

Genome of Staphylococcal Phage K: a New Lineage of *Myoviridae* Infecting Gram-Positive Bacteria with a Low G+C Content

S. O'Flaherty,^{1,2,3} A. Coffey,⁴ R. Edwards,⁵ W. Meaney,² G. F. Fitzgerald,^{3,6} and R. P. Ross^{1,6*}

*Dairy Products Research Centre*¹ and *Dairy Production Research Centre*,² Teagasc, Moorepark, Fermoy, Co. Cork, and *Department of Microbiology, University College Cork*,³ *Department of Biological Sciences, Cork Institute of Technology*,⁴ and *Alimentary Pharmabiotic Centre*,⁶ Cork, Ireland, and *University of Tennessee Health Sciences Center, Memphis, Tennessee 38163*⁵

Received 7 November 2003/Accepted 26 January 2004

Phage K is a polyvalent phage of the *Myoviridae* family which is active against a wide range of staphylococci. Phage genome sequencing revealed a linear DNA genome of 127,395 bp, which carries 118 putative open reading frames. The genome is organized in a modular form, encoding modules for lysis, structural proteins, DNA replication, and transcription. Interestingly, the structural module shows high homology to the structural module from *Listeria* phage A511, suggesting intergenus horizontal transfer. In addition, phage K exhibits the potential to encode proteins necessary for its own replisome, including DNA ligase, primase, helicase, polymerase, RNase H, and DNA binding proteins. Phage K has a complete absence of GATC sites, making it insensitive to restriction enzymes which cleave this sequence. Three introns (*lys-I1*, *pol-I2*, and *pol-I3*) encoding putative endonucleases were located in the genome. Two of these (*pol-I2* and *pol-I3*) were found to interrupt the DNA polymerase gene, while the other (*lys-I1*) interrupts the lysin gene. Two of the introns encode putative proteins with homology to HNH endonucleases, whereas the other encodes a 270-amino-acid protein which contains two zinc fingers (CX₂CX₂₂CX₂C and CX₂CX₂₃CX₂C). The availability of the genome of this highly virulent phage, which is active against infective staphylococci, should provide new insights into the biology and evolution of large broad-spectrum polyvalent phages.

Myoviridae are a diverse group of phages that are characterized by their contractile tails. *Myoviridae* are able to infect both gram-positive and gram-negative bacteria. The International Committee on Taxonomy of Viruses (ICTV) classification is so broad that the *Myoviridae* have been separated into several classes, i.e., T4-like viruses, P1-like viruses, P2-like viruses, Mu-like viruses, SP01-like viruses, PBS1-like viruses, and ϕ H-like viruses (42). These phage classes exhibit many differences in their lifestyles. For example, P1 lysogenizes as a plasmid, Mu is a transposon and does not undergo site-specific recombination like P2, and T4 is a lytic phage. Some of the phage infect gram-positive bacteria, some infect gram-negative bacteria, and some infect archaea (42). The genomes of these phages vary in size from at least 30,000 to 170,000 bp, and there is no molecular basis for these ICTV classifications. The phages are called *Myoviridae* merely because they look alike. In the post-genomic era, we need to move towards a classification system that takes into account the molecular characteristics that are shared between members of the different phage families. The data that we present below challenge the idea that these should all be members of the same family, using a molecular mechanism of categorizing the phage.

Phage K is a member of the family *Myoviridae* and has been the subject of previous preliminary studies which focused on the adsorption and infection processes, the effect of acridines on phage reproduction, morphology, and the nature of phage K DNA replication (17–19, 32–34). This bacteriophage has activity against a range of both coagulase-negative and coagu-

lase-positive staphylococci (37) and uses *N*-acetylglucosamine in cell wall teichoic acid for phage adsorption (10). Phage K is also identical to the phage named Au2 examined by Burnet and Lush (7), and restriction analysis in our laboratory revealed that it is indistinguishable from the polyvalent phage 812 (31).

Generally, phage genomes are organized in modular structures, with each module containing a set of genes which carry out a biological function (9). Phage evolution can occur by the exchange of modules between phages which have access to the same gene pool (5). The evolution of lactococcal and streptococcal phages of the family *Siphoviridae* has been studied extensively (6, 12–14, 25); however, this has not been the case with phages of the *Myoviridae*. As more *Myoviridae* genomes are sequenced and analyzed, it should become more evident that phage K may represent a new genus infecting gram-positive bacteria with a low G+C content.

A notable feature of some large phage genomes is the presence of introns. Phage-carried group I introns are typically found interrupting genes involved in DNA metabolism, in contrast to chromosomally located introns, which are generally found interrupting structural RNA genes (15).

In this communication we analyze the phage K genome with respect to its organization, regulation, and evolution, and we also confirm experimentally the presence of three introns and determine their insertion sites. Unusually, the DNA polymerase gene is interrupted by two distinct introns, and another is found interrupting the lysin gene. Each of the three introns also encodes putative endonucleases. Another surprising feature of the genome is that the structural module bears high similarity to the structural module from *Listeria* phage A511. The elucidation of the genome of this phage will provide an

* Corresponding author. Mailing address: Teagasc, Dairy Products Research Centre, Moorepark, Fermoy, Co. Cork, Ireland. Phone: 353-25-42229. Fax: 353-25-42340. E-mail: pross@moorepark.teagasc.ie.

insight into phage classification, genome organization, and intron invasion in a phage with considerable potential in phage therapy applications.

MATERIALS AND METHODS

Phage propagation. Phage K was obtained from the American Type Culture Collection (ATCC 19685-B1). Phage K was routinely propagated on *Staphylococcus aureus* DPC5246 in brain heart infusion broth. Concentrated phage K preparations were obtained by CsCl density gradient centrifugation following polyethylene glycol 8000 precipitation of brain heart infusion lysates. The protocols used were in accordance with those described previously (38). Phage preparations were dialyzed in 10 mM sodium phosphate buffer (pH 7) and filter sterilized prior to use.

DNA sequencing and sequence analysis. DNA sequencing was performed to 12-fold redundancy with the LI-COR 4200L automated DNA sequencer and dye primer chemistry with a cycle sequencing protocol (MWG-Biotech AG, Ebersberg, Germany). The sequence assembly was checked for correctness by comparison with restriction data obtained by using several restriction enzymes. The data indicated that while the assembly was correct, sequences from the ends were missing. Despite attempts at ligation-mediated PCR, the sequences of these ends could not be determined. Sequence analysis was performed as described earlier (30), using DNASTar (Madison, Wis.) software. Open reading frames (ORFs) preceded by a Shine-Dalgarno sequence at an appropriate distance (3 to 18 bp) from the initiation codon and coding for proteins of at least 100 amino acids were considered putative genes. ORFs were identified, translated, and searched against the protein database by using the BLAST (1) and PSI-BLAST (2) algorithms. Clustal alignments of sequences were performed with the DNASTar software.

Preparation of phage DNA. Phage K DNA was extracted from a CsCl-purified (see above) phage stock solution (300 µl; >10⁹ PFU/ml) as follows. DNA was incubated at 37°C for 30 min with RNase (10 mg/ml) (Sigma-Aldrich, Dublin, Ireland), the mixture was adjusted to 1% sodium dodecyl sulfate (SDS), and 6 µl of EDTA (0.5 M) and 6 µl of proteinase K (10 mg/ml) (Sigma-Aldrich) were added, followed by incubation for 1 h at 37°C. All proteinaceous material was removed by using phenol-chloroform-isoamyl alcohol (25:24:1) (Sigma-Aldrich), and DNA was extracted and precipitated as described previously (38).

