Characterization of a novel Panton-Valentine leukocidin-encoding staphylococcal phage and its natural PVL-lacking variant

Lynn El Haddad and Sylvain Moineau*

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

A new siphophage (LH1) was isolated from raw milk using a *Staphylococcus aureus* ST352 host. Its genome (46,048 bp, 57 ORFs) includes the two genes encoding for the Panton-Valentine leukocidin (PVL), a virulence factor usually harbored by *S. aureus* prophages. Nine structural proteins were identified, including a tail protein generated through a +1 frameshift. A phage lytic mutant was isolated and its analysis revealed the deletion of genes coding for the PVL and an integrase. The deletion likely occurred through recombination between direct repeats.

Staphylococcus aureus is a bacterial pathogen that can infect humans, animals, plants, and may contaminate foods. Some strains harbors several pathogenic and virulent components including exotoxins, such as enterotoxins (SE), exfoliatins, toxic-shock syndrome toxin (TSST), hemolysins, and Panton-Valentine leukocidin (PVL) (5, 7). The genes coding for this latter toxin are located on prophages (14). The PVL toxin belongs to the synergohymenotropic toxin family and is composed of two subunits, named LukF-PV and LukS-PV (18, 25, 29). These secretory proteins act together by forming pores in cell membranes, lysing neutrophils and macrophages, thus leading to pneumonia and eventual cell death (8, 19).

Since the discovery of the PVL toxin by Panton and Valentine in 1932 (25), at least eight PVL-encoding phages belonging to the *Siphoviridae* family have been characterized (14, 20, 22, 31). They can be classified into three different groups according to the replication/transcription region and morphogenesis module as well as their host range (15, 20, 31). Members of groups 1 and 3 have isometric capsids and members of group 2 have elongated capsids. The presence of the same toxin genes in morphologically distinct phages suggests that they can be readily exchanged, possibly during co-infection.

Contrarily to temperate phages, virulent phages can be used as biocontrol agents against pathogenic strains in medical and food applications. The presence of an integrase gene and toxin genes in phage genomes leads to poor antibacterial efficacy and safety concerns, respectively. Consequently, only virulent phages (or temperate phages replicating on indicator strains) lacking any virulence factor have been used as biocontrol agents against *S. aureus* in foods (9, 11, 12, 23) and in animal models (6, 21, 26). In an effort to isolate new virulent phages to control *Staphylococcus* contaminations in dairy products, we analyzed raw milk samples for the presence of staphylococcal phages.

One raw milk sample was obtained from six dairy farms. One hundred microliters of an overnight culture of *S. aureus* 01 (ST352) grown in TSB at 37ºC was added to 5 ml of 2X tryptic soy broth (TSB) and 5 ml of the milk sample. The culture was incubated overnight at 37ºC. Five ml of the first amplification was added to 5 ml of 2X TSB and incubated overnight at 37ºC. The latter step was repeated one more time. Then, the presence of phages was tested by depositing 5 µl of the last amplification on a tryptic soy agar (TSA) plate containing *S. aureus* 01. Phages infecting *S. aureus* 01 were isolated from only one sample out of the six analyzed. Several clear phage plaques were picked, purified, and characterized. The same restriction profile was obtained for all the isolates and one was randomly selected and named phage LH1. The isolation of LH1 confirms raw milk as a source of staphylococcal phages. The *S. aureus* strain used as host was previously isolated from a raw milk cheese.

While clear phage plaques could be observed on TSA plates, LH1 did not produce a clear lysate in TSB. However, while testing conditions to amplify this phage in broth, a clear lysate was spontaneously obtained following an incubation of infected cells at 37ºC for 8 hours, an overnight storage of the infected cells at 4ºC, and a reincubation of the mixture at 37ºC. Analysis of several plaques from this clear lysate led to the isolation of a phage variant with a different restriction profile. This phage was named LH1-MUT.

Phage LH1 has an elongated capsid with a length of 96 ± 19 nm and a width of 45 ± 2 nm. It also has a non-contractile tail of 323 ± 11 nm in length, with a width of 11 ± 1 nm (see Figure S1 in the supplemental material). Similar measurements were obtained for LH1-MUT suggesting that the mutation(s) did not occur in the morphogenesis genes. These two phages belong to group B2 of the *Siphoviridae* family (1).

69 The infection cycles of both phages LH1 and LH1-MUT were evaluated during growth on
70 the host strain at 37ºC in TSB as described elsewhere (28). The latent period of LH1 was
71 estimated at 70 minutes and the burst size at 5 ± 1 pfu per infected cell. The low burst size of
72 phage LH1 may explain the difficulty in obtaining a clear lysate during amplification in TSB. On
73 the other hand, LH1-MUT had a slightly longer latent period of 80 minutes but the burst size was
74 4-fold higher, with 21 ± 4 new virions per infected cell. Other reported *S. aureus* phages have
75 shorter latent periods (25 to 45 minutes) and larger burst sizes (27 to 100) (10, 21, 26).
76 LH1 and LH1-MUT were further characterized by determining their host range on a panel
77 of 14 strains of *S. aureus*. Two of the 14 *S. aureus* strains were isolated from Canadian raw
78 cheeses, six strains were obtained from the Canadian Bovine Mastitis Research Network, four
79 reference strains were obtained from the Félix d'Hérelle Center (www.phage.ulaval.ca), and the
80 last two were MRSA strains obtained from the Public Health Agency of Canada. Both phages
81 infected the two raw cheese isolates as well as two strains isolated from mastitis. Phage LH1-
82 MUT infected three additional mastitis isolates (total of 7/14 strains). It seems that these two
83 phages preferably infect *S. aureus* strains isolated from milk environments, with phage LHI-
84 MUT having the wider host range. However, compared to staphylococcal phages belonging to the
85 two other *Caudovirales* families (*Myoviridae*, *Podoviridae*), siphophages LH1 and LH1-MUT
86 have a narrow host range (24, 30).
87 In order to determine if phages LH1 and LH1-MUT carry undesirable characteristics for
88 biocontrol purposes, we sequenced their genomes. The protocols for DNA isolation, sequencing,
89 and analysis are reported elsewhere (28). The linear genome of LH1 is 46,048 bp in length
90 (GenBank accession number JX174275). It has a GC content of 33.2%, which is similar to that of
91 other PVL phages (15, 22) and of its *S. aureus* hosts (32.8%) (17). The genome possesses

92 cohesive extremities (*cos*-type), made of ten complementary bases (5'-CCGGAGAGGC-3'). No
93 tRNA was found in the genome. Using various bioinformatic tools, putative functions were
94 attributed to 26 of the 57 ORFs (46%) identified in the genome of LH1 (Table S1). Of interest
95 was the presence of a gene likely coding for an integrase as well as two genes coding for the
96 subunits of the PVL toxin, lukS-PV and lukF-PV, indicating that phage LH1 belong to the group
97 of PVL-encoding staphylococcal phages. With their elongated capsids, they also belong to group
98 2 of PVL phages (4, 31).

