View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CorpusUL

Evolution of *Lactococcus lactis* Phages Within a Cheese Factory

Geneviève M. Rousseau[1] and Sylvain Moineau[1,]*

[1]Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie

Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire

Félix d'Hérelle Reference Center for Bacterial Viruses

Université Laval, Québec, Canada, G1V 0A6

* Corresponding author. Mailing address: Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire, Université Laval, Quebec City, Quebec, Canada, G1V 0A6

Phone: 418-656-3712. Fax: 418-656-2861. E-mail: Sylvain.Moineau@bcm.ulaval.ca

ABSTRACT

We have sequenced the double-stranded DNA genomes of six lactococcal phages (SL4, CB13, CB14, CB19, CB20, GR7) from the 936 group that were isolated over an 8-year period from whey samples obtained from a Canadian cheese factory. These six phages infected the same two industrial *L. lactis* strains out of 30 tested. The CB14 and GR7 genomes were found to be 100% identical even though they were isolated 14 months apart, indicating that a phage can survive in a cheese plant for over a year. The other four genomes were related but notably different. The length of the genomes varied from 28,144 to 32,182 bp and coded for 51 to 55 ORFs. All five genomes possessed a 3' overhang *cos* site, 11 nucleotides long. Several structural proteins were also identified by nano-HPLC-MS/MS, confirming bioinformatics analyses. Comparative analyses suggested that the most recently isolated phages (CB19 and CB20) were derived, in part, from older phage isolates (CB13, CB14/GR7). The organization of the five distinct genomes was similar to the previously sequenced lactococcal phage genomes of the 936 group, and from these sequences, a core-genome was determined for lactococcal phages of the 936 group.

INTRODUCTION

The manufacture of cheeses requires the inoculation of carefully selected bacterial cultures, known as starter cultures, at concentrations of at least 10^7 live bacteria per ml of heat-treated milk. The purpose of this process is to control the fermentation and to obtain high-quality fermented products (29). Starter cultures are a combination of Lactic Acid Bacteria (LAB), of which one of the most important species is *Lactococcus lactis*. *L. lactis* is a low GC Gram-positive bacterium used to metabolize lactose into lactic acid during the production of several cheese varieties. Because large amounts of lactococcal cells are cultivated each day in large-scale fermentation vats and because these cells are susceptible to bacteriophage infection, it is not surprising that most cheese factories have experienced problems with phage contamination (13). Even a single phage infecting a starter strain is enough to begin a chain reaction that can eventually inhibit bacterial growth, cause production delays, taste and texture variations and even complete fermentation failures (1, 29).

Phage infections are unpredictable in food fermentations. Their presence and persistence in a dairy factory can be explained in many ways. First, raw milk can introduce new phages into an industrial plant (25). Madera *et al*. (22) also reported that newly isolated lactococcal phages were more resistant to pasteurization. Whey, a liquid by-product of cheese manufacturing, is another reservoir that can spread phages in a factory environment (25). Airborne phage dissemination may also be important since concentrations of up to 10^6 PFU/m^3 have been observed close to a functional whey separation tank (32).

For decades the dairy industry has been working to curtail the propagation of virulent phages using a variety of practical strategies including, among others, sanitation, optimized factory design, air filtration units, rotation of bacterial strains and the use of phage resistance

systems (13). Yet, new virulent phages emerge on a regular basis. Indeed, large-scale industrial milk fermentation processes can be slowed down by virulent phages of the *Caudovirales* order. Members of three lactococcal phage groups are mostly found in dairy plants, namely, 936, c2 and P335. The 936-like phages are by far the most predominant worldwide (3, 18, 22, 27).

Phages of the 936 group have a double-stranded DNA genome and possess a long non-contractile tail connected to a capsid with icosahedral symmetry, characteristic of the *Siphoviridae* family. Currently, six complete phage genomes of the lactococcal 936 group are available in public databases, including sk1 (6), bIL170 (10), jj50, 712, P008 (23), and bIBB29 (16). Their comparative analysis revealed a conserved gene organization despite being isolated from different countries. Most of the differences have been observed in the early genes module, where insertions, deletions, and point mutations likely occurred (16, 23). Moreover, it is assumed these phages can also exchange DNA through recombination with other bacterial viruses present in the same ecosystem.

Because new members of this lactococcal phage group are regularly isolated, a better understanding of their evolution is warranted to better control them. A cheese factory is a particular man-made niche where rapidly growing bacterial strains encounter ubiquitous phages. Such active environments provide ample opportunities for phage evolution, especially to dodge phage resistance mechanisms that may be present in host cells. Nonetheless, the evolutionary dynamics that shape the diversity of lactococcal phage populations are still not well understood.

In this study, we analyzed the genome and structural proteome of six 936-group phages (SL4, CB13, CB14, CB19, CB20, GR7) that infected the same *L. lactis* strains and were isolated over an eight-year period from a cheese factory.

MATERIALS AND METHODS

Isolation of virulent lactococcal phages. Whey samples were obtained from a single cheese plant using defined starter cultures. The phages present in the samples were amplified using *L. lactis* subspecies *cremoris* SMQ-404 as an indicator strain, as described elsewhere (3, 28), and were propagated according to the method of Jarvis (17). The species of the lactococcal phage isolates was obtained using multiplex PCR (19).

Bacterial strains and phages. *L. lactis* subsp. *cremoris* SMQ-404 and SMQ-438 were grown at 30°C in M17 (Difco) supplemented with 0.5% lactose. For propagation of phages SL4, CB13, CB14, CB19, CB20, and GR7, host cells were incubated until the $OD_{600\ nm}$ reached 0.1. Phages and $CaCl_2$ were added to the growing culture at a final concentration of 10^6 phages/ml and 10 mM, respectively. The phage-infected culture was incubated until complete bacterial lysis was obtained, then filtered through a 0.45-µm-syringe-filter (Fisher Scientific). To obtain highly concentrated phage preparations, lysates were mixed with polyethylene glycol (34) and purified on a discontinuous CsCl gradient followed by a continuous one-step CsCl gradient. The first centrifugation was performed at 35,000 rpm for 3 h in a Beckman SW41 Ti rotor. The second ultracentrifugation was performed using a Beckman NVT65 rotor at 60,000 rpm for 18 h (7, 14).

DNA sequencing and sequence analysis. The DNA of virulent lactococcal phages SL4, CB13, CB14, CB19, CB20 and GR7 was isolated from high-titer lysates using a Lambda Maxi kit (Qiagen) and the modifications for the 936 group suggested by Deveau *et al.*, (11). To confirm the identity of the isolated DNA, digestions with EcoRI and EcoRV endonucleases were performed. Restriction profiles were then matched with their corresponding patterns in our

database. Restriction endonucleases (Roche Diagnostics) were used as recommended by the manufacturer. Primers previously designed to sequence the genome of lactococcal phage P008 (23) were used for direct sequencing of the conserved regions in the genomes of the six new phages. Primers were then designed to complete the sequencing of both strands using an ABI Prism 3100 apparatus from the genomic platform at the Centre Hospitalier de l'Université Laval. The *cos* site was determined as reported elsewhere (23). Sequence assembly was performed with Staden software (http//:staden.sourceforge.net/) and BioEdit 7.0.5.3 software (http//:www.mbio.ncsu.edu/BioEdit/bioedit.html) was used for alignment editing. Open reading frames (ORFs) were predicted using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The assignment of ORFs was performed using criteria that were described previously (23). The translated ORF products were compared with known protein sequences using BLASTP (2). The estimated molecular masses and pIs were obtained using the tool Compute pI/Mw (ca.expasy.org).