Preparation of phage proteins, SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and N-terminal sequencing. A concentrated CsCl phage K preparation (1 ml) was precipitated by adding 4 volumes of ice-cold acetone. Samples were centrifuged at 1,600 × g for 20 min, the supernatant was discarded, and the pellet was allowed to air dry. The pellet was then resuspended in 100 µl of sample buffer (2 ml of 10% SDS, 0.2 ml of 0.5% bromophenol blue, 1.25 ml of 0.5 M Tris HCl [pH 6.8], and 2.5 ml of glycerol, made up to 9.5 ml with deionized water; 50 µl of β-mercaptoethanol was added to 950 µl of this solution prior to use). Samples were boiled for 3 min before being loaded onto polyacrylamide gels. Proteins were electrotransferred from polyacrylamide gels onto polyvinylidene difluoride membranes (Bio-Rad Corp., Richmond, Calif.), in buffer A (25 mM Tris, 192 mM glycine, 20% methanol [pH 8.3]), using a Trans-Blot cell (Bio-Rad, Alpha Technologies, Dublin, Ireland), according to the manufacturer's instructions. Proteins were stained with Coomassie brilliant blue R250, cut out of the membrane, and sequenced on a Beckman LF 3000 microsequencer (Molecular Biology Unit, University of Newcastle upon Tyne). Database searches were performed with the program BLASTP and against the phage K genome.

Total RNA isolation. Staphylococcal strain DPC 5246 was grown at 37°C to an optical density at 600 nm of 0.50 and then infected with phage K at a multiplicity of infection of >10. Twenty-milliliter samples were removed, pelleted, and frozen immediately in a -80°C ethanol bath at 10, 20, and 30 min after phage infection. Total RNA was isolated by resuspending each pellet in 2 ml of TRI reagent (Sigma-Aldrich) and transferring the mixture to two 2.0-ml screw-cap microcentrifuge tubes containing 0.8 g of acid-washed beads (106 µm in diameter) (Sigma-Aldrich). The slurry was sheared with a Mini-Beadbeater-8 cell disrupter (Strattech Scientific Ltd., Bedfordshire, United Kingdom) for three 1-min cycles (and chilled on ice between cycles). The remainder of the RNA isolation procedure was carried out according to the manufacturer's instructions. Residual DNA was removed by using a DNA-free kit (Ambion Ltd., Cambridgeshire, United Kingdom). Standard procedures to minimize RNase contamination were used (38).

Synthesis of cDNA. Total RNA was isolated as described above, 30 min after phage infection. First-strand cDNA synthesis was performed as follows. Two micrograms of RNA was incubated with 3 µg of random primer (Invitrogen, Paisley, United Kingdom.) per ml in a final volume of 11 µl for 10 min at 70°C. Samples were snap frozen in a -80°C ethanol bath and briefly spun by centrif-

ugation. The primer extension reaction was carried out at 42°C overnight following the addition of 1 µl of a 10 mM deoxynucleoside triphosphate master mix (Bioline Ltd., London, United Kingdom), 1 µl (200 U) of Superscript II reverse transcriptase, 2 µl (0.1 M) dithiothreitol, and 4 µl of 5× first-strand buffer (Invitrogen). The primer extension reaction was inactivated by heating the samples to 70°C for 15 min.

Determination of intron splice junctions. PCR amplifications were performed in a 100-µl reaction mix with 2 µg of cDNA-DNA, 50 mM MgCl₂, 100 pmol of each primer, and 1U of *Taq* polymerase (Bioline Ltd.). The PCR primers used in this study were purchased from MWG-Biotech UK, Ltd. (Milton Keynes, United Kingdom). The sequences of relevant primers used for PCR are as follows: LysF, 5' AAC TGC AGT ATT ACG GAG GAT TTA AAA TGG CTA AGG AC TCA AGC 3'; LysR, 5' GCT CTG ACT ATT TGA ATA CTC CCC AGG C 3'; PolF, 5' AAC TGC AGA GGA GGA ATT AAA TGA AAG TAT TAA TC 3'; and PolR, 5' GCT CTA GAT ATT AAA TTT CTT GAT AAA TAT G 3'. Negative control PCRs with no template DNA were also performed. Samples were subjected to denaturation (94°C for 1 min), annealing (50°C for 1 min), and elongation (72°C for 1 min) for 35 cycles with a Hybad PCR express unit. Amplified DNA fragments from cDNA were purified from a 1% agarose gel by using a QIAquick gel purification kit (Qiagen, West Sussex, United Kingdom), cloned in the pCR2.1-TOPO vector, and transformed in One ShotTOP 10 chemically competent *Escherichia coli* by using a TOPO TA cloning kit according to the instructions of the manufacturer (Invitrogen). Primers from pCR2.1-TOPO (M13F and M13R) were used to determine the splice junctions for the intron interrupting the lysin gene. Internal primers that correspond to sequence surrounding the second two introns interrupting the DNA polymerase gene were designed as follows: Int1F, 5' GAT ATT ACC GCA TGG ACT TA 3'; Int1R, 5' AAC ATC ATA CTC TTT CTT AGC 3'; Int2F, 5' GCT AAT GTT AAA GAA GCA GAC 3'; and Int2R, 5' ACT CAT GTA CAT TG T CAA TAG 3'. Sequencing reactions were performed twice on two different pCR2.1-TOPO clones for each intron by Lark Technologies Inc., Essex, United Kingdom.

Phylogenetic analysis. The phylogenetic analysis was performed essentially as described previously (36). Briefly, every ORF in phage K was compared to every other ORF from the complete phage genome database maintained by Rohwer and Edwards. The database currently contains 375 phage genomes, including both free-living phages and prophages. All similar proteins are aligned, and a protein distance matrix is calculated from each alignment. The matrices are averaged, and the tree is calculated from the average protein distance matrix (R. Edwards and F. Rohwer, unpublished data).

Nucleotide sequence accession number. The genome sequence of phage K has been deposited in the GenBank database under accession number AY176327.

RESULTS AND DISCUSSION

General features of the genome of phage K. The phage K genome is presented as a 127,395-bp contiguous sequence of linear double-stranded DNA which carries at least 118 putative ORFs, which were capable of encoding peptides of at least 100 amino acids in all six reading frames preceded by a potential Shine-Dalgarno sequence at a distance of at least 3 to 18 bp from a start codon (AUG, GUG, or UUG) (Table 1). The majority of the ORFs (112) initiate translation with the AUG start codon, whereas only 5 (ORFs 38, 40, 41, 42, and 96) initiate translation with the UUG start codon and 1 (ORF 63) initiates with a GUG start codon (Table 1). Bioinformatic analysis of ORFs revealed that the majority exhibited low identities with proteins from the database (Table 1), which often is the case with new genomes. We suspect that the phage has terminally redundant ends, based on the following lines of evidence. First, it has previously been shown that the genome is linear, is not circularly permuted, and does not possess cohesive ends (32). Second, the extreme ends of this phage could not be sequenced, which may be due to the physical nature of the ends of the genome. The genome can be divided into two distinct regions, which are divergently transcribed as indicated by bioinformatic analysis. In this respect, of the 118 ORFs, 85 are transcribed in one orientation and 33 are tran-

TABLE 1. General features of putative ORFs from phage K with best matches in the database