99 Starting from one extremity (*cos*-site), the linear genome of phage LH1 was divided into
100 five regions (Fig. 1A) including the genes coding for the: 1) morphogenesis/structural proteins
101 (packaging, capsid, tail), 2) lysis module, 3) virulence factors (two subunits of the PVL toxin,
102 lukS-PV and lukF-PV), 4) lysogeny module, and 5) replication/transcription region. This genome
103 organization is conserved among the PVL-encoding phages (31). Figures S2 an S3 in
104 supplementary material provides an alignment of the PVL phage genomes available on NCBI,
105 showing the five regions as well as a phylogenic tree of staphylococcal phages, respectively.

106 The structural proteome of phage LH1 was determined by separating a CsCl purified
107 phage preparation on a 12% SDS-polyacrylamide gel. The 11 bands visualized using Coomassie
108 blue were sent for identification by LC/MS-MS along with the complete purified phage (Fig. 1B).
109 Overall, 8 structural proteins were identified, namely ORF3 (portal), ORF5 (MCP), ORF8,
110 ORF10 (MTP), ORF11, ORF14 (TMP), ORF16, and ORF18 (receptor binding protein RBP)
111 (Fig. 1BC). No new protein was detected when analyzing the complete phage sample. Most of
112 the observed molecular masses of the phage proteins matched the theoretical values except for
113 the major capsid protein (MCP) and the major tail protein (MTP). The ORF5 (MCP), with a
114 calculated molecular mass of 45 kDa, was associated with two protein bands with estimated

115 molecular masses of 32 and 28 kDa (Fig. 1BC). This suggested that the major capsid protein was
116 likely processed (2). The ORF10 (MTP) was also found in two protein bands (bands 8 and 9)
117 with molecular masses of 26 and 22 kDa, respectively. The shorter version corresponds to the
118 expected size (23 kDa) of ORF10 based on bioinfomatic analysis. The longer version could be
119 explained by a +1 frameshift, which added 19 amino acids to the protein. This extra peptide was
120 found by LC/MS-MS analysis of band 8 but was not found in band 9. The frameshift was likely
121 facilitated by a "slippery zone" of CCC.AGC (32). Furthermore, the ORF18 was recognized as
122 the receptor binding protein because of its high similarity to the ORF636 of phage ΦSLT acting
123 as an adhesion protein for a chain of lipoteichoic acid on the cell surface of *S. aureus* (16). Taken
124 altogether, our proteomic data found resembled that of phage ΦSLT (22). In fact, there is a 75%
125 of identity between phages ΦSLT and LH1 at the genomic level.

126 As for the lysis module and the region containing the two subunits LukS-PV and LukF-
127 PV, they are much conserved among the PVL-encoding phages (31). The lysis module was
128 composed of a 115-aminoacid holing, which was characterized by *i)* the presence of two
129 hydrophobic transmembrane domains (TM) which puts it in the class II holins, *ii)* highly charged
130 hydrophilic C-terminal domain, and *iii)* no dual-start motif. This module also contains its
131 associated endolysin having a catalytic domain (N-terminal) and a cell-wall binding domain (C-
132 terminal). The former contains a CHAP domain (Cysteine, Histidine-dependent
133 amidohydrolases/peptidase) that cleaves the peptidoglycan peptide chain between the acetyl-
134 group of the N-acetylemuramic acid and the proximal L-alanine (EC 3.5.1.28). Based on
135 sequence similarity, this endolysin is likely an N-acetylmuramoyl-L-alanine amidase.

136 We compared the two subunits LukS-PV and LukF-PV in all the PVL-encoding phages
137 (Fig. S2 in supplementary material). The PVL genes are located in the same genomic region for

138 all the PVL phages, between the lysis module and the attachment site (attP) within the lysogeny
139 genes (Fig. 1). It is unclear at this time why these virulence factors are anchored in this region,
140 although it is likely optimal for toxin expression.

141 Genome analysis of the virulent phage LH1 also led to the unexpected identification of a
142 lysogeny module. This region was found to contain an integrase gene, *int* (*orf27*), a repressor
143 gene, *rep* (*orf30*), and an anti-repressor gene, *ant* (*orf32*). This region is highly homologous
144 among PVL prophages. Downstream of the *int* gene was a possible phage attachment site, *attP*
145 (Fig. 2), which is identical to the *attP* of phage ΦPVL (15). It included five direct repeats of 5'-
146 AGGGCNN-3' where NN stands for AA, AG or GG, as well as the 29-nucleotide core
147 attachment site. Furthermore, a 9-base inverted repeat with a 3-base loop
148 (ATTTAGTACtagGTACTAAAT) was found between the integrase (*orf27*) and *orf28*. This
149 sequence is observed in other staphylococcal siphophages suggesting a regulatory function (13).

150 We could not obtain any PCR products using LH1 specific primers and DNA isolated
151 from the bacterial host indicating that this phage did not originate from *S. aureus* strain 01. In
152 addition, the remaining *S. aureus* 01 cells in the turbid broth during the amplification of phage
153 LH1 were not lysogenized by this phage and they were still phage-sensitive (data not shown).
154 Nonetheless, the presence of the two genes *lukS-PV* and *lukF-PV* along with the genes *int, rep*,
155 and *ant* suggest that the ancestor of LH1 may have been a prophage.

156 Finally, the DNA replication module, along with the transcriptional regulation module,
157 included mostly genes coding for DNA binding proteins, a DNA polymerase, a gene coding for
158 VirE found in other staphylococcal phages (10, 20), an helicase, a transcriptional regulator
159 (RinA-type), and a HNH endonuclease. Altogether, the most conserved regions among the three
160 groups of PVL-encoding phages were the lysis module, the toxin genes, and the integrase.

The genome of LH1-MUT was also sequenced for comparison purposes with phage LH1. It has 41,949 bp and a GC content of 33.6%. Phage LH1-MUT lost a 4,099-bp region while the rest of its genome was 100% identical to LH1. A detailed analysis of this deleted region (from position 23,711 to 27,810 according to LH1 genome) indicated the loss of three genes, namely those coding for the two PVL subunits as well as the putative integrase. The deletion site was precisely flanked by a heptanucleotide direct repeat, TTTTACA, and only one repeat was retained in LH1-MUT. In fact, these two repeats may even be expanded to 12 nucleotides (TTAACTTTTACA) with only one mismatch (Fig. 2). This suggests that the deletion may have occurred through intra-molecular recombination between direct repeats of homologous DNA (27). Interestingly, the conserved inverted repeats are located in vicinity of one of the repeats. Many spontaneous deletions between short dispersed DNA sequence repeats are found associated with nearby inverted repeat sequences (3).

