Available online at www.sciencedirect.com

ScienceDirect

Virology 365 (2007) 1–9

VIROLOGY

www.elsevier.com/locate/yviro

KSY1, a lactococcal phage with a T7-like transcription

Alain Chopin [a,*], Hélène Deveau [b,c], S. Dusko Ehrlich [a], Sylvain Moineau [b,c,d], Marie-Christine Chopin [a]

[a] *Laboratoire de Génétique Microbienne, INRA, 78352 Jouy-en-Josas, France*

[b] *Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Université Laval, Quebec City, Quebec, Canada G1K 7P4*

[c] *Groupe de Recherche en Ecologie Buccale (GREB), Faculté de Médecine Dentaire, Université Laval, Quebec City, Quebec, Canada G1K 7P4*

[d] *Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Quebec City, Quebec, Canada G1K 7P4*

Received 17 October 2006; returned to author for revision 12 March 2007; accepted 23 March 2007

Available online 30 April 2007

Abstract

The virulent lactococcal phage KSY1 possesses a large elongated capsid (223 nm long, 45 nm wide) and a short tail (32 nm). This phage of the *Podoviridae* group (C3 morphotype) has a linear 79,232-bp double-stranded DNA genome, which encodes 131 putative proteins and 3 tRNAs. This is the first description of the genome of a phage of this morphotype. KSY1 possesses a T7-like transcription system, including an RNA polymerase and a series of specific promoters, showing sequence homology to other known T7-like RNA polymerase promoters. Late stages of KSY1 multiplication are resistant to rifampicin. Otherwise, KSY1 shares limited similarity with other *Podoviridae* phages. Fourteen KSY1 structural proteins were identified by SDS–PAGE analysis. Among these proteins, those forming the distal tail structure and likely involved in host recognition are encoded by a 5-kb genomic region of KSY1. This region consists of a mosaic of DNA segments highly homologous to DNA of other lactococcal phages, suggesting an horizontal gene transfer.

Keywords: Bacteriophage; *Lactococcus*; *Podoviridae*; Genome sequence; T7-like RNA polymerase

Introduction

It is now recognized that bacteriophages are the most abundant organisms in the biosphere. Although largely unexplored, their genetic diversity is apparently immense since about half of the newly described phage proteins have no database match. Model systems have shown that every bacterial species can be infected by a variety of phages, leading to the estimation that the number of phages is at least one order of magnitude higher than the number of bacteria (Rohwer, 2003). However, it appears that this myriad of phages may be organized into a relatively small number of types (Bamford, 2003; Casjens et al., 1992).

Lactococci are low G+C Gram-positive bacteria used for the manufacture of an array of fermented milk products and are highly susceptible to phage infections. Due to their nuisance for the dairy industry, lactococcal phages have been highly documented. All known lactococcal phages possess a double-stranded DNA genome and a tail, thus belonging to the Caudovirales order (Ackermann, 1999). The overwhelming majority of them fall within three unrelated groups of DNA homology. They are designated: 936 (virulent, isometric-headed, genome length of 28–32 kb), c2 (virulent, prolate-headed, genome length of 22 kb), and P335 (temperate/virulent, isometric-headed, genome length of 33–42 kb). The genome sequence was determined for 18 of these wild-type lactococcal phages, including 11 from the P335 group (Labrie and Moineau, 2007), 5 from the 936 (Mahony et al., 2006), and 2 from the c2 (Schouler et al., 1994; Lubbers et al., 1995). These genomes show a similar pattern of gene order loosely resembling that of coliphage λ, which led to the proposal that lactococcal phages (particularly the 936 and P335-like phages) belong to a λ supergroup of phages within the *Siphoviridae* family (Brüssow and Desiere, 2001; Brüssow and Hendrix, 2002; Proux et al., 2002). This contrasts with *Escherichia coli* phages, which belong to many different types (λ, T7, T4, P1, Mu, M13, etc.).

* Corresponding author. Fax: +33 134652521.
E-mail address: Alain.Chopin@jouy.inra.fr (A. Chopin).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.03.044

The rare remaining lactococcal phages form at least seven different homology groups (Deveau et al., 2006). Characterization of the genome of these phages would more accurately assess the natural diversity of the lactococcal phage community and possibly, shed light on their origin, evolution, and relationships with other phages. The recent characterization of the genome of the unique member of the lactococcal phage group Q54 exemplifies such diversity as it contains an unusual configuration of modules, which suggested past recombination events between c2- and 936-like phages (Fortier et al., 2006).