ORF ^a	bp		Protein size (kDa)	Representative similarity to proteins in database	e value	% Identity (% match of length)	Accession no.
	Start	Stop					
1*	2934	2449	18.4	Hypothetical protein			
2*	3358	2927	16.7	Hypothetical protein			
3	3914	3372	21.5	Hypothetical protein			
4*	4414	3926	19.5	<i>Pseudomonas aeruginosa</i> conserved hypothetical protein	3e-24	39 (40)	NP_253270
5*	4825	4427	16.1	Hypothetical protein			
6*	5529	4822	27.7	<i>Agrobacterium tumefaciens</i> serine/threonine protein phosphatase I	2e-13	29 (26)	NP_534038
7*‡	6183	5629	21.2	Hypothetical protein			
8*‡	8050	7502	21.9	Hypothetical protein			
9*	9194	8457	28.7	Hypothetical protein			
10*	10003	9614	15.2	Hypothetical protein			
11*‡	10798	10316	18.8	Hypothetical protein			
12*‡	11390	10848	20.4	Hypothetical protein			
13*	11923	11390	20.7	Hypothetical protein			
14*	13213	12368	31.7	Hypothetical protein			
15*	13809	13225	21.9	<i>Mycoplasma penetrans</i> AAA family ATPase	5e-20	40 (32)	NP_757966
16*	15233	14817	16	Hypothetical protein			
17*‡	15669	15367	11.3	<i>Enterococcus hirae</i> ArpR	6e-18	45 (45)	CAA90708
18*	18110	16062	79.8	<i>Enterococcus faecalis</i> plasmid pAD1 unknown protein	3e-04	28 (6)	AAL59463
19*‡	19226	18648	21.4	Hypothetical protein			
20*	19845	19219	23.8	Hypothetical protein			
21*	20734	19838	35	<i>Clostridium acetobutylicum</i> homolog of eukaryotic DNA ligase III	3e-05	27 (14)	NP_347388
22*	21767	21027	28.6	<i>Neisseria meningitidis</i> PhoH-related protein	1e-14	36 (38)	NP_273886
23*	22433	21819	23	<i>Staphylococcus aureus</i> temperate phage phiSLT ORF 636	0.51	26 (18)	NP_075517
24*	22874	22449	15.8	<i>Clostridium perfringens</i> ribonuclease H1	7e-14	34 (34)	NP_562366
25*	23719	23078	24.6	Hypothetical protein			
26*‡	24971	24279	24.8	<i>S. aureus</i> hypothetical protein	7e-07	29 (25)	NP_372619
27*	25793	25158	24.8	<i>Streptococcus pyogenes</i> hypothetical phage protein	2e-22	36 (37)	NP_607354
28*‡	26651	25860	29.3	<i>Synechocystis</i> sp., unknown protein	1e-15	24 (20)	NP_440390
29*	26959	26651	12.2	Hypothetical protein			
30‡	27701	27072	23.1	<i>S. aureus</i> autolysin (<i>N</i> -acetylmuramoyl-L-alanine amidase)	1e-54	52 (52)	P24556
31*	28472	27972	19.2	<i>Bacteroides thetaiotaomicron</i> endonuclease	1e-12	39 (33)	NP_810939
32*	29435	28632	29.8	<i>S. aureus</i> phage Twort <i>N</i> -acetylmuramoyl-L-alanine amidase	1e-38	40 (37)	CAA69021
33*	29938	29435	18.1	<i>S. aureus</i> phage Twort holTW (holin)	9e-38	61 (53)	CAA69020
34*	34507	34833	12.3	Hypothetical protein			
35*	34848	36665	70.2	<i>Ralstonia solanacearum</i> probable bacteriophage-related protein	0.002	25 (8)	NP_518974
36*	36658	37479	30.7	Hypothetical protein			
37*‡	37636	38115	18.5	Hypothetical protein			
38*‡‡	38157	39350	43.6	<i>Bacillus cereus</i> cell surface protein	3e-05	28 (24)	NP_346725
39*	39435	39776	12.8	Hypothetical protein			
40†	39794	40165	14.5	Hypothetical protein			
41*‡‡	40169	41860	64.0	<i>Escherichia coli</i> bacteriophage P27 putative portal protein	0.004	22 (10)	NP_543090
42*†	42054	42827	28.6	<i>Listeria monocytogenes</i> bacteriophage A511 ORF 1	2e-62	54 (63)	CAA62538
43*	42846	43796	35.7	Hypothetical protein			
44‡	43912	45303	51.2	<i>S. aureus</i> phage Twort capsid protein	0.0	82 (81)	AAQ6728
				<i>L. monocytogenes</i> bacteriophage A511 major capsid protein	c-174	66 (67)	CAA62540
45*	45704	46612	31.2	<i>L. monocytogenes</i> bacteriophage A511 ORF 3	e-104	60 (36)	CAA62541
46*	46626	47504	33.7	<i>L. monocytogenes</i> bacteriophage A511 ORF 4	8e-56	26 (9)	CAA62542
47*	47504	48124	23.8	<i>L. monocytogenes</i> bacteriophage A511 ORF 5	2e-25	46 (35)	CAA62543
48*	48143	48979	31.8	<i>L. monocytogenes</i> bacteriophage A511 ORF 6	6e-53	38 (39)	CAA62544
49*	49223	50986	54.5	<i>L. monocytogenes</i> bacteriophage A511 major tail sheath protein	0.0	57 (58)	CAA62546
50*‡	51059	51487	15.9	<i>S. aureus</i> Phage twort unknown protein	3e-62	85 (84)	AF132670
				<i>L. monocytogenes</i> bacteriophage A511 ORF 8	5e-48	68 (67)	CAA62547
51*	51767	52225	18.1	Hypothetical protein			
52*	52514	52825	12.3	Hypothetical protein			
53*	52957	53415	18.1	<i>L. monocytogenes</i> bacteriophage A511 ORF 9	1e-23	39 (39)	CAA62548
54*‡	53459	53995	20.9	Hypothetical protein			
55	54051	58106	144	<i>Lactococcus lactis</i> bacteriophage CP-1 ORF 18	4e-08	32 (3)	NP_047298
56	58185	60611	91.2	<i>Staphylococcus epidermidis</i> secretory antigen SsaA-like protein	4e-32	49 (13)	NP_765044
57*	60625	61512	34.6	Hypothetical protein			
58*	61512	64058	96.1	<i>Bacillus subtilis</i> glycerophosphoryl diester phosphodiesterase	1e-41	38 (11)	NP_388095
59*	64165	64956	29.3	Hypothetical protein			
60*	64956	65480	20	Hypothetical protein			
61*	65480	66184	26.6	<i>Salmonella enterica</i> serovar Typhi putative bacteriophage baseplate protein	0.020	45 (7)	NP_456045
62	66199	67245	39.2	<i>Clostridium tetani</i> phage-like element PBX protein XkdT	5e-06	24 (43)	NP_782676
63†	67266	70325	116.2	Hypothetical protein			
64	70436	70957	19.2	Hypothetical protein			
65*‡	70978	74436	129	<i>Dichelobacter nodosus</i> YrIC protein	9e-33	27 (10)	T17382
66*	74644	76566	72.5	<i>S. aureus</i> SLT ORF 488-like protein	0.021	29 (4)	AAL82326
67*	76589	76963	14.6	Hypothetical protein			
68*‡	76970	78346	50.4	<i>L. monocytogenes</i> lmo1188	0.034	24 (8)	NP_464713