Nonetheless, the entire *lukS-PV–lukF-PV–int* sequence was removed from the LH1 genome, generating a lytic mutant phage lacking the PVL subunits and possibly incapable of integrating into the bacterial chromosome of some strains. This deletion somehow led to an increased burst size and host range. Unlike the wild-type phage LH1, LH1-MUT may be used as a potential candidate as biocontrol agent against specific *S. aureus* strains. Our study also indicates that raw milk may carry staphylococcal phages encoding virulence factors.

Acknowledgments. We thank B.D. Conway for editorial assistance, S. Labrie for milk samples, D. Tremblay for electron microscopy, and J. Dumaresq for insightful discussion. L.H. was a recipient of a scholarship from the CIDA. We are grateful to the MAPAQ, CBMRN, and NML of Canada for staphylococcal strains. This work was funded by the FQRNT, MAPAQ, AAC, and Novalait. S.M. is chairholder of the Canada Research Chair in Bacteriophages.

References

1. **Ackermann, H.W.** 2009. Phage classification and characterization. Methods Mol. Biol. **501**:127-140.
2. **Briggiler Marcó, M., J.E. Garneau, D. Tremblay, A. Quiberoni, and S. Moineau.** 2012. Characterization of two virulent phages of *Lactobacillus plantarum.* Appl. Environ. Microbiol. **78**:8719-8734.
3. **Bzymek, M., and S.T. Lovett.** 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci U S A. **98**:8319-8325.
4. **Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brussow.** 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. **67**:238-276, table of contents.
5. **Cao, R., N. Zeaki, N. Wallin-Carlquist, P.N. Skandamis, J. Schelin, and P. Radstrom.** 2012. Elevated enterotoxin A expression and formation in *Staphylococcus aureus* and its association with prophage induction. Appl. Environ. Microbiol. **78**:4942-4948.
6. **Capparelli, R., M. Parlato, G. Borriello, P. Salvatore, and D. Iannelli.** 2007. Experimental Phage Therapy against *Staphylococcus aureus* in Mice. Antimicrob. Agents Chemother. **51**:2765-2773.
7. **Dale, B., and A. Brown**. 2007. Panton-Valentine leukocidin producing *Staphylococcus aureus*. in *the 37th annual infection control conference*. Brighton.
8. **Finck-Barbancon, V., G. Duportail, O. Meunier, and D.A. Colin.** 1993. Pore formation by a two-component leukocidin from *Staphylococcus aureus* within the membrane of human polymorphonuclear leukocytes. Biochim. Biophys. Acta. **1182**:275-282.
9. **Garcia, P., C. Madera, B. Martinez, A. Rodriguez, and J. Evaristo Suarez.** 2009. Prevalence of bacteriophages infecting *Staphylococcus aureus* in dairy samples and their potential as biocontrol agents. J. Dairy Sci. **92**:3019-3026.
10. **Garcia, P., B. Martinez, J.M. Obeso, R. Lavigne, R. Lurz, and A. Rodriguez.** 2009. Functional genomic analysis of two *Staphylococcus aureus* phages isolated from the dairy environment. Appl. Environ. Microbiol. **75**:7663-7673.
11. **Garcia, P., L. Rodriguez, A. Rodriguez, and B. Martinez.** 2010. Food biopreservation: promising strategies using bacteriocins, bacteriophages and endolysins. Trends Food Sci. Technol. **21**:373-382.
12. **Gill J.J., P.M.S., K.E. Leslie, M.W. Griffiths,.** 2006. Bovine whey proteins inhibit the interaction of *Staphylococcus aureus* and bacteriophage K. J. Appl. Microbiol. **101**:377-386.
13. **Iandolo, J.J., V. Worrell, K.H. Groicher, Y. Qian, R. Tian, S. Kenton, A. Dorman, H. Ji, S. Lin, P. Loh, S. Qi, H. Zhu, and B.A. Roe.** 2002. Comparative analysis of the genomes of the temperate bacteriophages phi11, phi12 and phi13 of *Staphylococcus aureus* 8325. Gene. **289**:109-118.
14. **Kaneko, J., T. Kimura, Y. Kawakami, T. Tomita, and Y. Kamio.** 1997. Panton-valentine leukocidin genes in a phage-like particle isolated from mitomycin C-treated *Staphylococcus aureus* V8 (ATCC 49775). Biosci. Biotechnol. Biochem. **61**:1960-1962.
15. **Kaneko, J., T. Kimura, S. Narita, T. Tomita, and Y. Kamio.** 1998. Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton-Valentine leukocidin genes. Gene. **215**:57-67.

16. **Kaneko, J., S. Narita-Yamada, Y. Wakabayashi, and Y. Kamio.** 2009. Identification of ORF636 in phage phiSLT carrying Panton-Valentine leukocidin genes, acting as an adhesion protein for a poly(glycerophosphate) chain of lipoteichoic acid on the cell surface of *Staphylococcus aureus.* J. Bacteriol. **191**:4674-4680.
17. **Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K.-i. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N.K. Takahashi, T. Sawano, R.-i. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu.** 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus.* The Lancet. **357**:1225-1240.
18. **Lina, G., Y. Piemont, F. Godail-Gamot, M. Bes, M.O. Peter, V. Gauduchon, F. Vandenesch, and J. Etienne.** 1999. Involvement of Panton-Valentine leukocidin-producing *Staphylococcus aureus* in primary skin infections and pneumonia. Clin. Infect. Dis. **29**:1128-1132.
19. **Lipinska, U., K. Hermans, L. Meulemans, O. Dumitrescu, C. Badiou, L. Duchateau, F. Haesebrouck, J. Etienne, and G. Lina.** 2011. Panton-Valentine leukocidin does play a role in the early stage of *Staphylococcus aureus* skin infections: a rabbit model. Plos One. **6**:e22864.
20. **Ma, X.X., T. Ito, Y. Kondo, M. Cho, Y. Yoshizawa, J. Kaneko, A. Katai, M. Higashiide, S. Li, and K. Hiramatsu.** 2008. Two Different Panton-Valentine Leukocidin Phage Lineages Predominate in Japan. J. Clin. Microbiol. **46**:3246-3258.
21. **Matsuzaki, S., M. Yasuda, H. Nishikawa, M. Kuroda, T. Ujihara, T. Shuin, Y. Shen, Z. Jin, S. Fujimoto, M.D. Nasimuzzaman, H. Wakiguchi, S. Sugihara, T. Sugiura, S. Koda, A. Muraoka, and S. Imai.** 2003. Experimental protection of mice against lethal *Staphylococcus aureus* infection by novel bacteriophage phi MR11. J. Infect. Dis. **187**:613-624.
22. **Narita, S., J. Kaneko, J. Chiba, Y. Piemont, S. Jarraud, J. Etienne, and Y. Kamio.** 2001. Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, phiSLT. Gene. **268**:195-206.
23. **O'Flaherty, S., A. Coffey, W.J. Meaney, G.F. Fitzgerald, R.P. Ross.** 2005. Inhibition of bacteriophage K proliferation on *Staphylococcus aureus* in raw bovine milk. Lett. Appl. Microbiol. **41**:274-279.
24. **O'Flaherty, S., J. Flynn, A. Coffey, G. Fitzgerald, B. Meaney, and P. Ross.** 2005. Molecular Characterisation of Bacteriophage K Towards Applications for the Biocontrol of Pathogenic Staphylococci, In *Molecular Characterization of Bacteriophage K and the Lysin LysK towards Applications for the Biocontrol of Pathogenic Staphylococci.* Department of Microbiology. **Doctorate**:39 pp.
25. **Panton, P.N., and F.C.O. Valentine.** 1932. Staphylococcal toxin. Lancet. **1**:506-508.
26. **Pantucek, R., A. Rosypalová, J. Doskar, J. Kailerová, V. Ruzicková, P. Borecká, S. Snopková, R. Horváth, F. Götz, and S. Rosypal.** 1998. The Polyvalent Staphylococcal Phage phi812: Its Host-Range Mutants and Related Phages. Virology. **246**:241-252.
27. **Pierce, J.C., D. Kong, and W. Masker.** 1991. The effect of the length of direct repeats and the presence of palindromes on deletion between directly repeated DNA sequences in bacteriophage T7. Nucleic Acids Res. **19**:3901-3905.
28. **Samson, J.E., and S. Moineau.** 2010. Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. Appl. Environ. Microbiol. **76**:6843-6852.