Structural proteins identification. Approximately 8 µg of phage proteins (5) were added to a 12% SDS polyacrylamide gel (1.5 mm thick). The protein samples were mixed with 4X loading buffer (0.250 M Tris-HCl pH 6.8, 40% (w/v) glycerol, 8% (w/v) SDS, 20% (v/v) β-mercaptoethanol, 0.1% (w/v) bromophenol blue) and boiled for 5 minutes before loading. Proteins were detected using Coomassie blue staining. For protein identification, bands were cut from the gel, digested with trypsin and identified by nano-HPLC-MS/MS at the Génome Québec Innovation Centre at McGill University.

DNA-DNA hybridizations. The Southern blot DNA analysis of phage and bacterial genomes was done using the method described by Deveau *et al.* (12). Phage genomic DNAs were used as probes. Bacterial DNA was isolated by the method described by Fortier and Moineau (15) with the following modifications: cell pellets were suspended in 1 ml saline (0.85% NaCl), transferred to an Eppendorf tube and centrifuged for 10 minutes at full speed in a microcentrifuge. The pellets were then suspended in 200 µl of a 25% sucrose solution containing 30 mg/ml lysozyme, and incubated at 37°C for 15 minutes. Finally, 400 µl of 3% sodium dodecyl sulfate was added and the preparation incubated at room temperature with agitation for 7 minutes.

Electron microscopy. A 1.5 ml sample from a high-titer phage lysate was centrifuged (23,500 x g) at 4°C for 1 h. The supernatant was discarded and the residual 100 µl washed twice with 1.5 ml ammonium acetate (0.1 M). A final volume of 100 µl was saved for observation. Grids were prepared by adding 7.5 µl of washed phages to a copper Formvar-carbon-coated grid (200 mesh, Pelco International). Uranyl acetate (7.5 µl of a 2% solution) was immediately added and mixed by pipetting up and down. The liquid was removed after 30 to 60 seconds by touching the edge of the grid with blotting paper (12). Phages were observed at 80 kV with a JEOL 1230 transmission electron microscope. Dimensions given were the mean of 10 specimens. Phage dimensions were measured using purified phage preparation diluted 1:10, rather than washed phage lysate.

Nucleotide accession numbers. The genome sequences were submitted to the GenBank database under the following accession numbers: FJ848881 (phage SL4), FJ848882 (phage CB13), FJ848883 (phage CB14), FJ848884 (phage CB19), and FJ848885 (phage CB20).

RESULTS AND DISCUSSION

Isolation of the phages. The six virulent phages analyzed in this study (SL4, CB13, CB14, CB19, CB20, and GR7) were isolated from different cheddar cheese whey samples from the same Canadian cheese factory over an 8-year period (1996 to 2003) and at the beginning of the study were the only one infecting the industrial *L. lactis* strain SMQ-404 (Table 1). Host range analysis indicated that all six phages infected the same two *L. lactis* subsp. *cremoris* strains (SMQ-404 and SMQ-438) out of 30 *L. lactis* subsp. *cremoris* strains tested. So far, *L. lactis* SMQ-404 and SMQ-438 are sensitive to only 936-like phages. The EcoRV and EcoRI restriction patterns of isolated phage DNAs were compared and found to be related (data not shown). Multiplex PCR analysis confirmed that they all belonged to the 936 species (12, 19). The genomes were sequenced to shed more light on their origins.

Analysis of phage genomes. Overall, the genome length of the six lactococcal phages ranged from 28,144 to 32,182 bp (Table 1). Their GC content was 34.9%, which is similar to other lactococcal phages (Table 1) and *L. lactis* strains (35.7%) for which the complete genomes are available (4, 24, 40). These phage genomes possessed 51 to 55 open reading frames and, overall, shared 82.1% nucleic acid identity. Interestingly, comparative genomic analyses revealed that phages CB14 and GR7 contained identical genomes (100%), even though the two phages were isolated 14 months apart, indicating that a phage can be stable in a cheese plant for a long period of time. Phages CB19 and CB20 were 99.5% identical (Table 2) and were isolated from the same whey sample. The main sequence differences were found in *orf*1 and *orf*3 (Table 3), which likely encode terminase subunits. The genome of phage SL4 possessed four additional *orfs* when compared to the other four distinct phage genomes. Two of the *orfs* (*orf*10 and *orf*24) coded for

proteins containing a putative HNH endonuclease motif, indicating possible roles in replication, recombination, maturation or encapsidation of phage DNA (10). It has been suggested that the homing endonuclease is also involved in the gene diversity observed in the early expressed genomic region of the 936-like phages (23). Phage CB13 also has four distinct genes including one (*orf*31) that may code for an endonuclease.

Table 1 summarizes the main characteristics of the six phages as well as their comparisons with the six other 936-like phages for which complete sequences are available in GenBank. These latter six phages were isolated outside North America and their host strains (laboratory strains *L. lactis* subsp. *lactis* IL1403 and *L. lactis* subsp. *cremoris* MG1363) were different from the one used here (industrial strain *L. lactis* subsp. *cremoris* SMQ-404). The overall genome organization, however, was highly conserved in all 11 virulent lactococcal phage sequences of the 936 group (Fig. 1). The DNA packaging module was always found next to the morphogenesis module, followed by the lysis genes and the replication cluster. Comparative genomic analyses indicated that these 11 genomes possessed 62.2% to 99.5% identity at the nucleotide level, even though the isolates came from seven different countries (Table 2).

All 11 phages possessed cohesive genomic extremities (*cos*-type). We identified three unique *cos* sites, although only one or two nucleotides separated these three genomic extremities (Table 1). Phages SL4, CB19, CB20, sk1, bIL170, 712, jj50, and bIBB29 possessed the same *cos* site, even though some of them were isolated 30 years apart. Phages CB13 and CB14/GR7 shared the same *cos* site sequence while phage P008 had distinct genomic extremities (Table 1).

Origin of the phage genes. Phage SL4 was isolated in 1996 while CB13, CB14, CB19 and CB20 emerged between July 2003 and September 2003. Phage CB20 was the last distinct phage isolated from our whey samples that infected the same *L. lactis* strains. Thus, we hypothesized

194 that CB20 may be derived from the other four phages. Pairwise comparisons suggested that *orf*4
195 to *orf*51 of phage CB20 may have derived from the genome of CB19 while *orf*1, *orf*2 and *orf*3
196 originated from the genome of phage CB13 (Fig. 2).

197 Phage CB19 was the second last phage isolated from the cheese whey samples. The
198 sequences *orf*17*, orf*19*, orf*20*,* as well as *orf*22 to *orf*44 of phage CB19 may have originated from
199 phage CB14, while *orf*4 and *orf*14 may have originated from phage CB13 (Fig. 2). However, the
200 origins of *orf*5 to *orf*21 (except for *orf*14, *orf*17, *orf*19 and *orf*20) and *orf*45 to *orf*51 of phage
201 CB19 remain elusive. Nonetheless, these data suggest that most recently isolated lactococcal
202 phages are derived, in part, from older phage isolates. The isolation of phages CB13, CB14/GR7,
203 CB19, and CB20 in a short period of time in 2003 may explain why we could identify these
204 shared modules. Phage SL4 was isolated 8 years previously and it was less related to these
205 phages.

206 *L. lactis* host strains are known to carry prophages (4, 24, 40) that participate in strain
207 diversity and the evolution of virulent phages (20), although these prophages belong to the P335
208 group and are genetically distinct from the 936-like phages analyzed here (12). Nonetheless, to
209 verify if the unknown phage DNA came from the host strain, DNA-DNA hybridization
210 experiments were performed against total DNA from *L. lactis* strains SMQ-404 and SMQ-438
211 using total phage DNA as the probe. No hybridization signals were observed (data not shown),
212 indicating that these two host strains did not contribute to the genomic diversity of the 936-like
213 phages. Taken together, these data suggest that some phage modules were swapped from other
214 virulent lactococcal phages already in the cheese plants but not analyzed in this study.