Continued on following page

TABLE 1—Continued

ORF ^a	bp		Protein size (kDa)	Representative similarity to proteins in database	e value	% Identity (% match of length)	Accession no.
	Start	Stop					
69*	78438	80186	37.2	<i>S. enterica</i> serovar Typhi putative helicase	2e-36	33 (16)	NP_569550
70*	80198	81811	63.2	<i>E. faecalis</i> putative Rep protein	0.031	40 (5)	CAC29157
71*	81804	83246	54.6	<i>S. enterica</i> serovar Typhi putative DNA helicase	4e-11	21 (15)	NP_569513
72	83325	84362	40.1	<i>L. monocytogenes</i> protein similar to putative exonuclease SbcD	1e-09	23 (19)	NP_465171
73*	84362	84739	14.9	Hypothetical protein			
74*	84739	86658	73.4	<i>Yersinia pestis</i> plasmid pMT1 probable exonuclease	2e-21	34 (9)	T14925
75*	86658	87254	23.2	Hypothetical protein			
76*	87269	88336	40.9	<i>Helicobacter pylori</i> DNA primase	5e-07	24 (17)	NP_206814
77*	88741	89193	17	Hypothetical protein			
78	89180	89788	23.6	Bacteriophage T5 D14 protein	1e-05	26 (17)	O48499
79*	89805	90197	14.7	<i>Staphylococcus epidermidis</i> NrdI protein	3e-13	35 (50)	NP_764067
80	90212	92326	80.2	<i>S. aureus</i> phage Twort large ribonucleotide reductase subunit	0.0	72 (73)	AAM00816
81	92340	93389	40.4	<i>S. aureus</i> ribonucleoside reductase minor subunit	3e-84	53 (48)	NP_371256
82*	93407	93736	12.4	Hypothetical protein			
83	93720	94040	12	<i>Alicyclobacillus acidocaldarius</i> thioredoxin	0.013	33 (31)	P80579
84	94247	94843	23.5	Hypothetical protein			
85*‡	94853	95158	11.9	<i>Chlamydomophila pneumoniae</i> integration host factor alpha	0.003	26 (26)	NP_224616
86*	95234	96106	33.2	<i>B. subtilis</i> bacteriophage SPO1 DNA polymerase	0.054	25 (12)	P30314
87*	96314	96784	18.6	Hypothetical protein			
88	96920	98263	52.8	<i>Thermotoga maritima</i> DNA-directed DNA polymerase I	1e-06	22 (13)	NP_229419
89	98429	99238	31.2	Bacteriophage bastille HNH endonuclease I-BasI	6e-22	40 (40)	AAO93095
90	99472	100332	32.9	<i>B. subtilis</i> phage SPO1 DNA-directed DNA polymerase	2e-31	36 (31)	JC1269
91*	100660	101142	18.9	Hypothetical protein			
92*‡	101229	102500	46.9	Hypothetical protein			
93*	102560	103816	46.8	<i>T. maritima</i> DNA repair protein	2e-27	30 (24)	NP_229655
94*	104160	104822	26.6	<i>B. subtilis</i> bacteriophage SPO1 RNA polymerase sigma GP34 factor	3e-05	32 (17)	P06227
95*	104950	105582	23.2	Hypothetical protein			
96*‡†	105605	106117	17.8	<i>S. aureus</i> temperate phage phiSLT major tail protein	4e-14	40 (49)	NP_075510
97*	106719	107474	29.2	Hypothetical protein			
98	107467	108717	47.5	Hypothetical protein			
99	108731	109099	14	Hypothetical protein			
100*‡	109086	109397	12	Hypothetical protein			
101*	109990	110757	30	<i>S. aureus</i> hypothetical protein	0.001	36 (16)	CAB60746
102*	110735	111181	17.3	Hypothetical protein			
103*	112416	113417	28.4	Hypothetical protein			
104*	113165	113623	17.9	Hypothetical protein			
105*	113688	114131	17.5	Hypothetical protein			
106*‡	114148	114852	27.4	Hypothetical protein			
107*‡	114914	115312	15.4	Hypothetical protein			
108*‡	115706	116176	17.9	Hypothetical protein			
109*	117041	117349	12	Hypothetical protein			
110*	117346	118254	35.2	<i>Pyrococcus horikoshii</i> ribose-phosphate pyrophosphokinase	8e-09	27 (24)	NP_143754
111*‡	118272	119741	56.1	<i>Haemophilus ducreyi</i> putative nicotinamide phosphoribosyl transferase	e-141	52 (52)	AAK06405
112*	120082	120477	15.5	Hypothetical protein			
113	121090	121401	11.7	Hypothetical protein			
114*	121407	121706	11.6	Hypothetical protein			
115*	122253	122606	13.9	Hypothetical protein			
116*	122625	123011	15.6	Hypothetical protein			
117*	124587	124913	12.7	Hypothetical protein			
118*	125811	126158	13.7	Hypothetical protein			

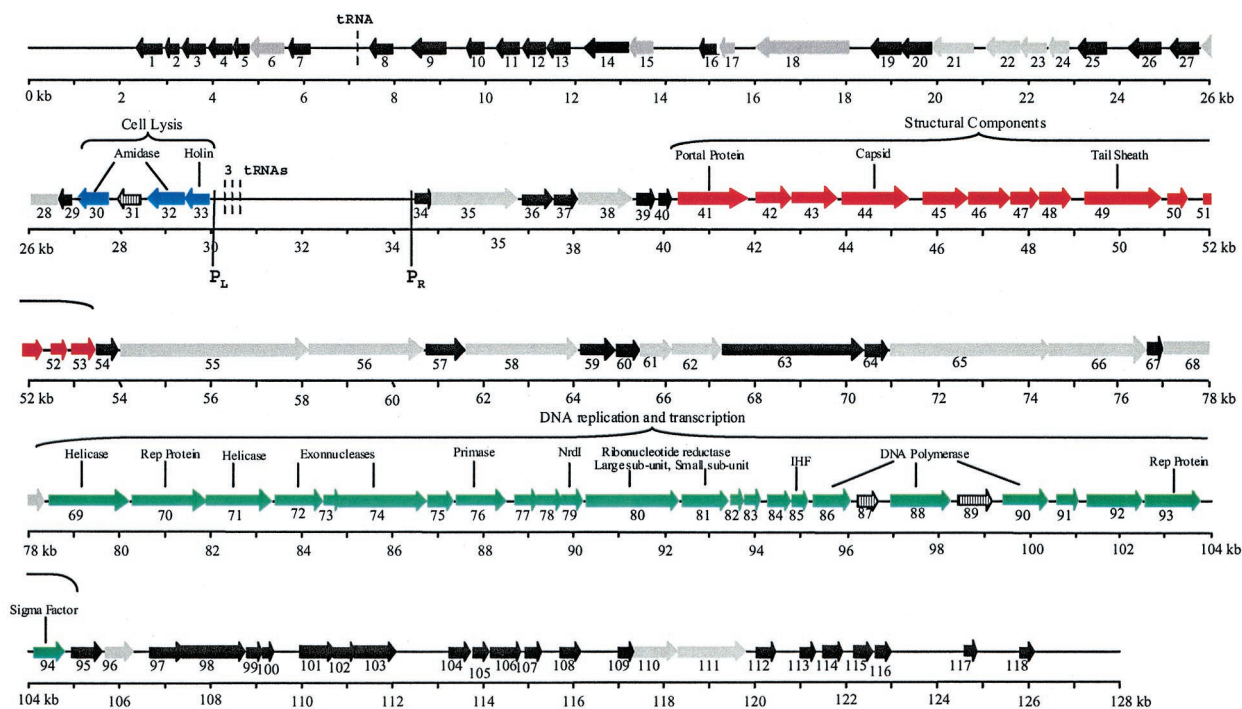
^a †, non-AUG start codons; ‡, presence of putative rho-independent terminators downstream of the ORF; *, good ribosome binding sites upstream of the start codon.

scribed in the opposite orientation, with all of the latter grouped together in the first 30 kb, as illustrated in Fig. 1. Phage K has a G+C content of 30.6%, which is significantly lower than that generally associated with the staphylococcal bacterial genome (22).

Lack of restriction sites for host-encoded endonucleases. Analysis of the phage genome revealed no significant homology to any DNA methylases (Table 1). However, an interesting feature of the phage K genome is that it completely lacks GATC sites and so should not be restricted by Sau3A1, BamHI, PvuI, and DpnI. A paucity of restriction sites has also been observed in phages which infect lactococci (29). Phage K and a second staphylococcal phage (phage CS1, isolated recently in our laboratory) were subjected to digestion with the

restriction enzyme Sau3A1. Phage CS1 was digested by Sau3A1, yielding numerous fragments, whereas, as expected, phage K was not digested (data not shown). *S. aureus* is known to encode a Sau3A1 restriction-modification system which recognizes the 5'-GATC-3' DNA sequence (40). A second site-specific endonuclease from *S. aureus*, which recognizes the sequence 5'-GGNCC-3', was identified (41). Interestingly, there is only one 5'-GGTCC-3' site in the genome of phage K; none of the other possible recognition site combinations (5'-GGGCC-3', 5'-GGACC-3', or 5'-GGCCC-3') are present. Hence, this phage appears to have a very efficient mechanism of counter defense against these specific endonucleases.

Phage K has its genes arranged in modules. Temperate staphylococcal phages are generally organized in a modular



P_L 29,996 **ACCGACCTACTGTTATATTATTGTTAGAAATAAATATAATAGAAAGGTCGGTTTTTTTAATG** 29,935
P_R 34,448 **AATTAACAAGAAAAAGTTAGAAGAAGAGGATACAAGAAAAATATATAGCTGATGGGTTTATG** 34,509

FIG. 1. ORF organization of phage K. ORFs 1 to 118 are indicated by arrows; the numbering corresponds to that in Table 1. Blue arrows, putative lysis module; red arrows, structural module; green arrows, DNA replication and transcription module; grey arrows, proteins with a putative function; black arrows, hypothetical proteins. Arrows with black vertical lines indicate three intron-carried ORFs. Arrows are roughly drawn to scale. Vertical lines mark two putative promoters. L and R, direction of transcription (left or right). Start codons and ribosome binding sites are indicated in boldface, and putative -10 and -35 sites are underlined. Dashed vertical lines represent the positions of four putative tRNAs.

form (20) which include modules for lysogeny, DNA replication, transcriptional regulation, packaging, structural proteins, and lysis. The organization of these temperate staphylococcal phage genomes is similar to that of temperate streptococcal phage genomes (6). Phage K also appears to have its genes arranged in modules, but the order differs from that for the temperate phages and the two sequenced lytic phages, which have their lysis module embedded in the structural region (43). The modules of phage K are not as well defined as those of the temperate staphylococcal and streptococcal phages; for example, there is a lack of intergenic regions between the structural and DNA replication and transcription modules. Putative rho-independent terminators were identified (Table 1) by using the TransTerm program (16). Three further terminators were located upstream of ORFs 1, 56, and 118 on the strand divergent to the ORFs. These terminators are characterized by a stem-loop in the mRNA followed by a U-rich sequence and allow for a punctuation of the 3' ends of multicistronic mRNA.