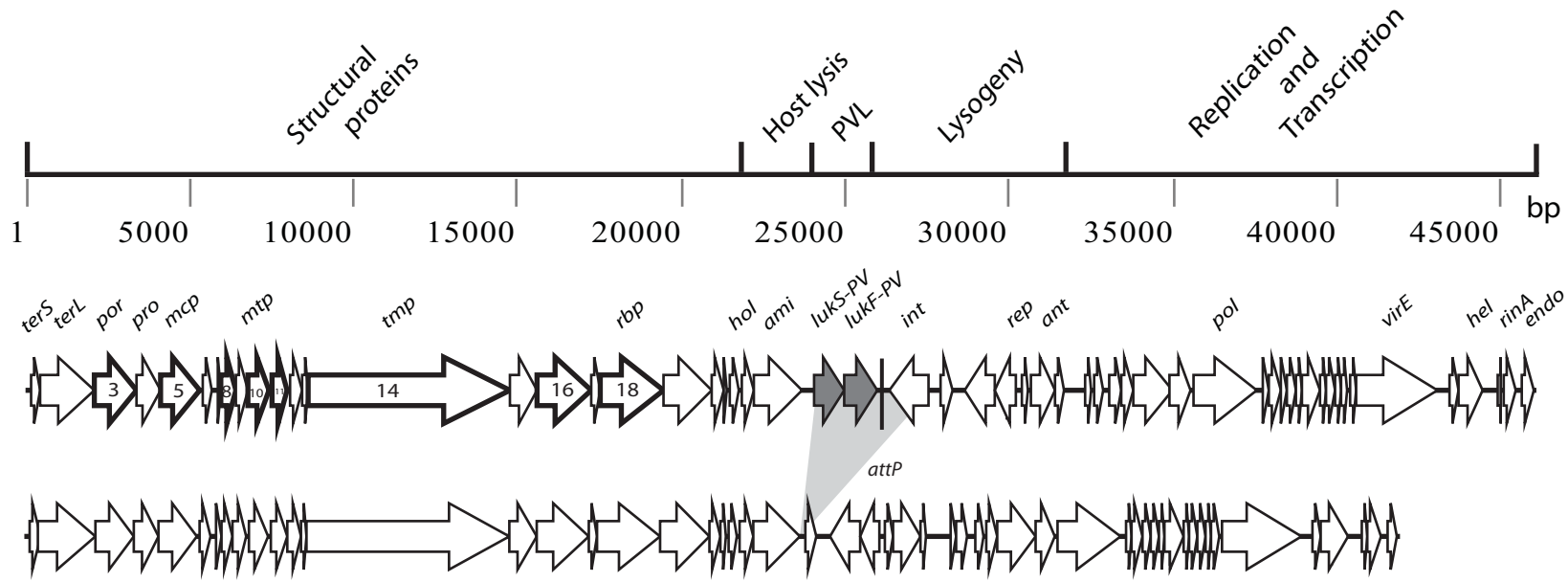
29. **Supersac, G., G. Prevost, and Y. Piemont.** 1993. Sequencing of leucocidin R from *Staphylococcus aureus* P83 suggests that staphylococcal leucocidins and gamma-hemolysin are members of a single, two-component family of toxins. Infect. Immun. **61**:580-587.
30. **Synnott, A.J., Y. Kuang, M. Kurimoto, K. Yamamichi, H. Iwano, and Y. Tanji.** 2009. Isolation from sewage influent and characterization of novel *Staphylococcus aureus* bacteriophages with wide host ranges and potent lytic capabilities. Appl. Environ. Microbiol. **75**:4483-4490.
31. **Zhang, M., T. Ito, S. Li, J. Jin, F. Takeuchi, T.L. Lauderdale, M. Higashide, and K. Hiramatsu.** 2011. Identification of the third type of PVL phage in ST59 methicillin-resistant *Staphylococcus aureus* (MRSA) strains. FEMS Microbiol. Lett. **323**:20-28.
32. **Zimmer, M., E. Sattelberger, R.B. Inman, R. Calendar, and M.J. Loessner.** 2003. Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed + 1 translational frameshifting in structural protein synthesis. Mol. Microbiol. **50**:303-317.

Legends of Figures

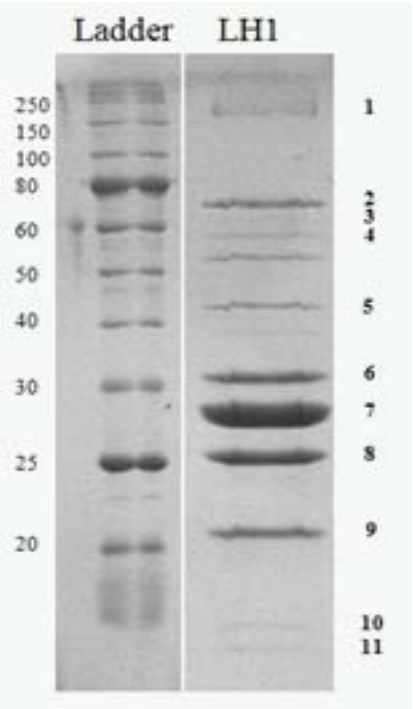
Figure 1. A) Genome alignments of LH1 and LH1-MUT. The genomes are divided into five regions including the structural module, host lysis, the genes coding for LukS-PV and LukF-PV (in LH1 genome), the lysogeny module (in LH1 genome), and the replication-transcription region. Each arrow represents an ORF. The two ORFs that encode for LukS-PV (left) and LukF-PV (right) are shown in dark gray. A light grey shadow shows the deletion of the three genes *LukS-PV*, *LukF-PV*, and *int* in LH1-MUT genome compared to LH1. The ORFs that encode for the structural proteins of LH1 identified by LC/MS-MS are numbered and presented in bold. **B)** Proteins bands shown on a SDS-PAGE followed by Coomassie blue staining (Protein Ladder 10-250 kDa, New England Biolabs). **C)** Identification of the different structural proteins by mass spectrometry.

