state followed by subsequent release upon the phage holin-induced shut down of transmembrane proton potential (Park et al., 2007; Xu et al., 2005), is quite different between two phages. In G7C genome

ORF61 is preceded by two small ORFs, named ORF60 and ORF60', sharing some homology (36% a.a. identity for gp60) with similarly organized ORFs 60 and 60' of the phage N4. The organization of this genome locus is reminiscent of Rz/RzI locus in λ bacteriophage, placed immediately downstream of the endolysin gene position (Summer et al., 2007). Putative spanin function was attributed to these proteins on the basis of the genetic organization and PSI-BLAST search (R. Young, personal communication). Gp62 of G7C (as well as its homolog in N4) is identified by PSI-BLAST as a putative holin. This identification is supported by the genome position of this ORF right upstream to the identified endolysin gene. No possible antiholin protein was identified in the G7C genome. So phage G7C possesses an almost complete (except for identified antiholin) set of lysis proteins that ensures effective release of the phage progeny at the end of the phage live cycle.

The question why N4 infection cycle is lysis-inhibited while that of G7C is not is still open. It is noteworthy that the roseophages DSS3ΦF2 and EE36ΦF1 are both lacking the endolysin gene, at least direct homologs of N4 gp61. Their life cycle includes a long (about 2 h) eclipse period, followed by a gradual and slow (10-15 h) release of new phage particles (Zhao et al., 2009).

Conclusions

The complete genome sequence of bacteriophage G7C is closely related to that of the phage N4. However, despite the high level of the main proteins sequence conservation, the accumulated divergence in genome sequence and organization is sufficient to change significantly the physiology of G7C infection in comparison to N4. Complete and non-homologous replacement of the adsorption module seems to be a characteristic feature in evolution of this group of viruses, since the similar polymorphism is observed in other N4-related phage genomes (Ceyssens et al., 2010; Zhao et al., 2009). In case of G7C this replacement provides the phage with very highly efficient host recognition machinery, allowing for an extremely fast adsorption to the host cells. The minor changes in lysis genes compared to G7C lead to a completely different growth style of G7C phage. The new data also suggest substantial differences in transcription regulation between coliphages G7C and N4. Further analysis of molecular mechanisms connecting observed genetic differences to these physiological alterations might contribute to understanding of the evolution of the phage genome even more than the analysis of more distantly related bacteriophages. The bacteriophage G7C is also a valuable template model for investigation of the molecular background of ecologically meaningful phenotypic differences identifiable in genetically related isolates obtained from the same environment. According to the novel universal system of bacteriophage naming (Kropinski et al., 2009b), the full suggested name of the bacteriophage G7C will be vB_EcoP_G7C.

Materials and Methods

Phage propagation and purification

Bacteriophage G7C was propagated on *E. coli* 4 s in LB medium containing 10 g tryptone, 5 g yeast extract, 5 g NaCl and distilled water up to 1 L. High titer lysates preparation, CsCl step gradient purification and phage DNA isolation were performed by conventional methods as described by Sambrook (Sambrook and Russell, 2001).

Determination of the phage adsorption curve

The appropriate dilution of the bacteriophage stock was added to 5 ml of the log phase broth host culture up to final concentration of about 10^6 PFU/ml. An aliquot of the culture was diluted and plated out before adding phage to estimate the exact CFU count. The phage-cell mixture was incubated on shaker at $+37^0$ and 100 RPM. 20 mkl aliquots were taken at different time points and diluted hundredfold in ice-cold medium to slow down the adsorption, then placed on ice. The samples centrifuged at 13000 rpm in tabletop centrifuge for 30 s to pellet the cells, and, 50 mkl of supernatant were plated to count free phage PFUs. After 10 min of incubation 250 mkl aliquots were plated instead of 50 mkl. 3 min sample was also plated without centrifugation to check the initial phage titer ($t=0$). The adsorption constant was calculated as $K=-\ln(P/P_0)/Bt$, where P – the free phage PFU count at the end of the time interval used, P_0 – the initial phage PFU count, B – the host cell count (CFU^*ml^{-1}) and t – time, min.