Promoters and tRNAs are located in an intergenic region. The 4.5-kb region between the divergently translated ORFs 33 and 34 does not carry ORFs according to the criteria employed in this study. However, two putative promoters are located in this region (Fig. 1).

Interestingly, three regions resembling tRNAs are located in

the 4.5-kb intergenic region (from bp 30,600 to 30,370 [Fig. 1]). These were identified by using the tRNA scan SE program (24) and encode Asp-tRNA, Phe-tRNA, and a pseudo-tRNA gene. A fourth tRNA gene (bp 7,222 to 7151 [Fig. 1]), encoding Met-tRNA, was located in a noncoding region between ORFs 7 and 8. These tRNA genes are common in large phages such as coliphage T4 (28), vibriophage KVP40 (27), and *Pseudomonas aeruginosa* phage phi KZ (26), which are three sequenced lytic *Myoviridae*, each of which also has tRNA genes located intergenically.

Taxonomy and comparative genomics. The need for a genome-based taxonomy tree has recently been identified (36). After studying 3,981 proteins of 105 genomes, no single gene that could be used as a basis for a classification system was found in all phages. Instead, a taxonomic system was based on a predicted phage proteome. The current phage proteome taxonomy is based on both complete phage genomes and prophages identified from within bacterial genomes (8). The database consists of 16,260 proteins from 375 genomes. Members of the ICTV family *Myoviridae* were grouped together in this system, with the exception of the T4 and P4 coliphages. These two phages represent their own groups in the proteomic tree due to the fact that they are the only sequenced representatives. Likewise, phage K does not fall within a defined group,

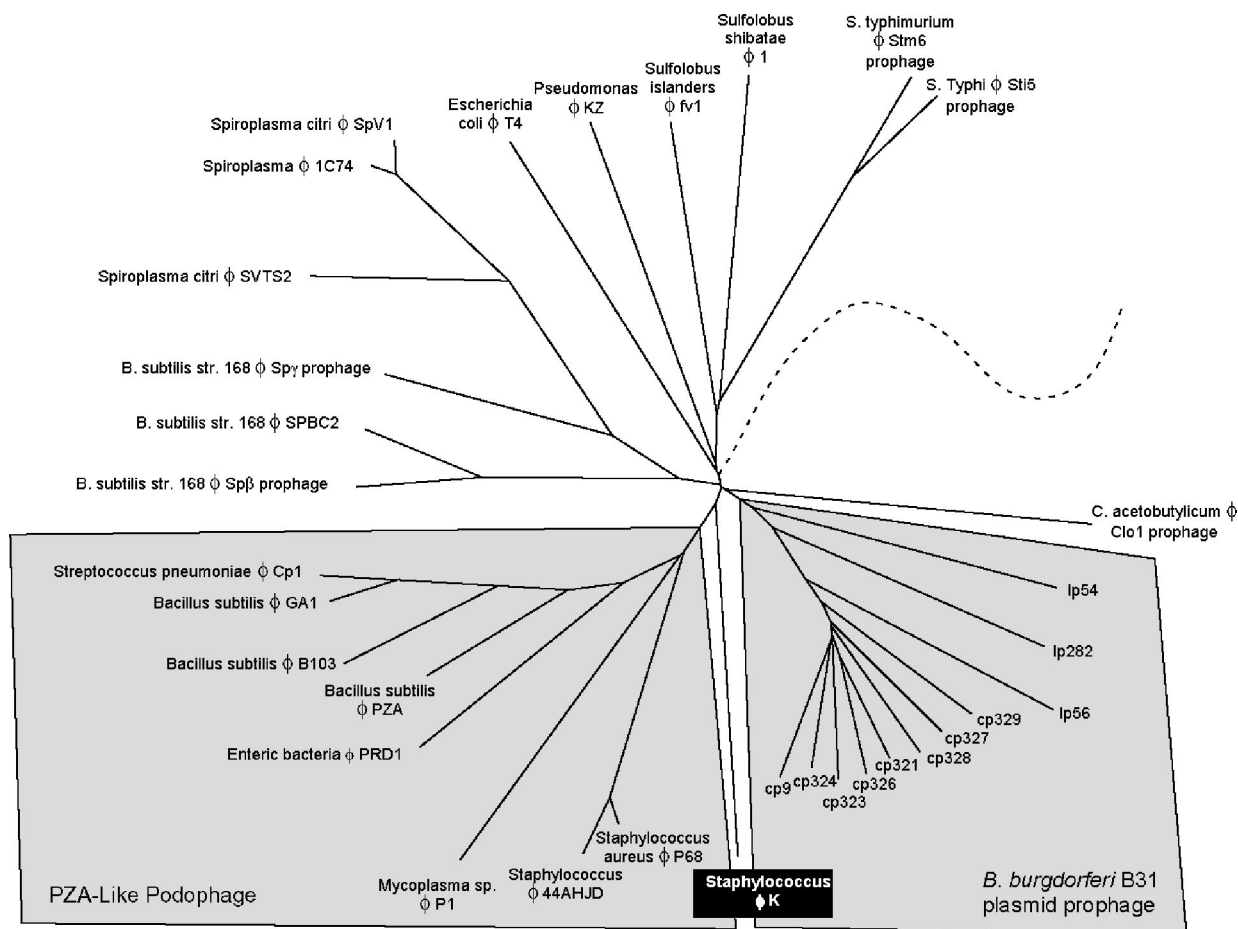


FIG. 2. Section of the phage proteomic tree illustrating the relationship between phage K and other sequenced *Myoviridae*. The tree is based on 375 sequenced phage genomes and prophages. Only the section of the tree corresponding to phage K is shown for clarity. The tree joins the remainder of the phages at the dashed line. The PZA-like *Podoviridae* (Podophage) are highlighted for correlation to reference 36. Phage K is weakly related to both the PZA-like *Podoviridae* and the *Borrelia burgdorferi* plasmid prophage, which form unique taxonomic groups.

confirming that it is the founding member of a new taxonomic group and that the *Myoviridae* are more diverse than their visual characterizations suggest.

Phage K falls closest to the PZA-like *Podoviridae* (Fig. 2). Figure 2 does not consider morphology or genome size but was generated by using a molecular characterization of the similarities between the phages. The weak similarities that currently bring phage K together with this group are mainly in the structural and replication modules; however, phage K is clearly in a taxonomic group of its own. The disparate locations of both phage T4 and phage K in Fig. 2 underscore the differences between the ICTV characterization of phages and the molecular characterization of phages. Indeed, we anticipate that as more phages are added to this tree, phage K will no longer fall on its own. Phage K and as-yet-unsequenced phages, such as those described by Jarvis et al. (21) and phage A511 (see below), could represent a new lineage of *Myoviridae* infecting gram-positive bacteria.

The lysis module is located in the first divergently transcribed 30 kb. The lysis module (ORFs 30 to 33) is located at the end of the first 30 kb, where all ORFs are divergently transcribed in relation to the rest of the genome (Fig. 1). ORF

33 encodes a putative holin of 167 amino acids (18.1 kDa) whose stop codon overlaps by 1 bp in a different reading frame, ORF 32. Both ORFs 33 and 32 have recognizable ribosome binding sites (Table 1). The putative holin of phage K exhibited 61% identity with a holin from phage Twort (Table 1) and probably functions by generating pores in the bacterial cell membrane. The lysin (spliced products of ORFs 30 and 32 [see below]) contains the recently described CHAP domain, which is characterized by three conserved motifs (3, 35). The putative *N*-acetylmuramoyl-L-alanine amidase domain is located in the center of the protein (residues 204 to 335), and a second amidase domain is located in the N terminus (residues 45 to 142).