Figure 2. Part of the LH1 genome showing 1) a magnified view of both extremities of the deleted region in LH1-MUT genome compared to LH1 genome (on the left). The alignment was made using ClustalW2 software. The direct repeats are underlined and the palindromic sequence forming a stem loop is boxed; 2) the nucleotide sequence of the attachment site attP (on the right). The attachment site is numbered according to its position on LH1 genome. It is located between the lukF-PV gene (*lukF-PV*) and the integrase gene (*int*). The five direct repeats are shown in bold characters and the core sequence of 29 nucleotides is underlined and in bold characters.

A

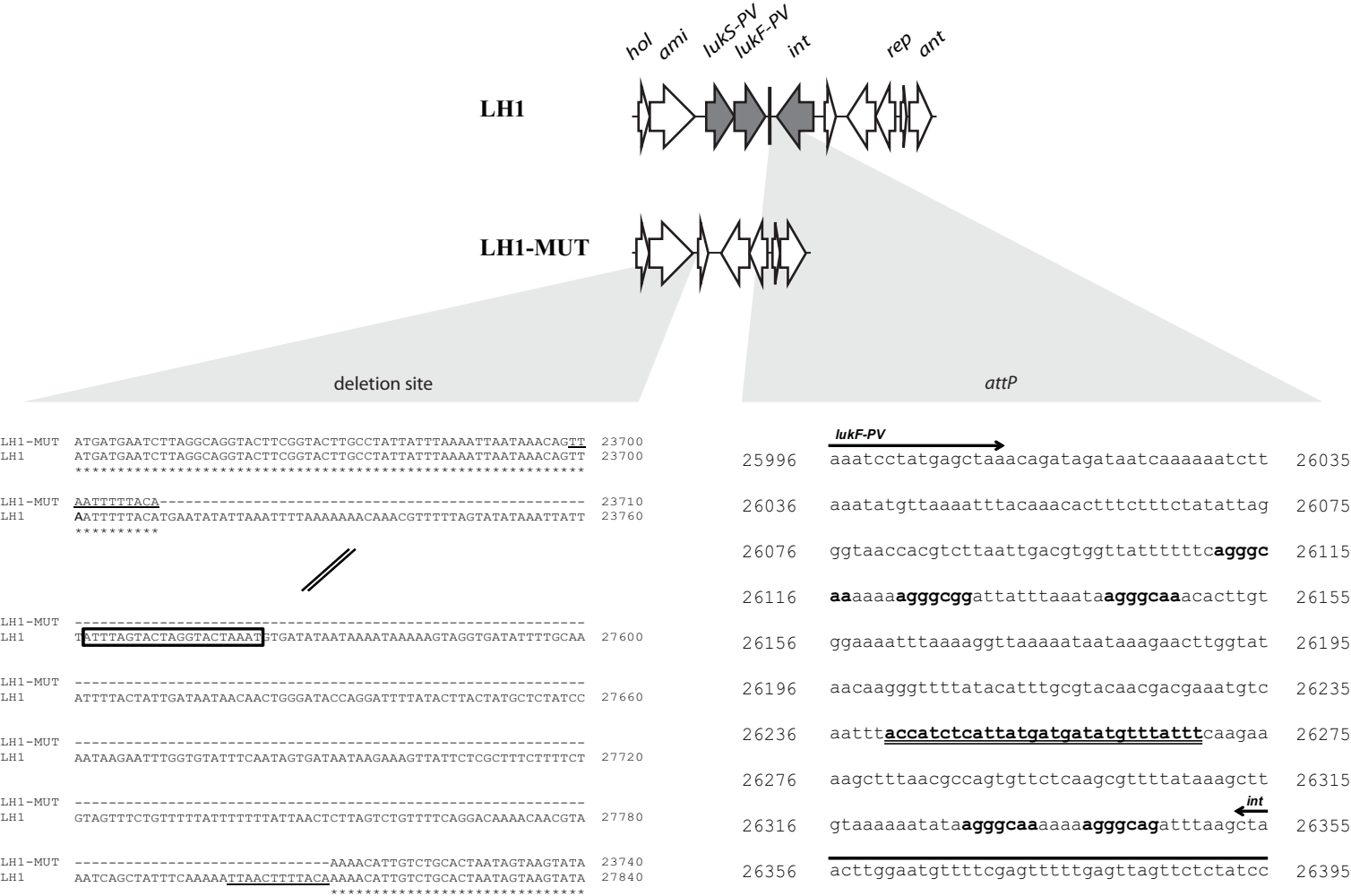


B



C

Band	Molecular Mass (kDa) SDS-PAGE	Molecular Mass (kDa) Calculated	ORF	Putative function
1	>200	226	14	Tail length tape measure protein
2	75	73	18	Receptor binding protein
3	62	61	16	Minor structural protein
4	59	61	16	Minor structural protein
5	43	48	3	Portal protein
6	32	45	5	Major capsid protein
7	28	45	5	Major capsid protein
8	26	23	10	Major tail protein
9	22	23	10	Major tail protein
10	15	15	8	Minor structural protein
11	15	16	11	Minor tail protein


hol
ami
lukS-PV
lukF-PV
int
rep
ant
LH1
LH1-MUT
deletion site
attP
LH1-MUT ATGATGAATCTTAGGCAGGTACTTCGGTACTTGCCTATTATTTAAAATTAATAAACAGTT 23700
LH1 ATGATGAATCTTAGGCAGGTACTTCGGTACTTGCCTATTATTTAAAATTAATAAACAGTT 23700
LH1-MUT AATTTTTACA 23710
LH1 AATTTTTACATGAATATATTAAATTTTAAAAAAACAAACGTTTTTAGTATATAAATTATT 23760
LH1-MUT 27600
LH1 TATTTAGTACTAGGTACTAAATGTGATATAATAAAATAAAAAGTAGGTGATATTTTGCAA 27600
LH1-MUT 27660
LH1 ATTTTACTATTGATAATAACAACTGGGATACCAGGATTTTATACTTACTATGCTCTATCC 27660
LH1-MUT 27720
LH1 AATAAGAATTTGGTGTATTTCAATAGTGATAATAAGAAAGTTATTCTCGCTTTCTTTTCT 27720
LH1-MUT 27780
LH1 GTAGTTTCTGTTTTTATTTTTTTATTAACTCTTAGTCTGTTTTCAGGACAAAACAACGTA 27780
LH1-MUT AAAACATTGTCTGCACTAATAGTAAGTATA 23740
LH1 AATCAGCTATTTCAAAAATTAACTTTTACAAAAACATTGTCTGCACTAATAGTAAGTATA 27840
lukF-PV
25996 aaatcctatgagctaaacagatagataatcaaaaaatctt 26035
26036 aaatatgttaaaatttacaaacactttctttctatattag 26075
26076 ggtaaccacgtcttaattgacgtggttatttttttcagggc 26115
26116 aaaaaaagggcggattatttaaataagggcaaacacttgt 26155
26156 ggaaaatttaaaaggttaaaaataataaagaacttggtat 26195
26196 aacaagggttttatacatttgcgtacaacgacgaaatgtc 26235
26236 aatttaccatctcattatgatgatatgtttatttcaagaa 26275
26276 aagctttaacgccagtgttctcaagcgttttataaagctt 26315
int
26316 gtaaaaaatataagggcaaaaaaagggcagatttaagcta 26355
26356 acttggaatgttttcgagtttttgagttagttctctatcc 26395