DNA manipulations and sequencing strategy

The G7C genomic DNA was isolated by phenol-chloroform extraction as described in (Sambrook and Russell, 2001), dissolved in MilliQ deionized water and submitted for 454 pyrosequencing to the McGill University and Genome Quebec Innovation Centre (Montreal, QC, Canada). The 28 contigs resulting from pyrosequencing and automated assembly were joined using direct Sanger sequencing of the phage genomic DNA. The sequences of the primers used are available from the authors upon request. Sanger dideoxynucleotide sequencing was performed using an Applied Biosystems Avant 3100 capillary DNA sequencer with BigDye reagents.

ORFs prediction and genome annotation

Final sequence assembly was performed with SeqMan software (DNASTAR Inc., Madison, WI). ORFs were predicted using Kodon (Applied Maths, Austin, TX) with the calls verified using GeneQuest (DNASTAR Inc., Madison, WI). Genome comparisons were done using MAUVE and CoreGenes according to usual routine (Kropinski et al., 2009a; Zafar et al., 2002). Ribosome binding sites (RBS) were verified using TIGR RBSfinder (http://www.tigr.org/software/). Transcriptional terminators were predicted using the Terminator tool from GCG package, the Wisconsin Sequence Analysis Package version 10, using algorithm of Brendel and Trifonov (Brendel and Trifonov, 1984), and MFOLD (Zuker, 2003) (http://frontend.bioinfo. rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi). Protein functions and similarity were identified using BLAST (http://www.ncbi.nlm.nih.gov) and HHpred (http://toolkit.tuebingen.mpg.de) (Soding et al., 2005) searches.

Protein cloning and expression

Genes 63.1 and 66 were amplified by PCR using primers N4.gp63.1.F1 ACATGCTAGCAAAACAAAGAGGTTTATATGAAC (NheI) and N4.gp63.1.R1 ATACCTCGAGAGCGGGGCATAATTTTAGT (XhoI) for gene 63.1 and the primers N4.gp66.F1 ATATTCTAGAGAACAACGAGGATGGCGATG (XbaI) and N4.gp66.R1 ATATCTCGAGCTAAGAAACCCCGCCGAAGC (XhoI) for gene 66, containing introduced restriction sites for cloning. The PCR products were digested with appropriate restriction enzymes (NheI/XhoI for gene 63.1 and XbaI/XhoI for gene 66) and ligated with pET-23a vector digested by the enzymes XbaI/XhoI (NheI and XbaI produce compatible cohesive ends) using T4 DNA ligase (Sibenzym, Russia). The ligated mixtures was used for transformation of E. coli DH5alpha cells, the clones containing desired constructions were identified by PCR screening, and the plasmids were purified and controlled by sequencing of the inserts. Finally the plasmids p63.1 and p66 were obtained. These plasmids contain genes 63.1 and 66 respectively with their natural ribosome binding sites under the control of T7 promoter.

The culture of E. coli BL21(DE3) T7 expression strain (NEB) containing appropriate plasmid was induced by addition of IPTG up to 1 mM concentration. The expression was carried out at 37 °C for 3 h. The cells were harvested by centrifugation at 3000 g for 15 min, resuspended in 20 mM Tris–HCl pH 8.0 and disintegrated by ultrasonication. The protease inhibitor PMSF (Serva) was added up to 1 mM concentration. The crude lysates were cleared by centrifugation at 12 000 g for 15 min and stored at +4.°C.

Analysis of the cell binding activity of gp 63.1 protein

The cells from mid-log phase culture of G7C host E. coli 4 s strain were harvested by centrifugation and resuspended in 20 mM Tris–HCl pH 8.0. A 50 mkl portion of this suspension containing approximately 10^9 cells was added to 200mkl of crude lysate containing about 0.5 mg/ml of recombinant protein and mixture was incubated at room temperature. Bovine serum albumin (BSA, fraction V, Serva) was added to the samples to a concentration 0.3 mg/ml as an internal control. The 50 mkl aliquots were taken at desired time points, centrifuged immediately in the microcentrifuge at 13 000 rpm for 30 s to pellet the cells, and the 20 mkl samples of the supernatants were collected from the upper part of the liquid. These samples were mixed with Laemmli loading buffer and analysed by conventional SDS-PAGE as it described in (Sambrook and Russell, 2001). In some experiments, another aliquot of 10^9 4 s cells was added after 15 min of incubation and the cycle of binding – dissociation was followed as it described above.