Other putative ORFs within this 30-kb region include ORF 6, which encodes a 235-amino-acid (27.7-kDa) protein with an incomplete protein phosphate domain; ORF 15, which encodes a protein exhibiting 40% identity with an ATPase from the AAA family (39); and ORF 22, which encodes a 246-amino-acid (28.6-kDa) putative ATPase (Table 1).

Phage K may encode its own replisome and sigma factors. As is evident from Table 1, the identity scores for ORFs of phage K are very low, which is typical of new phage genomes

and makes bioinformatic interpretation difficult. Therefore, it is important to note that the identities of the ORFs discussed below are around 30%, with the exception of the subunits of the ribonucleotide reductase gene (ORFs 80 and 81 [Table 1]). Phage K has the potential to encode most of the proteins required for its own replisome, viz., DNA ligase (ORF 21), primase (ORF 76), helicase (ORF 69 and 71), polymerase (spliced products of ORFs 86, 88, and 90), RNase H (ORF 24), and DNA binding proteins (ORFs 17 and 85) (Fig. 1 and Table 1). Further ORFs include those encoding two exonucleases (ORFs 72 and 74), an integration host factor (ORF 85), enzymes required for nucleotide metabolism (ORFs 79, 80, and 81), and a thioredoxin protein (ORF 83), which could function in posttranslational modification or act as a chaperone (Fig. 1 and Table 1). Indeed, of the 52 ORFs assigned a putative function, approximately one-third are involved in DNA replication, metabolism, and repair. The majority of these proteins exhibit homology to bacterial but not phage proteins. Hence, phage K has an advantage in that it can potentially replicate its DNA without too much reliance on host functions. This may suggest that phage K has evolved to a broader host range. Along with T4 and the T4-related phage KVP40 (27, 28), phage K is one of the few examples of large phage genomes that encode so many DNA replication proteins.

When phage enter the host bacterial cell, they take control of many of the host proteins to use to their advantage, one of these being RNA polymerase. Phage K carries a putative sigma-like factor ORF (ORF 94), which encodes a protein of 220 amino acids (Table 1). When ORF 94 was scanned against the genome of phage K to determine if it shared homology with any of the unknown proteins, none was found. Reverse transcription-PCR analysis indicated that this protein is expressed at the same levels 10, 20, and 30 min after phage infection (data not shown). This sigma factor (ORF 94) could function to modify the host core RNA polymerase to recognize phage promoter regions, thereby regulating gene expression to express phage genes rather than host genes.

Introns with ORFs interrupting genes with crucial enzymatic functions. Analysis of the genome revealed that both the putative polymerase and lysin genetic determinants contained intron-like sequences. Indeed, the polymerase gene contained two such putative structures (*pol-I2* and *pol-I3*), each encoding endonucleases (ORF 87 [I-KsaII] and ORF 89 [I-KsaIII], respectively) (Fig. 3A). In contrast, the lysin gene contained one intron-like sequence (*lys-I1*), which also encodes a distinct endonuclease (ORF 31 [I-KsaI]) (Fig. 3B). Both I-KsaI and I-KsaIII exhibit homology to HNH endonucleases (Table 1) and contain an HNH motif. Interestingly I-KsaI also contains an intron-encoded nuclease repeat motif at the C-terminal end (data not shown). The functions of the nuclease repeats are unknown but could be involved in DNA binding via the helix-turn-helix motif (residues 116 to 164). I-KsaII exhibited no significant homology to any protein in the database. Closer examination of I-KsaII revealed the existence of two potential zinc binding motifs (CX₂CX₂₃CX₂C and CX₂C₂₂CX₂C), and thus it may belong to a subfamily of HNH endonucleases containing a zinc binding motif (11).

In order to confirm that these three distinct DNA regions represented introns, cDNA from each was sequenced. The results demonstrated that the polymerase gene had introns

of 775 and 1,082 bp (*pol-I2* and *pol-I3*, respectively) (Fig. 3A), whereas the lysin gene had an intron of 878 bp (*lys-I1*) (Fig. 3B). Sequence homology between the three introns (minus their corresponding endonuclease) indicates that they are unrelated. The alignment of phage K lysin with five similar sequences from the database revealed that intron *lys-I1* interrupted the lysin gene in the putative *N*-acetylmuramoyl-L-alanine amidase domain. The alignment of phage K translated polymerase likewise revealed that *pol-I2* is inserted into the N-terminal part of the protein, more specifically near an incomplete putative 3'-5' exonuclease activity domain, whereas *pol-I3* is inserted in the center of the putative polymerase A domain (alignments not shown). The interruption of the polymerase and lysin genes in putative conserved regions, together with the differences in G+C content of the three introns and the endonucleases encoded within them (Fig. 3), suggests that they evolved independently but are convergently targeting essential functions.

Overview of relationship to other phages: phage K and *Listeria* phage A511 have similar structural modules. When structural proteins of phage K were examined by SDS-PAGE (Fig. 4), four that correspond to predicted proteins of phage K (ORFs 44, 49, 50, and 95) were identified on the basis of N-terminal sequencing. N-terminal sequencing identified the putative major tail sheath protein encoded by ORF 49 of phage K, which is 54.5 kDa (Fig. 4). The amino acid sequence matched the first 7 N-terminal amino acids of the ORF 49 product except for the initial methionine. Ben-Bassat et al. (4) previously reported that the N-terminal methionine residue is often processed when the second residue is an alanine, as is the case for ORF 49 and also ORFs 50 and 95 (Fig. 4). Database searches revealed a 57% identity with the tail sheath protein of *Listeria* phage A511 (23) (Fig. 4), which is a member of the *Myoviridae* family and has a contractile tail, linear double-stranded DNA, and a large genome of 116 kb. Only its structural module and amidase have been sequenced to date, and it cannot, therefore, be included on the proteome tree. N-terminal sequencing of band B resulted in 15 amino acids which are identical to residues 25 to 39 of the deduced protein product of ORF 44, indicating posttranslational cleavage of the first 23 amino acids. The product of this ORF shares 82% identity with the capsid protein of phage Twort (Table 1). The product of ORF 44 shares 66% identity with the capsid protein of phage A511, which also exhibits posttranslational cleavage of the first 23 amino acids (23) (Fig. 4). Interestingly band C corresponds to a protein of unknown function (ORF 95) (Fig. 4). N-terminal sequencing revealed the first 17 amino acids of this protein, which has a predicted molecular mass of 23.2 kDa (Fig. 4). The amino acid sequence obtained from band D (Fig. 4) corresponds to the product of ORF 50, with a predicted molecular mass of 15.9 kDa. ORF 50 shares an identity of 68% with ORF 8 from *Listeria* phage A511, which has an unknown function (23).

The proteins within the structural module are not homologous to the equivalent proteins of the sequenced lytic and temperate staphylococcal phages, but they do exhibit homology to *Listeria* phage A511, with the exception of ORF 41. The product of ORF 41 exhibited a low (22%) identity with a portal protein from *E. coli* phage P27 (Table 1). The portal protein is usually found transcribed near the large and small terminase