215

216 **Structural proteins.** The structural protein profiles of the five lactococcal phages were analyzed
217 by SDS-PAGE (Fig. 3A). The five phages had similar protein profiles, including a single major

structural protein, along with several minor proteins, confirming their relatedness (Fig. 3A). Phage SL4 had the most divergent profile. A total of twelve proteins were identified by nano-HPLC-MS/MS, including seven from phage CB19, four from SL4 and one from CB13 (Fig. 3A, D). More structural proteins were selected from phage CB19 as it appeared to contain the most common proteins (based on molecular weight) among the five phages. All proteins identified by nano-HPLC-MS/MS could be linked to a phage gene. The molecular masses calculated from the SDS-PAGE were in agreement with the theoretical masses for 9 of the 12 phage proteins (Fig. 3D), including the portal protein (bands #3 and #10), the receptor binding protein (RBP, bands #4 and #12) and structural proteins of unknown function (bands #2, # 9, #11).

The putative tape measure protein (TMP, protein #6) of phage CB19 had an observed mass of 198.1 kDa, which was almost double the theoretical value estimated by bioinformatic analyses (105.6 kDa). Formation of a dimer could possibly explain this difference. Similarly, protein band #7 of phage CB19 had an estimated molecular mass of 143.5 kDa based on its migration on a SDS-PAGE gel, while its theoretical mass was calculated to be 71.3 kDa. The protein was identified as a putative neck passage structure (NPS) protein. Interestingly, the same protein was identified in bands #5 and # 8 as a monomer. Structural and functional analyses have shown that some phage structural proteins are found as dimers (8, 21, 30).

As indicated above, the structural protein profiles of phage SL4 differed from the other four phages, but proteins with similar functions also differed in size. For example, the TMP (band #1) of phage SL4 was much smaller than its counterpart (band #6) in phage CB19, suggesting that it may be processed in SL4. Similarly, the portal protein was also smaller in SL4 (band #3) compared to phage CB19 (band #10). On the other hand, the RBP was a similar size in both phages. This is not surprising since the RBPs of these phages have a conserved architecture of three protomers related by a 3-fold axis, and each protomer comprises three domains: the N-

terminus shoulders, the interlaced β-prism linker and the C-terminus head (33, 36, 38). Moreover, both phages infect the same *L. lactis* strains.

Neck passage structure protein. Bioinformatic and structural protein profile analyses suggested that phages CB13, CB14, CB19, and CB20 harbored a NPS protein. A gene (L12) coding for such a protein was previously observed in the 936-like lactococcal phage bIL41 (9). The NPS of phage bIL141 shared 69% identity with ORF12 of phages CB19 and CB20. It was also previously shown that the 936-like phages sk1 and jj50 do not carry a gene coding for a NPS protein (Fig. 1) (23). Similarly, we could not identify a gene coding for this protein in phage SL4. Observation of the five lactococcal phages (SL4, CB13, CB14, CB19 and CB20) by electron microscopy identified a collar structure for all phages but SL4 (Fig. 3B, C), confirming the presence of NPS in four of the five lactococcal phage analyzed.

Crutz-Le Coq *et al.* (9) previously showed that NPS does not seem to play an important role during the assembly of phages and that it is not essential for lactococcal phages of the 936 group. They hypothesized, however, that NPS may be involved in host recognition (9). That SL4 does not possess a NPS structure and has the same host range as CB13, CB14, CB19 and CB20 suggests that this protein may have another function. Others have determined that NPS forms a collar-whisker complex but is nonessential for phage assembly, stability and host range of lactococcal phages of the P335 group (39).

Core-genome of 936-like phages. A core genome is defined as a set of genes invariably present and conserved in a group of isolates (37). According to Muzzi *et al.* (31), a gene is considered conserved when two proteins can be aligned with a minimum of 50% sequence conservation over

265 50% of the protein length. Using these definitions and the 11 genomes known for the lactococcal
266 936-like phages, a core-genome was determined for this group of phages. A total of 33 ORFs
267 were conserved (363 proteins out of 597 proteins analyzed, 60.8%) and are part of the core-
268 genome of the 936 group. ORFs of phage CB19 that are part of the core-genome of the 936 group
269 are highlighted in Table 3. Most of these ORFs are likely structural proteins. The most conserved
270 protein was ORF14 from phages SL4, CB13, CB14, CB19, CB20, and P008. This ORF also
271 corresponded to L14 in phage bIL170, ORF13 in phages 712 and bIBB29, and ORF12 in phages
272 sk1, jj50 and p2. It was recently proposed that the non-structural phage protein ORF12 of phage
273 p2 might act as a chaperone, maintaining the phage TMP in solution during the tail assembly of
274 lactococcal phages (35).

275 A comparative analysis was also performed with the 33 core deduced proteins of the 936-
276 like phages and deduced proteins found in other lactococcal phage groups. None of the 936 core
277 proteins gave a significant match to other lactococcal phage proteins, except the endolysin, which
278 might be also conserved in the lactococcal Q54 and P087 phage groups.

279 Based on bacterial studies (26), the core-genome includes all genes/proteins responsible
280 for the basic biological characteristics of a species as well as its major phenotypic traits. Clearly,
281 structural proteins represent major constituents of any given phage group, leading to a conserved
282 morphotype. It remains to be seen if this core genome will be upheld as more phage genomes of
283 the 936 group become available.

284

285 **Conclusions.** To our knowledge, this is the first report on the genomic characterization of North
286 American lactococcal phages of the predominant 936 group. The analysis of these Canadian
287 lactococcal phages has almost doubled the number of genomes available for 936-like phages.
288 Sequence comparisons provided valuable information about the evolutionary history of these

289 phages. One phage was found to persist a in cheese factory for over a year, indicating that the
290 industrial practice of removing a starter culture for a short period of time (weeks/months) is
291 unlikely to be effective in the long term. This study also showed that the genome architecture of
292 lactococcal phages of the 936 group is highly conserved and probably reflects an optimal
293 organization to rapidly multiply in a dairy environment. Such a conserved genetic structure likely
294 facilitated the functional exchange of genes or groups of genes (modules) between virulent phage
295 genomes in response to various host and/or environmental factors. Despite the conserved
296 structure, our analysis identified considerable genetic flux between phage genomes, particularly
297 in the early expressed region. Future studies aimed at understanding this natural genomic
298 variation will likely provide clues to improved control strategies for lactococcal phage
299 populations.
300
301 **Acknowledgments**
302
303 We thank Hélène Deveau and Julie Samson for helpful discussions, Denise Tremblay for
304 assistance with the Figures, Claudia Bergeron and Steve Labrie for phage isolation, and Barbara-
305 Ann Conway for editorial assistance. We are grateful to M. Lamoureux (Agropur) for providing
306 whey samples. This work was funded by a strategic grant from the Natural Sciences and
307 Engineering Research Council (NSERC) of Canada.

REFERENCES

1. **Accolas, J. P., C. Peigney, G. K. Y. Limsowtin, P.-J. Cluzel, and L. Séchaud.** 1994. Lutte contre les bactériophages dans l'industrie laitière, p. 473-492. *In* H. de Roissart and F. M. Luquet (ed.), Bactéries Lactiques. Lorica, Uriage, France.

2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-3402.

3. **Bissonnette, F., S. Labrie, H. Deveau, M. Lamoureux, and S. Moineau.** 2000. Characterization of mesophilic mixed starter cultures used for the manufacture of aged cheddar cheese. J. Dairy Sci. **83:**620-627.

4. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp *lactis* IL1403. Genome Res. **11:**731-753.

5. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248-54.