Supplemental Materials

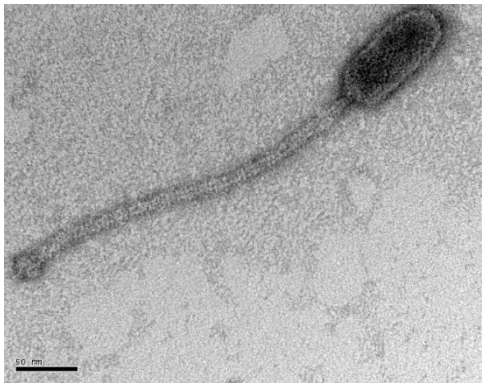


Figure S1. Electron micrograph of *S. aureus* phage LH1. The bar scale indicates 50 nm.

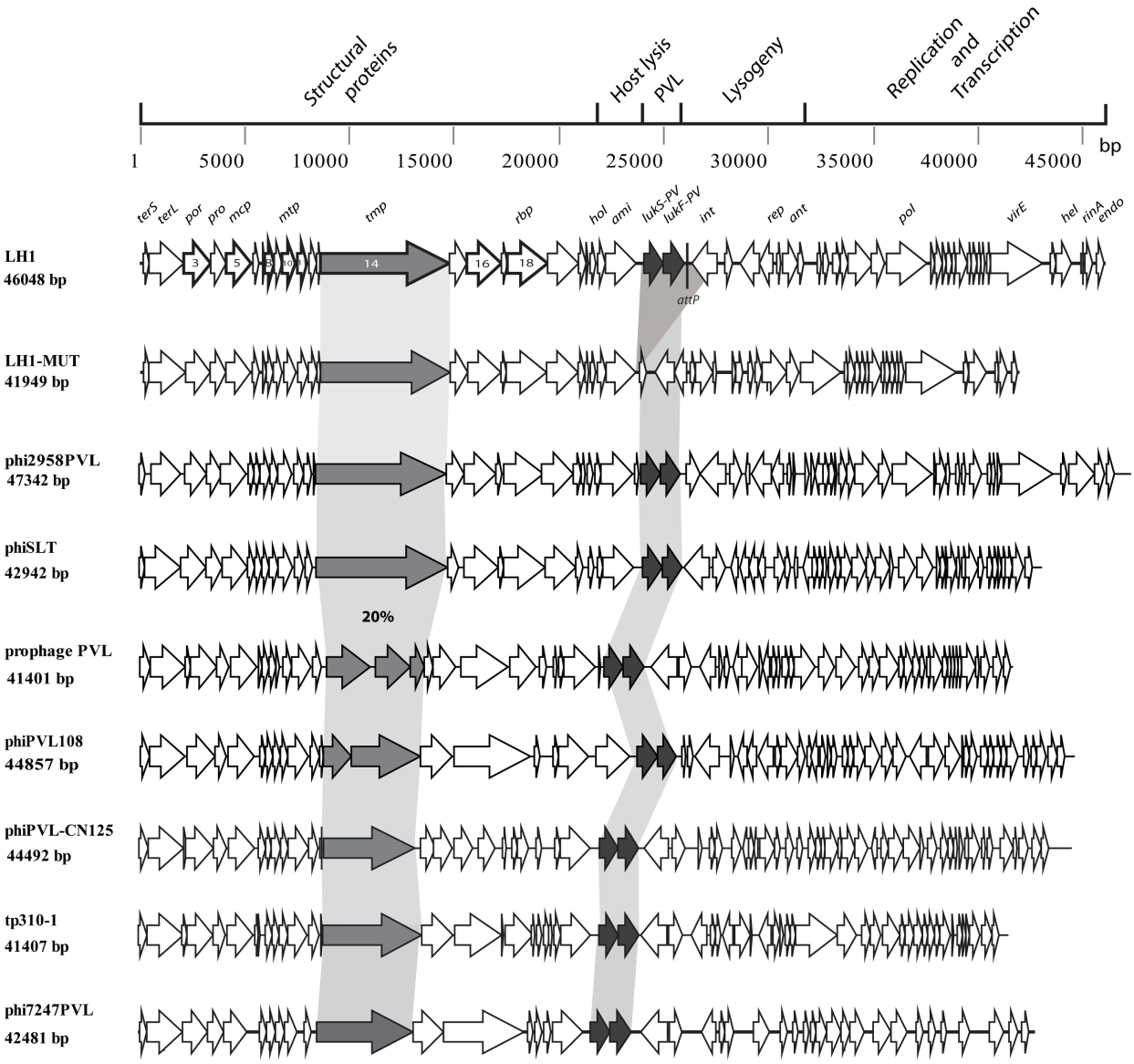


Figure S2. Genome alignments of all PVL-carrying phages available including LH1 and LH1-MUT. The genomes are divided into five regions including the structural module, host lysis, the genes coding for LukS-PV and LukF-PV, the lysogeny module, and the replication-transcription region. Each arrow represents an ORF. The two ORFs that encode for LukS-PV (left) and LukF-PV (right) are shown in dark gray. The ORF coding for the major tail protein is presented in light grey. These ORFs share more than 90% amino acid identity unless noted otherwise. A dark grey shadow shows the deletion of the three genes *LukS-PV*, *LukF-PV*, and *int* in LH1-MUT genome compared to all the other PVL-encoding phages.