Here, we report the characterization of the virulent phage KSY1, one of these rare lactococcal phages. KSY1 was isolated three decades ago from spoiled "Viili", a Finnish fermented ropy milk (Saxelin et al., 1979, 1986). KSY1 was recognized as the reference phage for the lactococcal phage species that bears its name (Deveau et al., 2006; Jarvis et al., 1991). We show that KSY1 possesses a single-subunit RNA polymerase, together with a series of specific transcription promoters showing recognizable homology to other T7-like RNA polymerase promoters. The late stages of KSY1 are resistant to rifampicin, indicating that host RNA polymerase is no longer required for phage development. This is the first description of a phage active against a Gram-positive host using this mode of transcription. Otherwise, KSY1 differs from other phages in the *Podoviridae* family by its genome size and arrangement and by most of its proteins.

Results

KSY1 morphology

Phage KSY1 has a very distinctive morphology with an elongated capsid (223 nm long, 45 nm wide), and a short tail (32 nm) (Fig. 1). KSY1 therefore belongs to the *Podoviridae* family, more precisely to the C3 morphotype of Ackermann and Eisenstark (1974), a type that has been observed in less than 1% of known phages (Ackermann, 2001).

KSY1 genome sequence

The KSY1 genome is made of a linear double-stranded DNA molecule of 79,232 bp. It has a G+C content of 35.1%, which is very close to that of *Lactococcus lactis* (35.3%) and of other lactococcal phages. Determination of the sequence of genome ends was hampered by the under-representation of these regions in the DNA bank. The sequence was therefore determined by using KSY1 genomic DNA as a template and known sequences near either end of the genome as sequencing primers. Sequence of genome ends was also confirmed by sequencing several junctions obtained after ligation of each genome end to a blunt-ended pUC19 molecule. Consequently, the KSY1 genome is blunt-ended. No terminal repeats or redundancy was identified at genome ends, which raises the question of how KSY1 DNA can be replicated.

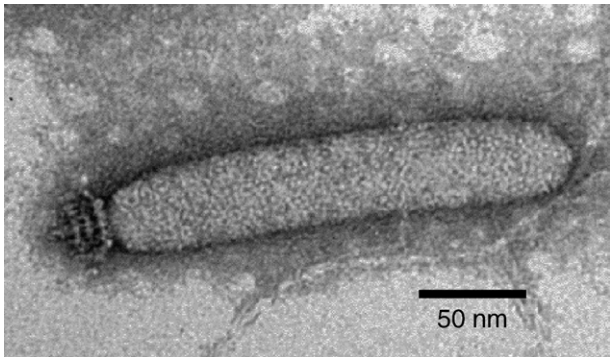


Fig. 1. Electron micrograph of phage KSY1.

Early attempts to clone KSY1 DNA fragments revealed that this genome was resistant to most type II restriction enzymes, except for those whose recognition site contains only A or T nucleotides. A computer search of restriction sites in KSY1 DNA sequence revealed the presence of several type II restriction sites, although they were less frequent than expected based on a purely random nucleotide distribution. This avoidance of restriction sites in phage DNA sequences is well known (Tock and Dryden, 2005) and has also been described in lactococcal phages (Moineau et al., 1993). It is likely that resistance to restriction of KSY1 DNA is due to a, yet unknown, chemical alteration of either C or G nucleotides.

KSY1 coding sequences

One hundred and thirty-one open reading frames were identified. They are all oriented in the same direction and close-packed. Almost 94% of the sequence is coding and 45 short gene overlaps are present. Three tRNA genes are also present ($tRNA^{Met}$, $tRNA^{Arg}$, and $tRNA^{Gln}$). Unexpectedly, the 3′ end of the non-template strand consists of a stretch of ten Cs. Schematic presentation of the genome is shown in Fig. 2.

Blast searches (Altschul et al., 1997) identified 38 KSY1 putative proteins (out of 131, 29%) with significant similarity to known proteins in the GenBank database. By using the same database, but restricted to viral genes, 3 additional homologous proteins were discovered. Among the 41 KSY1 proteins with similarity to known proteins, 20 (49%) are most similar to a protein from an organism (either a phage or a bacterium) from the dairy environment, 14 (34%) from a non-dairy terrestrial environment, and 7 (17%) from a marine environment (Supplementary Table 1). These homologies, together with further analyses described below, allowed the assignment of a probable function to 26 KSY1 proteins (Fig. 2).