6. **Chandry, P. S., S. C. Moore, J. D. Boyce, B. E. Davidson, and A. J. Hillier.** 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. Mol. Microbiol. **26**:49-64.

7. **Chibani, S., I. Fliss, R.E. Simard, and S. Moineau.** 1998. Monoclonal antibodies raised against native major capsid proteins of lactococcal c2-like bacteriophages. Appl. Environ. Microbiol. **64:**4255-4259.

8. **Cingolani, G., D. Andrews, and S. Casjens.** 2006. Crystallogenesis of bacteriophage P22 tail accessory factor gp26 at acidic and neutral pH. Acta Crystallogr. Sect. F. Struct. Biol. Cryst Commun. **62:**477-482

9. **Crutz-Le Coq, A.-M., F. Cantele, S. Lanzavecchia, and S. Marco.** 2006. Insights into structural proteins of 936-type virulent lactococcal bacteriophages. Arch. Virology **151:**1039-1053.

10. **Crutz-Le Coq, A. M., B. Cesselin, J. Commissaire, and J. Anba.** 2002. Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. Microbiology **148**:985-1001.

11. **Deveau, H., M.R. Van Calsteren, and S. Moineau.** 2002. Effect of exopolysaccharides on phage-host interactions in *Lactococcus lactis*. Appl. Environ. Microbiol. **68:**4364-4369.

12. **Deveau, H., S. J. Labrie, M.-C. Chopin, and S. Moineau.** 2006. Biodiversity and classification of lactococcal phages. Appl. Environ. Microbiol. **72:**4338-4346.

13. **Emond, E., and S. Moineau.** 2007. Bacteriophages and food fermentations. *In:* Bacteriophage: Genetics and Molecular Biology. S. McGrath and D. van Sinderen (eds.). Horizon Scientific Press/Caister Academic Press. p. 93-124.

14. **Fortier, L.-C., A. Bransi, and S. Moineau.** 2006. Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of *Lactococcus lactis*. J. Bacteriol. **188:**6101-6114.

15. **Fortier, L.-C., and S. Moineau.** 2007. Morphological and genetic diversity of temperate phages in *Clostridium difficile*. App. Environ. Microbiol. **73:**7358-7366.

16. **Hejnowicz, M. S., M. Golebiewski, and J. Bardowski.** 2008. Analysis of the complete genome sequence of the lactococcal bacteriophage bIBB29. Int. J. Food Microbiol. doi:10.1016/j.ijfoodmicro.2008.06.010.

17. **Jarvis, A. W.** 1978. Serological studies of a host range mutant of a lactic streptococcal bacteriophage. Appl. Environ. Microbiol. **36:**785-789.

18. **Josephsen, J., N. Andersen, H. Behrndt, E. Brandsborg, G. Christiansen, S. Hansen, and E. W. Nielsen.** 1994. An ecological study of lytic bacteriophages in *Lactococcus lactis* subsp. *cremoris* isolated in a cheese plant over a five year period. Int. Dairy J. **4:**123-140.

19. **Labrie, S., and S. Moineau.** 2000. Multiplex PCR for detection and identification of lactococcal bacteriophages. Appl. Environ. Microbiol. **66:**987-994.

20. **Labrie, S.J., and S. Moineau.** 2007. Abortive infection mechanisms and prophage sequences significantly influence the genetic make-up of emerging lytic lactococcal phages. J. Bacteriol. **189:**1482-1487.

21. **Leiman, P.G., M.M. Shneider, V.A. Kostyuchenko, P.R. Chipman, V.V. Mesyanzhinov, and M.G. Rossmann.** Structure and location of gene product 8 in the bacteriophage T4 baseplate. 2003. J. Mol. Biol. **328:**821-833.

22. **Madera, C., C. Monjardin, and J. E. Suarez.** 2004. Milk contamination and resistance to processing conditions determine the fate of *Lactococcus lactis* bacteriophages in dairies. Appl. Environ. Microbiol. **70:**7365-7371.

23. **Mahony, J., H. Deveau, G. S. Mc, M. Ventura, C. Canchaya, S. Moineau, G. F. Fitzgerald, and D. van Sinderen.** 2006. Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712, and P008: evolutionary insights into the 936 phage species. FEMS Microbiol. Lett. **261:**253-261.

24. **Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M.**

Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills. 2006. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. U. S. A. **103:**15611-15616.

25. **McIntyre, K., H. A. Heap, G. P. Davey, and G. K. Y. Limsowtin.** 1991. The distribution of lactococcal bacteriophage in the environment of a cheese manufacturing plant. Int. Dairy Journal **1:**183-197.

26. **Medini, D., C. Donati, H. Tettelin, V. Masignani, and R. Rappuoli.** 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. **15:**589-594.

27. **Moineau, S., M. Borkaev, B. J. Holler, S. A. Walker, J. K. Kondo, E. R. Vedamuthu, and P. A. Vandenbergh.** 1996. Isolation and characterization of lactococcal phages from U.S. buttermilk plants. J. Dairy Sci. **79:**2104-2111.

28. **Moineau, S., J. Fortier, H.-W. Ackermann, and S. Pandian.** 1992. Characterization of lactococcal bacteriophages from Québec cheese plants. Can. J. Microbiol. **38:**875-882.

29. **Moineau, S., D. Tremblay, and S. Labrie.** 2002. Phages of lactic acid bacteria: from genomics to industrial applications. ASM News **68:**388-393.

30. **Morais, M.C., S. Kanamaru, M.O. Badasso, J.S. Koti, B.A. Owen, C.T. McMurray, D.L. Anderson, and M.G. Rossmann.** 2003. Bacteriophage phi29 scaffolding protein gp7 before and after prohead assembly. Nat. Struct. Biol. **10:**572-576.

31. **Muzzi, A., V. Masignani, and R. Rappuoli.** 2007. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug Discov. Today. **12:**429-39.

32. **Neve, H., A. Laborius, and K. J. Heller.** 2003. Testing of the applicability of battery-powered portable microbial air samplers for detection and enumeration of air-borne dairy *Lactococcus lactis* bacteriophages. Kieler Milchwirtschaftliche Forschungsberichte **55:**301-315.

33. **Ricagno, S., V. Campanacci, S. Bangy, S. Spinelli, D. Tremblay, S. Moineau, M. Tegoni, and C. Cambillau. 2006.** Crystal structure of the receptor-binding protein head domain from *Lactococcus lactis* phage bIL170. J. Virol. **80:**9331-9335.

34. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning, a laboratory manual, Third ed. Cold Spring Harbor Laboratory Press.

35. **Siponen, M., G. Sciara, M. Villion, S. Spinelli, J. Lichière, C. Cambillau, S. Moineau, and V. Campanacci.** 2009. Crystal structure of ORF12 from *Lactococcus lactis* phage p2 identifies a tape measure protein chaperone. J. Bacteriol. **191:**728-734.

36. **Spinelli, S., A. Desmyter, C. T. Verrips, H. J. de Haard, S. Moineau, and C. Cambillau.** 2006. Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol **13:**85-9.

449 37. **Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S.V.**
450 **Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M.**
451 **Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R.**
452 **Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H.**
453 **Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K.**
454 **Watkins, K. J. B. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi,**
455 **L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraser.**
456 2005. Genome analysis of multiple pathogenic isolates of *Streptoccoccus agalactiae*:
457 Implications for the microbial ''pan-genome''. Proc. Natl Acad. Sci. U S A. **102:**13950-13955.
458
459 38. **Tremblay, D. M., M. Tegoni, S. Spinelli, V. Campanacci, S. Blangy, C. Huyghe, A.**
460 **Desmyter, S. Labrie, S. Moineau, and C. Cambillau.** 2006. Receptor-binding protein of
461 *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding
462 site. J Bacteriol **188:**2400-10.
463
464 39. **Vegge, C.S., H. Neve, L. Brøndsted, K.J. Heller, and F.K. Vogensen. 2006.** Analysis of
465 the collar-whisker structure of temperate lactococcal bacteriophage TP901-1. Appl. Environ.
466 Microbiol. **72:**6815-6818.
467
468 40. **Wegmann, U., M. O'Connell-Motherwa, A. Zomer, G. Buist, C. Shearman, C.**
469 **Canchaya, M. Ventura, A. Goesmann, M. J. Gasson, O. P. Kuipers, D. van Sinderen, and J.**
470 **Kok.** 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis*
471 subsp. *cremoris* MG1363. J. Bacteriol. **189:**3256-3270.