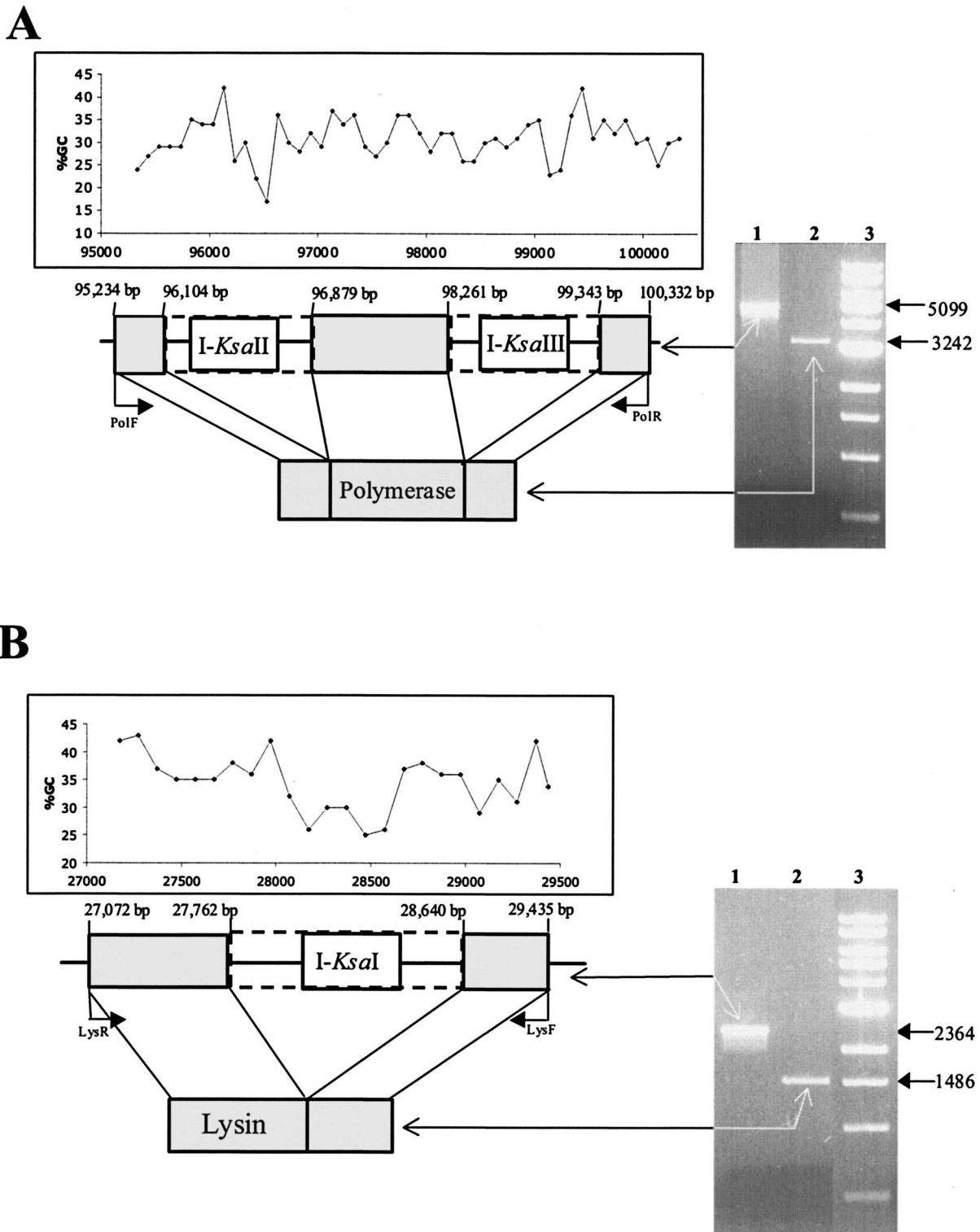


FIG. 3. (A) Schematic representation of the phage K polymerase gene interrupted by intron DNA. Dashed lines represent introns *pol-I2* and *pol-I3*, encoding I-KsaII and I-KsaIII, respectively. The in vivo splicing of phage K intron DNA from the polymerase gene is illustrated on a 1% agarose gel. Lane 1, PCR product obtained on phage K DNA with primers PolF and PolR; lane 2, PCR product obtained from cDNA of phage K DNA with primers PolF and PolR; lane 3, 1-kb ladder (New England Biolabs). Sizes of PCR products are indicated on the right in base pairs. The schematic diagram is overlaid with a graph illustrating the percent G+C content in this section of the genome. (B) Schematic representation of the phage K lysin gene interrupted by intron DNA. Dashed lines represent intron *lys-I1*, encoding I-KsaI. The in vivo splicing of phage K intron DNA from the lysin gene is illustrated on a 1% agarose gel. Lane 1, PCR product obtained on phage K DNA with primers LysF and LysR; lane 2, PCR product obtained from cDNA of phage K with primers LysF and LysR; lane 3, 1-kb ladder (New England Biolabs). Sizes of PCR products are indicated on the right in base pairs. The schematic diagram is overlaid with a graph illustrating the percent G+C content in this section of the genome.

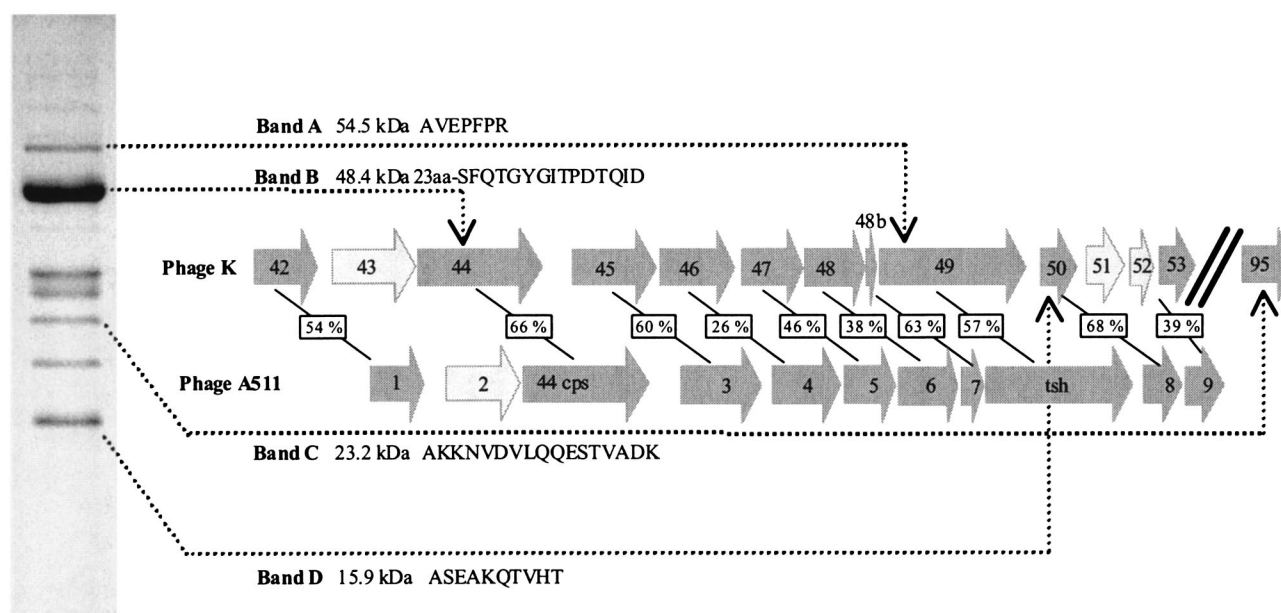


FIG. 4. One-dimensional SDS-PAGE of phage K proteins stained with Coomassie brilliant blue and schematic diagram of similarities with *Listeria* phage A511. Bands A, B, C, and D represent the four proteins that were N-terminally sequenced. The N-terminal sequences of each band and the ORF from the genome of phage K to which each band corresponds are shown to the right. Numbering of ORFs for phage A511 corresponds to that in reference 23. Dark grey arrows represent ORFs with a corresponding ORF in each phage, and light grey arrows represent ORFs that have no corresponding ORFs between phages K and A511. Percent identities shared between ORFs of phages K and A511 are indicated in boxes in the schematic diagram.

subunits (for example, in phage bII170 [8, 11]). ORF 35 is the putative large terminase subunit, although the homology is very weak and was apparent only with PSI-BLAST. No small terminase subunit was identified based on homology searches, but it is presumed to be encoded by one of the adjacent ORFs. Interestingly, this 11,361-bp (minus the portal protein) structural region of phage K shows significant homology with phage A511, not just at gene level but also in the arrangement of ORFs (23) (Fig. 4). Database searches with ORFs 42 to 53 revealed that all homologies were with the structural proteins of A511; with the exception of ORF44, no homologies were obtained with any other structural proteins of any other phage, including sequenced staphylococcal phages (Table 1). As Fig. 4 illustrates, all structural ORFs of A511, with the exception of ORF 2, have a homologous ORF in phage K. Phage K has three additional ORFs in the structural modules (ORF 43, 51, and 52), which have an unknown function and no corresponding ORF within phage A511 (Fig. 4). Most likely, ORF 43 has a complement in A511 (ORF2), where a gene replacement seems to have occurred, whereas in phage A511 a deletion event may have occurred, with the loss of ORFs corresponding to ORFs 51 and 52 from phage K. The dramatic similarity between these large (11-kb) regions of both phages suggests that phage K and A511 are related and could constitute a new lineage of *Myoviridae* infecting low-G+C-content gram-positive bacteria. The elucidation of the genomes of A511 and other large *Myoviridae* (21) may facilitate the classification of these large *Myoviridae* to the same lineage.