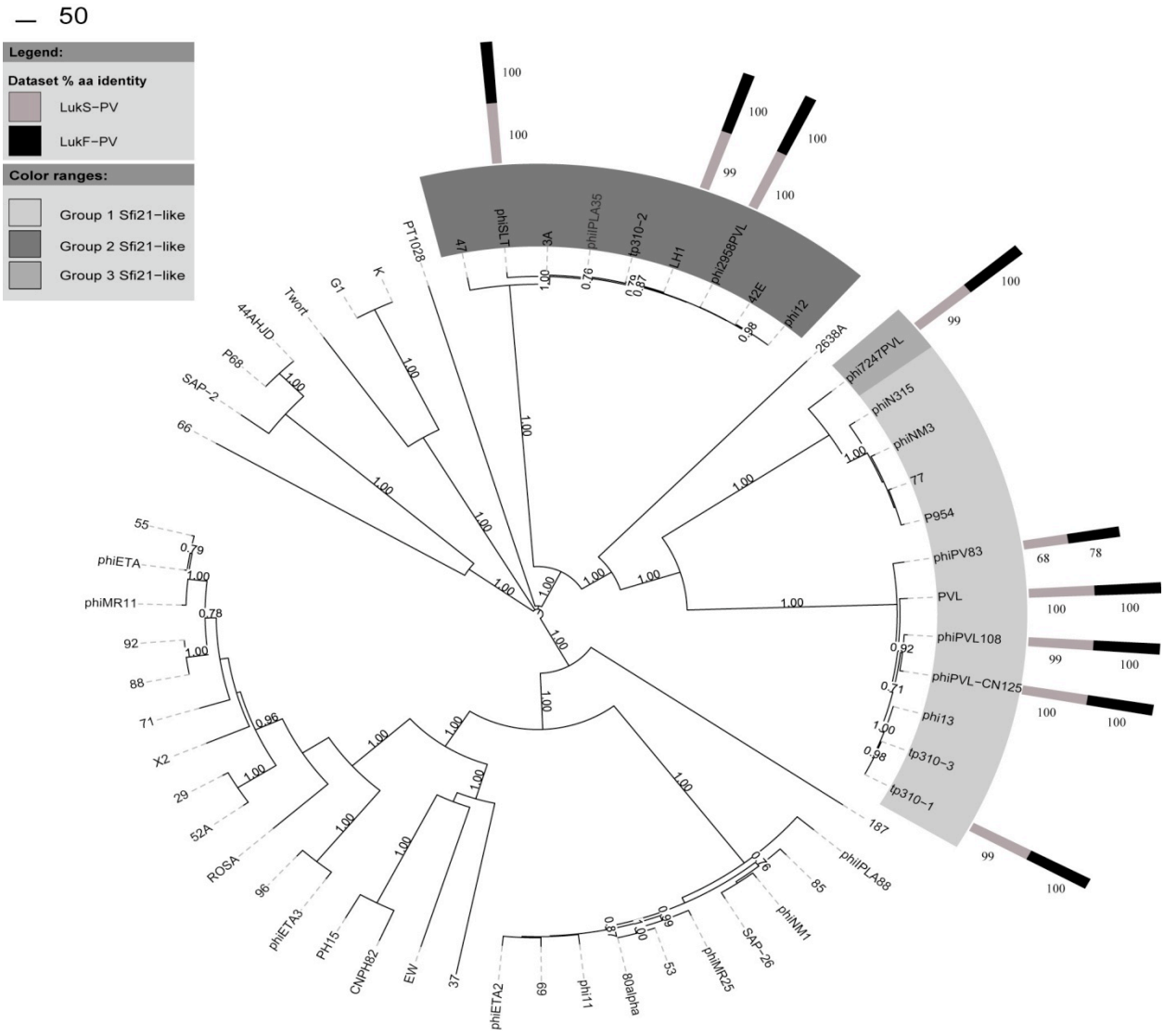


Figure S3. Comparison of staphylococcal phage complete proteome and phylogenic tree constructed using MEGA 5 software and optimized with the iTOL software and the neighbor-joining method. All phages available on the NCBI website are presented in this tree. The model used was the number of differences, the tree was constructed to be unrooted, and the parameters used to test the phylogeny were bootstraps from 1000 replicates with a random seed. The bar shows the difference between the phages in amino acids.

Table S1. ORF identification, putative function, and comparison with sequences available in public databases.