Table 1. Characteristics of the lactococcal phages analyzed in this study.

Phage	Date of isolation	Country	Host strain	Genome				References
				Length (pb)	%G+C	#ORFs	*cos* site	
P008	1971	Germany	IL1403	28,538	34.7	58	5'-CACAAAGGATT-3'	23
bIL170	1973	France	IL1403	31,754	34.4	64	5'-CACAAAGGACT-3'	10
sk1	Before 1976	Australia	MG1363	28,451	34.5	55	5'-CACAAAGGACT-3'	6
jj50	1985	Denmark	MG1363	27,453	34.9	50	5'-CACAAAGGACT-3'	23
712	Before 1988	New-Zealand	MG1363	30,510	33.9	55	5'-CACAAAGGACT-3'	23
SL4	1996	Canada	SMQ-404	28,144	35.0	52	5'-CACAAAGGACT-3'	This study
CB13	2003	Canada	SMQ-404	32,182	34.7	55	5'-CACAGAGGACT-3'	This study
CB14	2003	Canada	SMQ-404	29,459	34.8	52	5'-CACAGAGGACT-3'	This study
CB19	2003	Canada	SMQ-404	28,643	35.2	51	5'-CACAAAGGACT-3'	This study
CB20	2003	Canada	SMQ-404	28,625	35.0	51	5'-CACAAAGGACT-3'	This study
GR7	2004	Canada	SMQ-404	29,459	34.8	52	5'-CACAGAGGACT-3'	This study
bIBB29	Before 2007	Poland	IL1403	29,305	34.7	54	5'-CACAAAGGACT-3'	16

472

Table 2. Nucleic acid identity (%) between each lactococcal phage genome of the 936 group.

	P008	bIL170	sk1	jj50	712	**SL4**	**CB13**	**CB14/ GR7**	**CB19**	**CB20**	bIBB29
P008	100										
bIL170	77.8	100									
sk1	74.2	69.8	100								
jj50	74.9	69.8	93.9	100							
712	71.7	64.9	75.1	73.4	100						
SL4	74.1	69.9	73.2	74.4	69.1	100					
CB13	67.2	72.3	67.1	67.5	62.2	72.1	100				
CB14/GR7	71.9	73.7	72.4	72.6	66.7	77.1	79.2	100			
CB19	72.6	75.4	70.7	71.9	66.2	80.9	75.3	90.3	100		
CB20	72.7	75.6	70.7	71.9	66.3	80.9	75.8	89.9	99.5	100	
bIBB29	77.8	75.7	70.6	71.2	68.3	74.1	67.7	68.1	68.6	68.8	100

Legends of Figures and Table 3

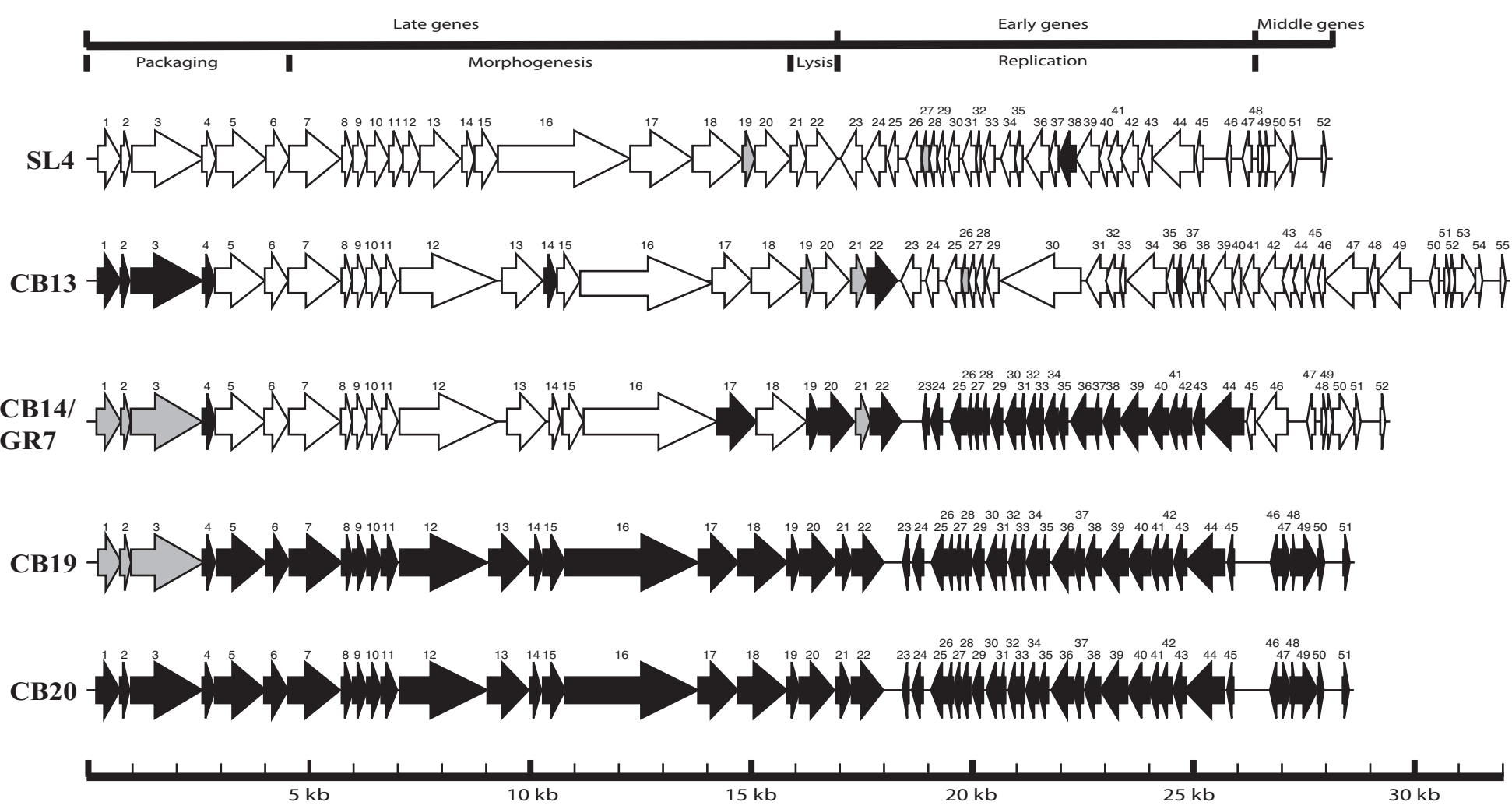
Figure 1. Schematic representation of the genomic organization of phages belonging to the 936 group. Each line represents a different phage genome and each arrow represents a putative protein. Each genome was compared only with the successive genome in this figure. ORFs of the same color represent those that share more than 80% amino acid identity. The percentages have been calculated for the smallest proteins. The white ORFs are unique. Grey shading connects genome regions conserved in all phages. Finally, arrows with thick outlines represent structural proteins observed on SDS-PAGE and identified by mass spectrometry or by N-terminal sequencing (for protein identification of phage bIL170, see references 9 and 33).