Nucleotide sequence accession number

The complete nucleotide sequence of phage G7C was deposited in the GenBank database (Genbank ID: HQ259105).

Analysis and MALDI-TOF profiling of G7C virion proteins

One-dimensional SDS gel electrophoresis was performed on 9-16% gradient polyacrylamide gels (Laemmli, 1970). The gels were silver-stained to detect the protein bands (PlusOne Silver Staining Kit, GE Healthcare, USA). The latter were excised, destained, dried and submitted for trypsinolysis. The desalted tryptic digest was mixed with trifluoroacetic acid and 2,5-dihydroxybenzoic acid as carrier, then it was separated on a tandem MALDI-time-of-flight mass spectrometer Ultraflex II BRUKER (Bruker Daltonics, Billerica, MA), equipped with a Nd:YAG UV laser in positive ion mode using reflectrone. The resulting peptide fingerprint was analyzed with Mascot software (Matrix Science Ltd., London, UK) against the database of conceptually translated G7C ORFs with needed correction for precision, potential methionine oxidation by air, and possible modifications of cysteine residues by acrylamide.

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2012.01.027.

Acknowledgements

We are grateful to Professor Ry Young from Texas A and M University for the help in lysis genes identification, and to Dr. Konstantin Miroshnikov for fruitful discussions. This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (A.M.K.). The work of Moscow group is supported by the RFBR grant No 06-04-48651-a and by the grants of Russian Ministry for Education and Science.