Conclusions. Phage K is a large, virulent bacteriophage which infects a broad range of staphylococci, including multiple-drug-resistant strains of *S. aureus*. Detailed genetic characterization of this phage has unveiled a number of features as

follows. (i) Phage K has been taxonomically placed in its own group because of overall uniqueness compared to other phage. (ii) The genome also contains introns in essential phage functions, two in the polymerase gene and one in the lysin gene. (iii) Phage K contains a large region with remarkable homology to *Listeria* phage A511. (iv) Finally, phage K has a remarkable paucity of GATC and GGCC, sites suggesting that phage K has evolved an efficient counter defense against host restriction-modification systems.

ACKNOWLEDGMENTS

This research was funded by the Irish Government under the FIRM program as part of the National Development Plan 2000-2006, by European Union structural funds, and by the Science Foundation Ireland. Sarah O'Flaherty is in receipt of a Teagasc Walsh Fellowship.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
- 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.
- Bateman, A., and N. D. Rawlings. 2003. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends Biochem. Sci.* **28**:234-237.
- Ben-Bassat, A., K. Bauer, S. Y. Chang, K. Myambo, A. Boosman, and S. Chang. 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bacteriol.* **169**:751-757.
- Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354**:484-490.
- Brussow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. *Mol. Microbiol.* **39**:213-222.
- Burnet, F. M., and D. Lush. 1935. The staphylococcal bacteriophages. *J. Pathol. Bacteriol.* **40**:455-469.
- Casjens, S. 2003. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* **49**:277-300.

9. Casjens, S., G. Hatfull, and R. Hendrix. 1992. Evolution of dsDNA tailed-bacteriophage genomes. *Semin. Virol.* **3**:383–397.
10. Chatterjee, A. N. 1969. Use of bacteriophage-resistant mutants to study the nature of the bacteriophage receptor site of *Staphylococcus aureus*. *J. Bacteriol.* **98**:519–527.
11. 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.
12. Desiere, F., S. Lucchini, C. Canchaya, M. Ventura, and H. Brussow. 2002. Comparative genomics of phages and prophages in lactic acid bacteria. *Antonie Leeuwenhoek* **82**:73–91.
13. Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brussow. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology* **283**:240–252.
14. Desiere, F., W. M. McShan, D. van Sinderen, J. J. Ferretti, and H. Brussow. 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic Streptococci: evolutionary implications for prophage-host interactions. *Virology* **288**:325–341.
15. Edgell, D. R., M. Belfort, and D. A. Shub. 2000. Barriers to intron promiscuity in bacteria. *J. Bacteriol.* **182**:5281–5289.
16. Ermolaeva, M. D., H. G. Khalak, O. White, H. O. Smith, and S. L. Salzberg. 2000. Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.* **301**:27–33.
17. Hotchin, J. E. 1951. The influence of acridines on the interaction of *Staphylococcus aureus* and *Staphylococcus* K phage. *J. Gen. Microbiol.* **5**:609–618.
18. Hotchin, J. E. 1954. The purification and electron microscopical examination of the structure of staphylococcal bacteriophage K. *J. Gen. Microbiol.* **10**:250–260.
19. Hotchin, J. E., I. M. Dawson, and W. J. Elford. 1952. The use of empty bacterial membranes in the study of the adsorption of *Staphylococcus* K phage upon its host. *Br. J. Exp. Pathol.* **33**:177–182.
20. 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 phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325. *Gene* **289**:109–118.
21. Jarvis, A. W., L. J. Collins, and H. W. Ackermann. 1993. A study of five bacteriophages of the Myoviridae family which replicate on different gram-positive bacteria. *Arch. Virol.* **133**:75–84.
22. Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. 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. Inoue, C. Kaito, K. Sekimizu, H. Hiramawa, 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*. *Lancet* **357**:1225–1240.
23. Loessner, M. J., and S. Scherer. 1995. Organization and transcriptional analysis of the *Listeria* phage A511 late gene region comprising the major capsid and tail sheath protein genes *cps* and *tsh*. *J. Bacteriol.* **177**:6601–6609.
24. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
25. Lucchini, S., F. Desiere, and H. Brussow. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* **73**:8647–8656.
26. Mesyanzhinov, V. V., J. Robben, B. Grymonprez, V. A. Kostyuchenko, M. V. Bourkaltseva, N. N. Sykilinda, V. N. Krylov, and G. Volckaert. 2002. The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *J. Mol. Biol.* **317**:1–19.
27. Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum, O. White, I. T. Paulsen, W. C. Nierman, J. Lee, B. Szczypinski, and C. M. Fraser. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J. Bacteriol.* **185**:5220–5233.
28. Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger. 2003. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**:86–156.
29. Moineau, S., S. Pandian, and T. R. Klaenhammer. 1993. Restriction-modification systems and restriction endonucleases are more effective on lactococcal bacteriophages that have emerged recently in the dairy industry. *Appl. Environ. Microbiol.* **59**:197–202.
30. O'Sullivan, D., D. P. Twomey, A. Coffey, C. Hill, G. F. Fitzgerald, and R. P. Ross. 2000. Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis*. *Mol. Microbiol.* **36**:866–875.
31. Pantucek, R., A. Rosypalova, J. Doskar, J. Kailerova, V. Ruzickova, P. Borecka, S. Snopkova, R. Horvath, F. Gotz, and S. Rosypal. 1998. The polyvalent staphylococcal phage phi 812: its host-range mutants and related phages. *Virology* **246**:241–252.
32. Rees, P. J., and B. A. Fry. 1981. The morphology of staphylococcal bacteriophage K and DNA metabolism in infected *Staphylococcus aureus*. *J. Gen. Virol.* **53**:293–307.
33. Rees, P. J., and B. A. Fry. 1981. The replication of bacteriophage K DNA in *Staphylococcus aureus*. *J. Gen. Virol.* **55**:41–51.
34. Rees, P. J., and B. A. Fry. 1983. Structure and properties of the rapidly sedimenting replicating complex of staphylococcal phage K DNA. *J. Gen. Virol.* **64**:191–198.
35. Rigden, D. J., M. J. Jedrzejas, and M. Y. Galperin. 2003. Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. *Trends Biochem. Sci.* **28**:230–234.
36. Rohwer, F., and R. Edwards. 2002. The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol.* **184**:4529–4535.
37. Rountree, P. M. 1949. The serological differentiation of Staphylococcal bacteriophages. *J. Gen. Microbiol.* **3**:164–173.
38. Sambrook, J., and D. W. Russell. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
39. Sasaki, Y., J. Ishikawa, A. Yamashita, K. Oshima, T. Kenri, K. Furuya, C. Yoshino, A. Horino, T. Shiba, T. Sasaki, and M. Hattori. 2002. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Res.* **30**:5293–5300.
40. Sussenbach, J. S., C. H. Monfoort, R. Schiphof, and E. E. Stobberingh. 1976. A restriction endonuclease from *Staphylococcus aureus*. *Nucleic Acids Res.* **3**:3193–3202.
41. Sussenbach, J. S., P. H. Steenbergh, J. A. Rost, W. J. van Leeuwen, and J. D. van Embden. 1978. A second site-specific restriction endonuclease from *Staphylococcus aureus*. *Nucleic Acids Res.* **5**:1153–1163.
42. van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner. 2000. *Virus taxonomy: the classification and nomenclature of viruses*. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, Calif.
43. Vybiral, D., M. Takac, M. Loessner, A. Witte, U. von Ahsen, and U. Blasi. 2003. Complete nucleotide sequence and molecular characterization of two lytic *Staphylococcus aureus* phages: 44AHJD and P68. *FEMS Microbiol. Lett.* **219**:275–283.