Figure 2. Schematic representation of the genomic organization of lactococcal phages SL4, CB13, CB14/GR7, CB19, and CB20. Each line represents a different phage genome and each arrow represents a putative protein. The ORFs in black or grey represent those that share 100% amino acid identity. Arrows in white indicate proteins that share less than 100% amino acid identity.

Figure 3. (A) Analysis of the structural proteins of phages SL4, CB13, CB14, CB19, and CB20 by SDS-PAGE. The M is the Broad Range Protein Marker (BioRad). The numbers represent the proteins identified by mass spectrometry. (B) Electron micrograph of phage CB19. (C) Electron micrograph of SL4. (D) Identification of the phage structural proteins by nano-HPLC coupled with tandem mass spectrometry.

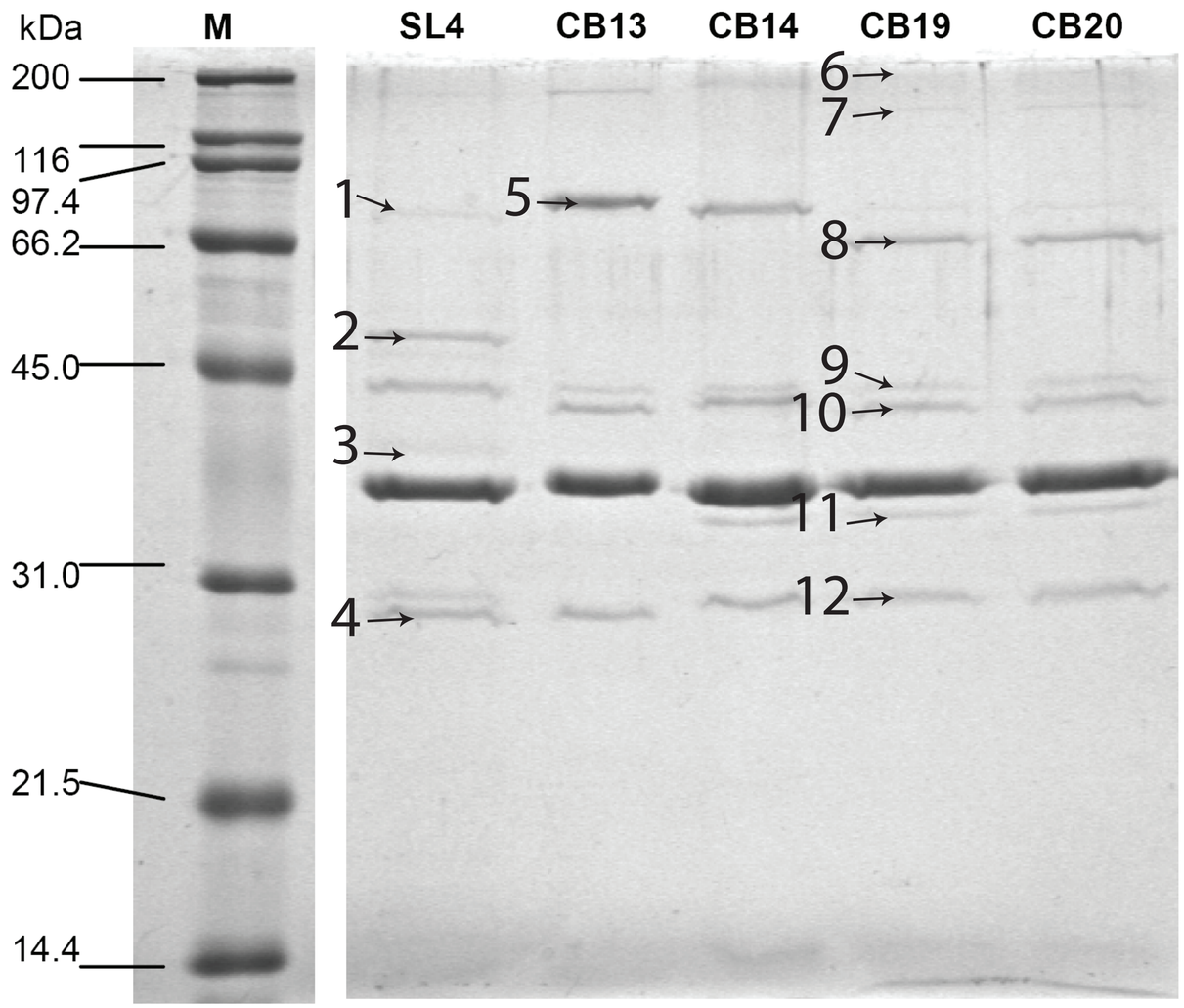
Table 3. Coordinates of phage CB19 ORFs and representative ORFs in the other 936-type phages. The percentages have been calculated for the smallest proteins. Bolded ORFs represent those of the CB19 genome that comprise part of the core-genome of the 936 group.