Phage KSY1 possesses a T7-like transcription system

The deduced gene product of the *orf14* (gp14) is similar to single-subunit RNA polymerases (RNAP) found in fungal mitochondria and in phages SSP7 and P60, two T7-like phages infecting marine cyanobacteria. Gp14 possesses all the amino acid residues important for the function of T7 RNAP, which are conserved in T7-like phage RNAPs (McAllister and Raskin, 1993; Sousa et al., 1993; Cermakian et al., 1997; Imburgio et al., 2002).

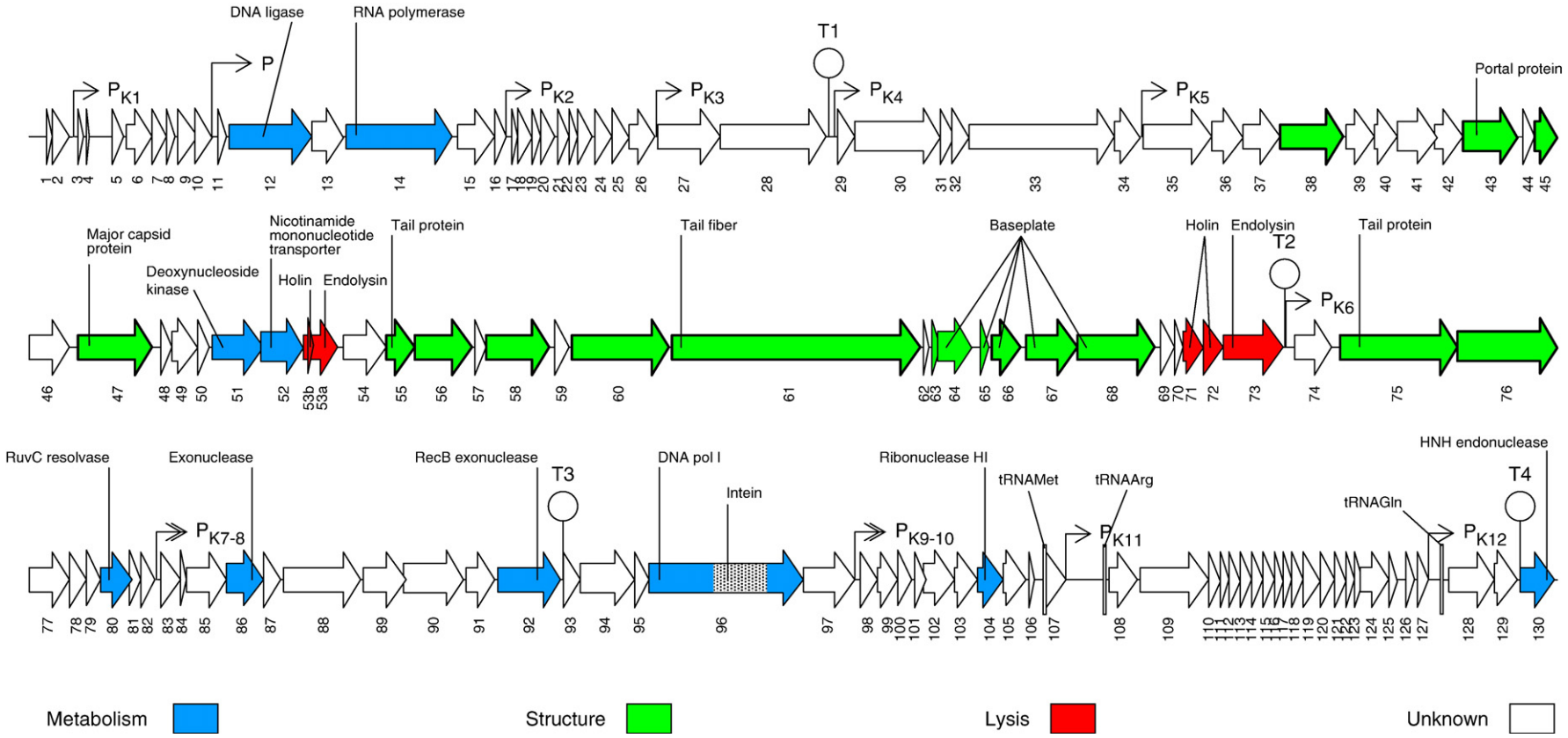


Fig. 2. KSY1 genome structure. *Orfs* 1 to 130 are indicated by arrows. Thick arrows indicate structural proteins that have been detected in the SDS–PAGE analysis. Putative functions are indicated. Putative transcription promoters are indicated by P (lactococcal vegetative promoter) and P_{K1} to P_{K12} (KSY1-specific promoters). Transcription terminators T1 to T4 are indicated by a hairpin-like symbol. Thin open bars represent the three tRNA genes. The dotted box represents the intein gene sequence within *orf96*.