References

Brendel, V., Trifonov, E.N., 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. 12, 4411–4427.
Burt, A., Koufopanou, V., 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr. Opin. Genet. Dev. 14, 609–615.
Ceyssens, P.-J., Brabban, A., Rogge, L., Lewis, M.S., Pickard, D., Goulding, D., Dougan, G., Noben, J.-P., Kropinski, A., Kutter, E., Lavigne, R., 2010. Molecular and physiological analysis of three Pseudomonas aeruginosa phages belonging to the "N4-like viruses". Virology 405, 26–30.
Choi, K.H., McPartland, J., Kaganman, I., Bowman, V.D., Rothman-Denes, L.B., Rossmann, M.G., 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J. Mol. Biol. 378, 726–736.
Clokie, M.R., Millard, A.D., Letarov, A.V., Heaphy, S., 2011. Phages in nature. Bacteriophage 1, 31–45.
Comeau, A.M., Bertrand, C., Letarov, A., Tétart, F., Krisch, H.M., 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. Virology 362, 384–396.
Falco, S.C., Zehring, W., Rothman-Denes, L.B., 1980. DNA-dependent RNA polymerase from bacteriophage N4 virions. Purification and characterization. J. Biol. Chem. 255, 4339–4347.
Fraser, J.S., Maxwell, K.L., Davidson, A.R., 2007. Immunoglobulin-like domains on bacteriophage: weapons of modest damage? Curr. Opin. Microbiol. 10, 382–387.
Glucksmann-Kuis, M.A., Dai, X., Markiewicz, P., Rothman-Denes, L.B., 1996. E. coli SSB activates N4 virion RNA polymerase promoters by stabilizing a DNA hairpin required for promoter recognition. Cell 84, 147–154.
Golomidova, A., Kulikov, E., Isaeva, A., Manykin, A., Letarov, A., 2007. The diversity of coliphages and coliforms in horse feces reveals a complex pattern of ecological interactions. Appl. Environ. Microbiol. 73, 5975–5981.
Henning, U., Sonntag, I., Hindennach, I., 1978. Mutants (ompA) affecting a major outer membrane protein of Escherichia coli K12. Eur. J. Biochem. 92, 491–498.
Kazmierczak, K.M., Rothman-Denes, L.B., 2006. Bacteriophage N4. In: Calendar, R. (Ed.), The Bacteriophages. Oxford University Press, Oxford New York, pp. 302–314.
Kazmierczak, K.M., Davydova, E.K., Mustaev, A.A., Rothman-Denes, L.B., 2002. The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. EMBO J. 21, 5815–5823.
Kropinski, A.M., Borodovsky, M., Carver, T.J., Cerdeño-Tárraga, A.M., Darling, A., Lomsadze, A., Mahadevan, P., Stothard, P., Seto, D., Van Domselaar, G., Wishart, D.S., 2009a. In silico identification of genes in bacteriophage DNA. Methods Mol. Biol. 502, 57–89.
Kropinski, A.M., Prangishvili, D., Lavigne, R., 2009b. Position paper: the creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. Environ. Microbiol. 11, 2775–2777.
Kulikov, E.E., Isaeva, A.S., Rotkina, A.S., Manykin, A.A., Letarov, A.V., 2007. Diversity and dynamics of bacteriophages in horse feces. Mikrobiologiia 76, 271–278.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11–16.
Letarov, A., Kulikov, E., 2009. The bacteriophages in human- and animal body-associated microbial communities. J. Appl. Microbiol. 107, 1–13.
Letarov, A., Manival, X., Desplats, C., Krisch, H.M., 2005. gpwac of the T4-type bacteriophages: structure, function, and evolution of a segmented coiled-coil protein that controls viral infectivity. J. Bacteriol. 187, 1055–1066.
Letarov, A.V., Golomidova, A.K., Tarasyan, K.K., 2010. Ecological basis for rational phage therapy. Acta naturae 2, 66–79.
McPartland, J., Rothman-Denes, L.B., 2009. The tail sheath of bacteriophage N4 interacts with the Escherichia coli receptor. J. Bacteriol. 191, 525–532.
Morona, R., Krämer, C., Henning, U., 1985. Bacteriophage receptor area of outer membrane protein OmpA of Escherichia coli K-12. J. Bacteriol. 164, 539–543.
Park, T., Struck, D.K., Dankenbring, C.A., Young, R., 2007. The pinholin of lambdoid phage 21: control of lysis by membrane depolarization. J. Bacteriol. 189, 9135–9139.
Paulsen, I.T., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J., Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M., Fraser, C.M., 2002. The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc. Natl. Acad. Sci. U. S. A. 99, 13148–13153.
Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., Gordon, J.I., 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466, 334–338.
Sambrook, J., Russell, D.W., 2001. Molecular cloning : a laboratory manual, 3 rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
Schito, G.C., 1974. Development of coliphage N4: ultrastructural studies. J. Virol. 13, 186–196.
Schito, G.C., Rialdi, G., Pesce, A., 1966. Biophysical properties of N4 coliphage. Biochim. Biophys. Acta 129, 482–490.
Soding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, 244–248.
Stojković, E.A., Rothman-Denes, L.B., 2007. Coliphage N4 N-acetylmuramidase defines a new family of murein hydrolases. J. Mol. Biol. 366, 406–419.
Summer, E.J., Berry, J., Tran, T.A.T., 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.
Walter, M., Fiedler, C., Grassl, R., Biebl, M., Rachel, R., Hermo-Parrado, X.L., Llamas-Saiz, A.L., Seckler, R., Miller, S., van Raaij, M.J., 2008. Structure of the receptor-binding protein of bacteriophage det7: a podoviral tail spike in a myovirus. J. Virol. 82, 2265–2273.
Xu, M., Arulandu, A., Struck, D.K., Swanson, S., Sacchettini, J.C., Young, R., 2005. Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science 307, 113–117.
Yano, S.T., Rothman-Denes, L.B., 2011. A phage-encoded inhibitor of Escherichia coli DNA replication targets the DNA polymerase clamp loader. Mol. Microbiol. 79, 1325–1338.
Yu, L., Wang, J., Zhao, J.-J., Wen, Z.-B., Yang, W.-H., Li, N., Lu, J.-C., Li, J.-S., 2008. Isolation of Escherichia coli phages and their biological characteristics. Bull. Acad. Mil. Med. Sci. 5, 1–10.
Zafar, N., Mazumder, R., Seto, D., 2002. CoreGenes: a computational tool for identifying and cataloging "core" genes in a set of small genomes. BMC Bioinformatics 3, 12.
Zhao, Y., Wang, K., Jiao, N., Chen, F., 2009. Genome sequences of two novel phages infecting marine roseobacters. Environ. Microbiol. 11, 2055–2064.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.