Strand	ORF	Start	Stop	Size (aa)	MW (kDa)	pI	SD sequence AGAAAGGAGGT	% aa identity SL4	CB13	CB14/GR7	CB20	P008	bIL170	sk1	jj50	712	bIBB29	Putative function
+	**1**	266	790	174	19.9	5.1	AGAAAGGATAAt**ATG**	ORF1, 98	ORF1, 98	ORF1, 100	ORF1, 98	ORF1, 96	I1, 97	ORF1, 97	ORF1, 97	ORF1, 97	ORF1, 95	Terminase small subunit
+	2	787	1017	76	9.0	4.0	GCACCAGAGGGatttga**ATG**	ORF2, 96	ORF2, 97	ORF2, 100	ORF2, 97	-	-	-	-	-	-	
+	**3**	1029	2651	540	63.0	6.0	AGAAAGGTAATga**ATG**	ORF3, 98	ORF3, 97	ORF3, 100	ORF3, 97	ORF2, 97	I2, 95	ORF2, 97	ORF2, 98	ORF2, 97	ORF2, 97	Terminase large subunit
+	**4**	2641	2925	94	11.2	9.1	AGAAATGGCGGtgtcag**ATG**	ORF4, 98	ORF4, 100	ORF4, 100	ORF4, 100	ORF3, 95	I3, 97	ORF3, 96	ORF3, 96	ORF3, 96	ORF3, 98	HNH endonuclease
+	**5**	2938	4074	378	43.3	5.0	AGAAAGGGGAAaaa**TTG**	ORF5, 96	ORF5, 97	ORF5, 95	ORF5, 100	ORF4, 94	I4, 94	ORF4, 92	ORF4, 92	ORF4, 92	ORF4, 96	Portal protein
+	**6**	4055	4591	178	19.9	4.6	AGAAAGGACGTaacaagcacag**ATG**	ORF6, 98	ORF6, 96	ORF6, 99	ORF6, 100	ORF5, 98	I5, 98	ORF5, 98	ORF5, 97	ORF5, 97	ORF5, 97	Prohead protease
+	**7**	4584	5765	393	43.7	5.5	ATTGAGGATATtaaaaagaaat**ATG**	ORF7, 93	ORF7, 98	ORF7, 96	ORF7, 100	ORF6, 95	I6, 95	ORF6, 95	ORF6, 95	ORF6, 96	ORF6, 94	Minor structural protein
+	**8**	5786	6049	87	10,0	6.2	AAAACGGAGGAagtaa**ATG**	ORF8, 98	ORF8, 96	ORF8, 96	ORF8, 100	ORF8, 98	I8, 98	ORF7, 98	ORF7, 96	ORF7, 97	ORF7, 96	
+	**9**	6049	6363	104	11.8	9.0	ATTATGGAGGTatttta**ATG**	ORF9, 95	ORF9, 94	ORF9, 92	ORF9, 100	ORF9, 94	I9, 90	ORF8, 92	ORF8, 96	ORF8, 93	ORF8, 95	
+	**10**	6353	6691	112	12.9	4.2	AGAGGGGGTCGtaagta**ATG**	ORF11, 94	ORF10, 96	ORF10, 92	ORF10, 100	ORF10, 91	I10, 93	ORF9, 96	ORF9, 91	ORF9, 91	ORF9, 93	
+	**11**	6682	7047	121	13.7	10.2	GTGCAGGTGGTcaacc**ATG**	ORF12, 97	ORF11, 95	ORF11, 93	ORF11, 100	ORF11, 96	I11, 94	ORF10, 98	ORF10, 95	ORF10, 95	ORF10, 95	
+	12	7104	9074	656	71.3	5.2	ACAATAATGGTatttttta**ATG**	-	ORF12, 33	ORF12, 40	ORF12, 100	ORF12, 63	I12, 85	-	-	ORF12, 53	-	Neck passage structure
																ORF16, 18		
+	**13**	9098	10003	301	32.6	4.9	AAAAAGGAAAAtaaaaa**ATG**	ORF13, 94	ORF13, 91	ORF13, 96	ORF13, 100	ORF13, 92	I13, 93	ORF11, 93	ORF11, 92	ORF11, 85	ORF11, 85	Major capsid protein
+	**14**	10041	10316	91	10.6	4.9	TAAAGGGATATaaaacaaa**ATG**	ORF14, 98	ORF14, 100	ORF14, 97	ORF14, 100	ORF14, 100	I14, 100	ORF12, 98	ORF12, 100	ORF13, 100	ORF13, 97	
+	**15**	10336	10848	170	19.9	4.8	TGAGAGGGCTGtga**ATG**	ORF15, 97	ORF15, 96	ORF15, 97	ORF15, 100	ORF15, 97	I15, 98	ORF13, 95	ORF13, 98	ORF14, 95	ORF14, 97	
+	**16**	10848	13838	996	105.6	9.1	AGAAAGGGTATgta**ATG**	ORF16, 97	ORF16, 86	ORF16, 85	ORF16, 100	ORF16, 75	I16, 76	ORF14, 82	ORF14, 81	ORF15, 74	ORF15, 77	Tape measure protein
+	**17**	13838	14734	298	34.4	5.5	ACTAGGGAGGGctta**ATG**	ORF17, 92	ORF17, 94	ORF17, 100	ORF17, 100	ORF17, 92	I17, 91	ORF15, 92	ORF15, 92	ORF16, 86	ORF16, 95	
+	**18**	14734	15861	375	42.8	5.1	AGAAAGGCGGActtcgttta**ATG**	ORF18, 94	ORF18, 94	ORF18, 95	ORF18, 100	ORF18, 92	I18, 91	ORF16, 91	ORF16, 91	ORF17, 91	ORF17, 92	
+	**19**	15851	16144	97	11.4	9.3	AGAAAGTGGAGacaaaccaa**ATG**	ORF19, 97	ORF19, 97	ORF19, 100	ORF19, 100	ORF19, 96	I19, 96	ORF17, 96	ORF17, 94	ORF18, 96	ORF18, 96	
+	**20**	16134	16943	269	29.1	5.7	TCAAGAAAGGTtaaaa**ATG**	ORF20, 88	ORF20, 87	ORF20, 100	ORF20, 100	ORF20, 45	I20, 46	ORF18, 75	ORF18, 76	ORF19, 59	ORF19, 32	Receptor-binding protein
+	**21**	16965	17318	117	13.5	6.3	AGAAAGCAAAAtaaa**ATG**	ORF21, 88	ORF21, 96	ORF21, 96	ORF21, 100	ORF21, 90	I21, 87	ORF19, 92	ORF19, 90	ORF20, 91	ORF20, 89	Holin
+	**22**	17315	18019	234	26.0	5.7	CAAACGGAGGAtaaaaaaga**ATG**	ORF22, 98	ORF22, 100	ORF22, 100	ORF22, 100	ORF22, 76	I22, 76	ORF20, 67	ORF20, 68	ORF21, 68	ORF21, 76	Endolysin
-	23	18613	18464	49	6.0	8.5	CAACTAGAAGTatagcg**TTG**	ORF23, 89	ORF23, 59	ORF23, 100	ORF23, 100	ORF23, 61	e36, 85	p21, 91	-	ORF23, 83	ORF22, 87	
-	**24**	18940	18674	88	10.2	4.6	AAATTAAAGGTatgaatag**ATG**	ORF25, 96	ORF24, 93	ORF24, 100	ORF24, 100	ORF25, 93	e33, 93	ORF21, 95	ORF21, 95	ORF24, 85	ORF25, 90	
-	**25**	19460	19113	115	13.5	9.1	ATAATTGAGGTtatagcata**ATG**	ORF26, 85	ORF25, 79	ORF25, 100	ORF25, 100	ORF26,66	e31, 65	ORF23, 90	ORF22, 90	ORF26, 73	ORF27, 81	
-	26	19627	19460	55	6.9	5.8	AGACAGGAGTAatcggata**ATG**	ORF27, 90	ORF26, 90	ORF26, 100	ORF26, 100	ORF27, 90	e30, 90	ORF24, 35	ORF23, 35	ORF27, 35	ORF28, 42	
-	27	19803	19627	58	6.8	10.3	AGAAAGCTAGTgaataat**ATG**	ORF28, 96	ORF27, 91	ORF27, 100	ORF27, 100	ORF28, 88	e29, 90	-	-	-	ORF29, 90	
-	28	19996	19805	63	7.1	10.2	AGAAAGTTTTGgtgaaaaaataa**ATG**	ORF29, 93	ORF28, 88	ORF28, 100	ORF28, 100	-	-	-	-	-	-	
-	**29**	20295	20038	85	9.8	5.0	GAAAAGGAGGTtaaata**GTG**	ORF30, 91	ORF29, 96	ORF29, 100	ORF29, 100	ORF29, 68	e28, 65	ORF25, 88	ORF24, 86	ORF28, 75	ORF30, 69	
-	30	20676	20368	102	12.1	4.6	GGCTTGGAGGTaacatcta**ATG**	ORF31, 94	ORF32, 96	ORF30, 100	ORF30, 100	ORF33, 82	e24, 91	ORF26, 80	ORF25, 78	-	ORF33, 83	
-	31	20795	20676	39	4.1	8.4	ATAAATATAGGagaacaaa**ATG**	ORF32, 92	ORF33, 89	ORF31, 100	ORF31, 100	ORF34, 62	e23, 89	ORF27, 79	ORF26, 79	-	ORF34, 79	
-	32	21095	20853	80	9.7	6.7	CGAAAGGAAGTaaatag**ATG**	ORF33, 88	ORF35, 88	ORF32, 100	ORF32, 100	-	e21, 85	ORF30, 86	ORF29, 83	ORF29, 88	ORF36, 85	
-	33	21206	21096	36	3.8	9.4	ATAAAGGAGCGataca**ATG**	-	ORF36, 100	ORF33, 100	ORF33, 100	-	e19, 91	sk1p32, 94	-	ORF30, 86	-	
-	34	21555	21241	104	11.9	4.6	ACTATGGTGGTatcgtcta**ATG**	ORF34, 92	ORF37, 90	ORF34, 100	ORF34, 100	ORF35, 86	e17, 88	ORF31, 83	ORF30, 86	-	ORF38, 82	
-	35	21746	21555	63	7.4	6.5	GGAAAGAGGAAaaa**ATG**	ORF35, 93	ORF38, 90	ORF35, 100	ORF35, 100	-	-	-	-	-	-	
-	**36**	22338	21826	170	20.3	9.7	ATATATGAGGGagtatt**ATG**	ORF36, 98	ORF39, 98	ORF36, 100	ORF36, 100	ORF37, 94	e15, 91	ORF32, 91	ORF31, 91	ORF32, 96	ORF39, 91	
-	**37**	22547	22335	70	8.3	4.7	GGAAAGTTGGTttc**ATG**	ORF37, 95	ORF40, 92	ORF37, 100	ORF37, 100	ORF38, 92	e14, 92	ORF33, 92	ORF32, 92	ORF33, 91	ORF40, 94	
-	**38**	22922	22563	119	13.0	5.0	AGAGGGAAAATaaaaa**ATG**	ORF38, 100	ORF41, 90	ORF38, 100	ORF38, 100	ORF39, 92	e13, 94	ORF34, 92	ORF33, 92	ORF34, 94	ORF41, 88	
-	**39**	23549	22926	207	23.9	6.7	CAAATGGAGAAaaaaa**ATG**	ORF39, 61	ORF42, 63	ORF39, 100	ORF39, 100	ORF40, 61	e12, 92	ORF35, 96	ORF34, 95	ORF35, 63	ORF42, 91	Sak protein
-	40	24046	23546	166	19.2	9.6	AGAAAGAGGAGaaa**ATG**	-	-	ORF40, 100	ORF40, 100	-	e11, 94	-	-	-	-	HNH endonuclease
-	41	24299	24033	88	10.5	7.8	TATTAGAAAGTtcattt**ATG**	ORF40, 91	ORF43, 90	ORF41, 100	ORF41, 100	ORF42, 86	e10, 94	ORF38, 76	ORF37, 84	ORF38, 76	ORF43, 88	
-	42	24543	24256	95	11.7	10.0	ATAAAGGAGAAata**ATG**	ORF41, 91	ORF44, 95	ORF42, 100	ORF42, 100	ORF43, 93	e9, 92	ORF39, 95	ORF38, 97	-	ORF44, 93	
-	**43**	24853	24596	85	10.1	9.1	AGCAAGGAAAGgtaacagaaa**ATG**	ORF43, 94	ORF45, 88	ORF43, 100	ORF43, 100	ORF45, 87	e7, 87	ORF41, 89	ORF40, 89	ORF40, 88	ORF46, 89	
-	44	25742	24846	298	34.2	5.7	CTAAAGGAGAAagaa**ATG**	ORF44, 45	ORF47, 45	ORF44, 100	ORF44, 100	ORF47, 96	e5, 45	ORF43, 43	ORF42, 44	ORF42, 95	ORF47, 46	DNA polymerase subunit
-	45	25947	25786	53	6.2	6.5	AGAAAAGGAGAAaaat**ATG**	ORF45, 90	-	-	ORF45, 100	ORF48, 92	-	-	-	ORF45, 76	-	
-	**46**	26957	26769	62	7.6	9.6	TTATAGGAGGTaact**ATG**	ORF47, 93	ORF50, 91	ORF47, 91	ORF46, 100	ORF52, 85	e1, 90	ORF50, 88	ORF45, 85	ORF49, 85	ORF50, 91	
+	**47**	26988	27215	75	8.5	4.9	ATAAAGAGTATaacataaa**ATG**	ORF48, 97	ORF51, 95	ORF48, 92	ORF47, 100	ORF53, 97	m1, 95	ORF51, 95	ORF46, 95	ORF51, 90	ORF51, 95	
+	**48**	27220	27351	43	5.2	4.9	TTAAGGGAGAAtaaagaa**ATG**	ORF49, 93	ORF52, 90	ORF49, 93	ORF48, 100	ORF54, 93	m2, 90	ORF52, 93	ORF47, 93	ORF52, 93	ORF52, 95	
+	**49**	27348	27827	159	17.9	6.8	TGAATTGAGTTctgattt**ATG**	ORF50, 96	ORF53, 95	ORF50, 96	ORF49, 100	ORF55, 96	m3, 92	ORF53, 96	ORF48, 94	ORF53, 94	ORF53, 91	Holliday junction endonuclease
+	**50**	27828	27995	55	6.2	6.5	AGAAAGTCAGGataagtaa**ATG**	ORF51, 96	ORF54, 91	ORF51, 96	ORF50, 100	ORF56, 98	m4, 98	ORF54, 100	ORF49, 100	ORF54, 92	ORF54, 100	
+	51	28416	28556	46	5.3	11.2	AGATAAGGGAGaagcaa**ATG**	ORF52, 97	ORF55, 86	ORF52, 86	ORF51, 100	ORF58, 76	-	-	-	-	-	