In many T7-like phages, the RNAP binds and initiates transcription at specific conserved promoter sequences. By computer analysis, a 26-nt motif was found in twelve copies, with some degeneracy, in the KSY1 genome (Table 1). This motif may represent specific promoters for the KSY1 RNAP. All twelve copies are oriented in the same direction and are strictly localized in intergenic regions. The consensus motif shows sequence homology to transcriptional promoters of other T7-like phages, especially in the highly conserved −8 to −3 core region, strongly suggesting that it represents KSY1 transcriptional promoters. These promoters have been denoted P_{K1} to P_{K12}. Upstream of *orf83* and *orf98*, two tandem copies of the promoter (named P_{K7-8} and P_{K9-10}) are separated by 8 and 18 nucleotides, respectively. Noteworthy is that the degeneracy in the promoter sequence depends on the location in the genome. Promoters P_{K6} to P_{K10} have the consensus sequence, with the exception of promoter P_{K8}, which contains one mismatch. By contrast, promoters P_{K1} to P_{K5}, P_{K11} and P_{K12} contain 2 to 9 mismatches relative to the consensus 26-nt sequence. These findings are reminiscent of phage T7 where promoters were classified in two classes, referred to as class II or class III, based on their location and nucleotide sequence (Dunn and Studier, 1983).

Other transcription signals were searched in the KSY1 genome. Four putative transcription terminators were identified by visual inspection and denoted T1 to T4 (Fig. 2). However, neither visual inspection of intergenic sequence nor computer analysis enabled the identification of any host consensus vegetative promoter. However, a single extended −10 promoter sequence (5′-AATATGTTATAAT-3′) was identified upstream *orf11*, and designated P (Fig. 2). This type of promoter lacks the consensus −35 sequence and follows the −10 yNtNTGy-TATAAT consensus (Kumar et al., 1993). Such promoter sequences have been shown to be functional in *L. lactis* and its phages (Bidnenko et al., 1995; Chopin, 1993). Therefore, following the injection of the phage DNA into the cell, this extended −10 promoter could serve as a site for transcription initiation by the host *L. lactis* RNAP, leading to the early synthesis of the RNAP (gp14) of phage KSY1. Afterwards, the phage RNAP would likely initiate most transcription from its specific promoter sequences (P_{K1} to P_{K12}).

The existence of a functional single-subunit RNAP in KSY1 was established by showing that late stages of KSY1 multiplication are resistant to rifampicin. A one-step growth experiment (Fig. 3A) revealed a burst of 30 phage particles after 5 h of incubation at 30 °C of cells infected at a multiplicity of infection of 0.1. These growth parameters are similar to those previously observed (Saxelin et al., 1979). When rifampicin was added within 60 min following infection, phage multiplication was completely inhibited (Fig. 3B). However, a gradual increase in phage yield was observed when rifampicin was added between 1 and 3 h after infection. These results indicate that the host RNA polymerase is only required during the early steps of KSY1 multiplication. Later on, phage multiplication becomes rifampicin resistant, suggesting that single-subunit RNAP is catalyzing all transcription.

KSY1 DNA replication and metabolism

KSY1 gp96 is highly similar to the DNA polymerase of *Aquifex aeolicus* (Supplementary Table 1). However, *orf96* contains an intein-coding sequence similar to the one present

Table 1
Sequence of KSY1 putative transcription promoters and homology to other T7-like phage promoters