ORF	Start	End	Length (a.a.)	pI/Mw (kDa)	SD sequence 5'- AGGAGG - 3'	Putative function of the deduced protein	Best hit with BLAST	Number of identical a.a./total number of a.a. (%)	Length (a.a.)	E-value	Accession number (GenBank)
1	48	353	101	7.8/11.72	AGGGGGtctttatATG	Small subunit of the terminase	3A, ORF037	101/101 (100)	101	2e-51	YP_239934.1
2	334	2034	566	5.7/65.5	AGAAGAagGTG	Large subunit of the terminase	3A, ORF005	562/563 (99)	563	0	YP_239935.1
3	2039	3277	411	5.8/47.7	ATGCGTtaaggagGTG	Portal protein	phiSLT, gp40	411/412 (99)	412	0	NP_075502.1
4	3249	4034	261	5.0/29.6	AGAAAAtctttgaaagGTG	Protease associated with the capsid maturation	phiSLT, gp41	256/257 (99)	257	1e-145	NP_075503.1
5	4001	5209	402	5.1/45.2	AAGAGAaaaaattaaacgcgaATG	Major capsid protein	phi2958PVL, gp38	397/402 (98)	402	0	YP_002268008.1
6	5278	5556	92	5.1/10.9	AGGGGTgatgaaATG		phiSLT, gp43	92/92 (100)	92	8e-46	NP_075505.1
7	5748	5900	50	8.06/5.9	AGGACTaaccataattATG		3A, ORF036	50/50 (100)	110	6e-21	YP_239941.1
8	5897	6298	133	9.7/15.2	AGTGGTtttatcagaaaaATG	Minor structural protein	3A, ORF028	133/133 (100)	133	2e-71	YP_239942.1
9	6299	6694	131	5.6/15.7	AGGAGTtggccagataaATG		3A, ORF029	131/131 (100)	131	5e-68	YP_239943.1
10	6729	7370	213	5.0/23.4	AGGAGGaaataagcaATG	Major tail protein	phi12, gp37	213/213 (100)	213	3e-120	NP_803343.1
11	7462	7917	151	4.6/16.0	AGGAGAataatttATG	Minor tail protein	phi12, gp38	151/151 (100)	151	3e-80	NP_803344.1
12	7975	8325	116	4.6/13.6	AGGAGCtaatacaATG		phiSLT, gp49	115/116 (99)	116	1e-58	NP_075511.1
13	8367	8525	52	4.9/6.2	AGAAAAtacaacgtttctgtATG		phi2958PVL, gp45	52/52 (100)	52	5e-21	YP_002268016.1
14	8539	14739	2066	9.8/22.6	AGGAGGttaatATG	Tail-length tape measure protein	47, ORF001	2062/2066 (99)	2066	0	YP_240016.1
15	14739	15563	274	5.1/31.2	GGGAGGtttgactaattaATG		phi12, gp41	274/274 (100)	274	3e-156	NP_803347.1
16	15572	17155	527	5.7/60.9	AGGTAGgtgatttaATG	Minor structural protein	phiSLT, gp53	526/227 (99)	527	0	NP_075515.1
17	17155	17445	96	5.2/10.5	AGGGAGtgtattaatataATG		3A, ORF044	96/96 (100)	96	5e-48	YP_239952.1
18	17461	19371	636	5.8/73.1	AGGAAGgtgcattATG	Receptor binding protein	phiSLT, ORF636	634/636 (99)	636	0	NP_075517.1
19	19371	20837	488	5.5/54.2	AGGGAGaattataagtta**ATG**		phiIPLA35, gp55	480/488 (98)	488	0	YP_002332418.1
20	20837	21226	129	4,8/14,8	AGGAGTgagaaaataATG		phi12, gp46	129/129 (100)	129	3e-68	NP_803352.1
21	21219	21383	54	4.3/6.5	AGGAGAgactgaaaATG		phi12, gp47	54/54 (100)	54	2e-22	NP_803353.1
22	21417	21728	103	6.8/12.4	AATTTGaataaaGTG		phiSLT, gp58	99/99 (100)	99	1e-48	NP_075520.1
23	21819	22166	115	9.6/13.1	AAGAGTcaGTG	Holin	PVL, gp23	99/100 (99)	100	1e-50	NP_058462.1
24	22177	23631	484	9.3/53.8	AGGTGTtgaccaATG	Amidase	3A, ORF007	479/484 (98)	484	0	YP_240025.1
25	24012	24959	315	9.1/35.7	AGAAAGgaaATG	Panton-Valentine leukocidin chain S precursor	PVL, ORF027	312/312 (100)	312	0	NP_058465.1
26	24961	25938	325	9.1/36.9	AGGACAtaattgatATG	Panton-Valentine leukocidin subunit F	PVL, ORF028	325/325 (100)	325	0	NP_058466.1
27	27485	26280	401	9.9/47.5	AGGAGGgatgtaaaATG	Integrase	47, ORF011	396/401 (99)	401	0	YP_240030.1
28	27868	28218	116	9.1/14.0	AGAATTcGTG		phiIPLA35, gp2	116/116 (100)	207	5e-59	YP_002332365.1
29	29561	28641	306	5.8/35.2	AAGGGGctgattataATG	DNA polymerase III subunit epsilon	47, ORF013	306/306 (100)	306	4e-179	YP_240034.1
30	30191	29577	204	7.8/22.9	AGGAGGaaatttaaaATG	Transcriptional repressor	47, ORF020	204/204 (100)	204	2e-112	YP_240035.1
31	30396	30590	64	8.0/8.8	ATTCTGctttagcgATG		47, ORF060	64/64 (100)	75	4e-29	YP_240036.1
32	30616	31392	258	5.7/29.5	AGGAGGcataaacaaATG	Anti-repressor protein	47, ORF017	258/258 (100)	258	2e-149	YP_240038.1
33	31408	31626	72	4.6/8.4	AGGAGGacttaaaaATG		47, ORF065	72/72 (100)	72	3e-33	YP_240039.1
34	32310	32525	71	8.0/8.1	AGGAGCataaacaaATG		phiSLT, gp11	71/71 (100)	71	1e-32	NP_075474.1
35	32552	32815	87	9.0/10.4	AGGAAAagatagaaATG	DNA binding protein	3A, ORF048	84/87 (96)	87	1e-42	YP_239975.1
36	33068	33391	107	6.3/12.6	AGGAGTtattaatATG		phi2958PVL, gp15	107/107(100)	107	5e-55	YP_002267985.1
37	33406	33768	120	4.8/13.8	AGGAGGagttaatcaATG		tp-310-2, gp18	116/120 (96)	120	2e-60	YP_001429913.1
38	33765	34931	388	5.5/44.3	TGGAAGcgagaattaatgcATG		3A, ORF012	382/388 (98)	388	0	YP_239979.1

39	34912	35514	200	5.2/21.9	AGCAATctgctgaagATG		phi2958PVL, gp18	184/185 (99)	185	5e-102	YP_002267988.1
40	35582	37534	650	6.8/73.5	AGGTGTcaagaatttgagatttATG	DNA polymerase domain A	47, ORF003	640/650 (98)	653	0	YP_240050.1
41	37700	37885	61	4.4/7.4	AGGAAGtgatttaATG		29, ORF083	57/61 (93)	61	2e-24	YP_240593.1
42	37886	38242	118	10.1/13.9	AGGTGGaataaATG		71, ORF044	118/118 (100)	118	3e-62	YP_240440.1
43	38246	38488	80	6.5/9.6	AAGAAGtagatcATG		71, ORF065	77/80 (96)	80	4e-38	YP_240441.1
44	38493	38756	87	4.1/9.8	TTGAGGagATG		69, ORF060	84/84 (100)	85	2e-40	YP_239630.1
45	38725	38913	62	3.9/6.6	AAGTGGtctataaatATG		tp310-2, gp31	52/56 (92)	56	1e-22	YP_001429926.1
46	38906	39442	178	4.6/20.8	AGGTGGaacaggaaaATG		phi2958PVL, gp25	177/178 (99)	178	3e-99	YP_002267995.1
47	39548	39652	34	9.5/7.0	ACAGGGtaaaaatATG		69, ORF089	32/34 (94)	57	4e-09	YP_239633.1
48	39669	39905	78	5.1/8.9	AGGAGTgatgagaaGTG		phiPV83, gp31	68/78 (87)	78	2e-22	NP_597920.1
49	39930	40166	78	4.7/9.1	AAGAGGggagataataATG		phiSLT, gp32	78/78 (100)	78	2e-38	NP_075494.1
50	40150	40311	53	4.6/6.3	TACAGGagATG		phi2958PVL, gp27	53/53 (100)	53	1e-22	YP_002267997.1
51	40448	40579	43	6.2/5.2	AGTAGCatttctcataagATG		3A, ORF059	43/43 (100)	66	1e-16	YP_239993.1
52	40631	43078	815	5.2/94.4	AGGAGCcgaacATG	Virulence associated protein E	47, ORF002	800/815 (98)	815	0	YP_240066.1
53	43419	43709	96	9.5/11.3	AGGTGAatatATG		phiIPLA35, gp33	94/96 (97)	96	1e-49	YP_002332396.1
54	43690	44493	267	6.4/31.4	AGAATGgtagGTG	Protein SNF2, Helicase	3A, ORF009	262/265 (98)	455	7e-154	YP_239997.1
55	44946	45056	36	9.6/4.3	ACGACTattattcatcacatcATG	Protein SNF2, Helicase	phi2958PVL, gp31	36/36 (100)	423	4e-12	YP_002268001.1
56	45069	45506	145	9.4/17.0	TGGAGGtataagATG	Transcriptional regulator RinA	3A, ORF025	145/145 (100)	145	4e-79	YP_239999.1
57	45663	45977	104	8.9/12.6	AAGAGGtgtaagagATG	HNH endonuclease	47, ORF039	104/104 (100)	104	2e-53	YP_240072.1