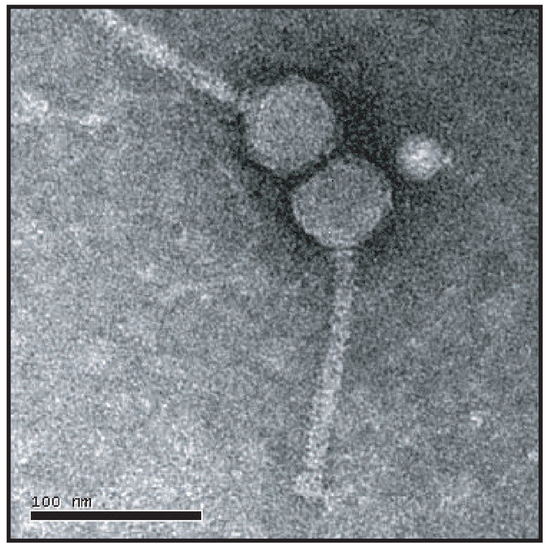

Late genes
Early genes
Middle genes
Packaging
Morphogenesis
Lysis
Replication
SL4
CB13
CB14/
GR7
CB19
CB20
5 kb
10 kb
15 kb
20 kb
25 kb
30 kb

Late genes
Early genes
Middle genes
Packaging
Morphogenesis
Lysis
Replication
jj50
sk1
712
P008
bIBB29
bIL170
SL4
CB13
CB14/
GR7
CB19
CB20
5 kb
10 kb
15 kb
20 kb
25 kb
30 kb

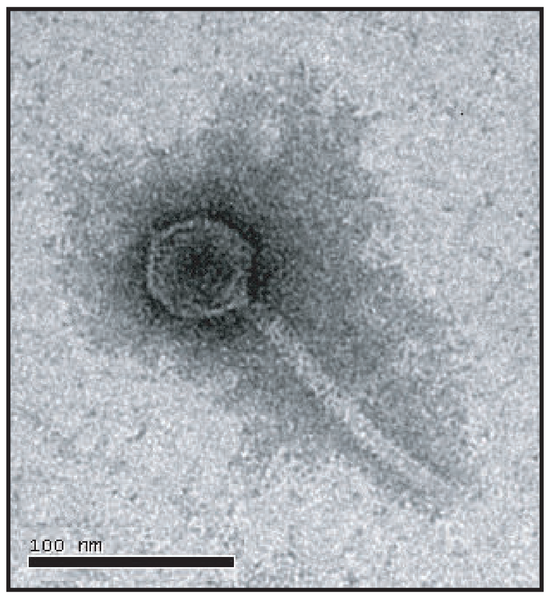
A



B



C



D

Protein #	ORF	MM (kDa) Estimated	MM (kDa) SDS-PAGE	Putative function
1	16	105.7	74.7	Tape measure protein
2	17	53.3	48.8	Unknown function
3	5	43.5	38.9	Portal protein
4	20	29.0	29.2	Receptor binding protein
5	12	79.4	80.3	Neck passage structure protein
6	16	105.6	198.1	Tape measure protein
7	12	71.3	143.5	Neck passage structure protein
8	12	71.3	62.4	Neck passage structure protein
9	18	42.8	44.0	Unknown function
10	5	43.3	42.3	Portal protein
11	17	34.4	34.6	Unknown function
12	20	29.1	29.8	Receptor binding protein