Promoter	Coordinates	Sequence
PK1	736–761	GAGATACA**TT**G**T**T**A**ACTAATA**GG**A**AC**[a]
PK2	8104–8129	**T**AGA**A**ACAA**AG**GT**A**ACTAATA**A**TATA
PK3	10,642–10,667	GAGATACAA**A**A**A**T**A**ACTA**T**TA**TATAG**
PK4	13,687–13,712	**T**AGATAC**T**ATAGTTACTAATAC**A**ATA
PK5	18,887–18,912	GAGATACAA**A**A**A**TTACTAATA**AC**AT**T**
PK6	47,897–47,922	GAGATACAATAGTTACTAATACTATA
PK7	54,864–54,889	GAGATACAATAGTTACTAATACTATA
PK8	54,898–54,923	GAGATACAA**A**AGTTACTAATACTATA
PK9	67,003–67,028	GAGATACAATAGTTACTAATACTATA
PK10	67,047–67,072	GAGATACAATAGTTACTAATACTATA
PK11	70,674–70,699	GAGATACAATAGTTACTAATACTA**AT**
PK12	76,966–76,991	GAGATACAATAGTTACTAATAC**ATA**A
KSY1 consensus		GAGATACAATAGTTACTAATACTATA
		| | |||
T7 consensus		TAATACGACTCACTATAGGGAGA[b]
		| ||||| |
ϕYeO3-12 consensus		AATTAACCCTCACTAAAGGGAGA[c]
		| ||||| |
gh-1 consensus		TAAAAACCCTCACTATGGC--C-[b]
		| |||| |
SP6 consensus		ATTTAGGTGACACTATAGAAGAA[b]
		↑+1[d]

[a]Mismatches in KSY1 promoters are indicated in bold letters.
[b]Consensus sequence from Pajunen et al. (2002).
[c]Consensus sequence from Imburgio et al. (2000).
[d]Transcription start.

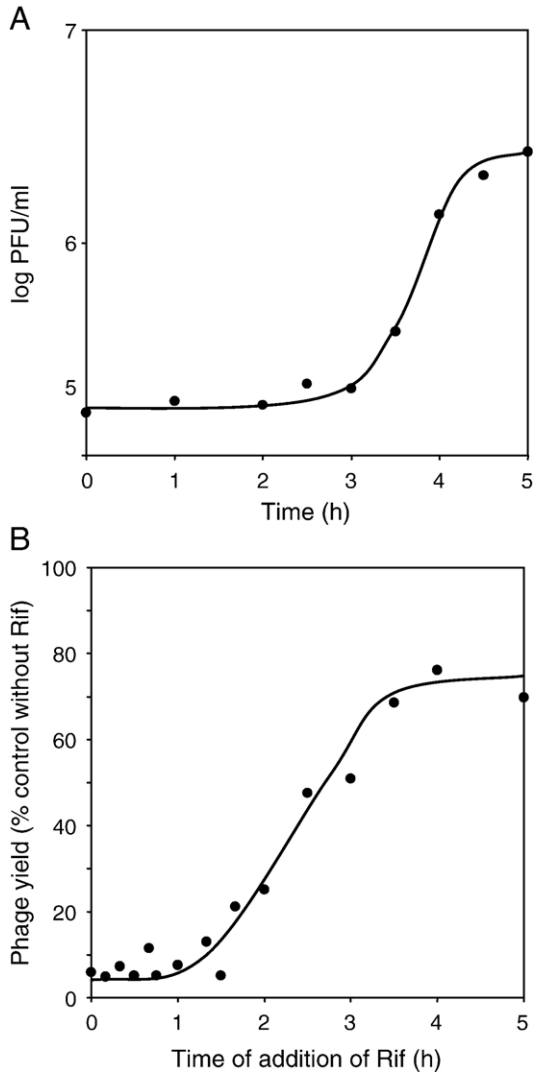


Fig. 3. Late stages of KSY1 multiplication are resistant to rifampicin. (A) One-step growth curve. KSY1 was used to infect *L. lactis* at a multiplicity of infection of 0.1. The number of PFU/ml was determined at indicated times. (B) Rifampicin was added to a parallel culture at various periods after infection by KSY1. Incubation was continued up to 5 h and the number of PFU/ml was determined. Phage yield was calculated by comparing the yield of phages from the cultures with rifampicin to the yield of phages from the culture without rifampicin. Results are mean of two independent experiments.

within the *nrdB* gene of *A. aeolicus*, coding for a ribonucleotide reductase. This sequence possesses typical intein features (Perler et al., 1997) and is likely to be spliced following translation of gp96. Several other putative proteins, sharing homology with DNA ligases, kinases, nucleases, or mononucleotide transporters, may participate in DNA replication and metabolism.

KSY1 genes coding for structural proteins

Sequence analysis revealed few genes encoding putative structural proteins. The structural protein composition of phage KSY1 was therefore characterized by SDS–PAGE coupled with LC–MS/MS. Mass spectrometry analysis of thirteen protein bands detected after Coomassie blue staining of the gel led to the identification of sixteen KSY1 structural proteins (Figs. 2 and 4). Ten protein bands in the SDS–PAGE each represented one KSY1 structural protein, but two different proteins were detected in three other seemingly single bands (5, 7, and 9) (Fig. 4). Band #7 contained the most predominant KSY1 structural protein. The vast majority of peptides detected in band #7 corresponded to the gp47 peptide sequence. Considering the morphology of this phage, gp47 is likely the major capsid protein. A second protein (gp58) was also detected with high coverage but the number of detected peptides was very small in comparison with gp47 (data not shown).

Overall, the molecular masses estimated by SDS–PAGE for the majority of proteins were in agreement with the calculated masses from the gene sequence. Discrepancies were observed only for proteins detected in bands #1 and #3 (Fig. 4). In fact, three protein bands (#1, #2, and #3) of different size were all linked to gp61. The predicted size of gp61 is 160 kDa, which is equivalent to band #2. It is possible that band #1 corresponds to a dimer of gp61 while band #3 may be a truncated version of gp61. The vast majority of peptides detected by LC–MS/MS in band #3 were mapped at the C-terminal part of the protein, suggesting that the N-terminal could be truncated by ~100 amino acids.

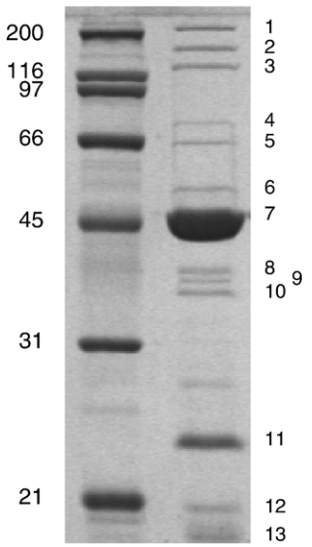
The structural gp61 protein possess several features of a tail fiber protein. *Orf61* is the largest KSY1 gene, a characteristic shared by many phage tail fiber genes; it is located upstream of the baseplate genes; it encodes a protein that shows amino acid sequence similarity to several phage tail proteins; and it carries a conserved polygalacturonase domain (PGU1) in its C-terminal region. Phage tail proteins often carry murein hydrolase activities (Kenny et al., 2004; Moak and Molineux, 2004). Such activity could be required for a more efficient access to phage receptor and/or for breaking down the peptidoglycan layer to allow the entry of the phage DNA into the bacterial cell. The polygalacturonase activity may also contribute to the reported host-capsule degrading activity of phage KSY1 (Saxelin et al., 1979).

Among the structural proteins identified, gp64 to gp68 are strongly homologous to proteins forming a double-disk baseplate and a central tail fiber of the distal tail structure of the lactococcal phage TP901-1 (Brondsted et al., 2001; Pedersen et al., 2000; Vegge et al., 2005). Indeed, the gp64, gp65, and gp67 of KSY1 are homologous to BppU of TP901-1, which encodes the upper base plate. The deduced N-terminal region of KSY1 gp66 is similar to the lower base plate protein BppL (Vegge et al., 2006). Interestingly, all these components of the KSY1 distal structure are encoded by a c.a. 5-kb DNA region made of DNA segments >80% identical to sequences from lactococcal phages of the polythetic P335 group (Supplementary Fig. 1).

KSY1 lysis systems

Two gene products with putative cell wall hydrolase activity were identified in the KSY1 deduced proteome, namely gp53 and gp73. Gp53 is similar to the endolysin amidase of many phages infecting pathogenic streptococci and *Staphylococcus*

A



B

Band	Gene product	Predicted function	Molecular mass (kD)	
			Calculated	SDS-PAGE
1	61	Tail fiber protein	160	>200
2	61	Tail fiber protein	160	160
3	61	Tail fiber protein	160	130
4	75	Tail protein	78	80
5a	60	-	65	65
5b	76	-	63	65
6	68	Baseplate	49	50
7a	47	Major capsid protein	48	45
7b	58	-	41	45
8	43	Portal protein	36	40
9a	56	-	36	38
9b	67	Baseplate	34	38
10	38	-	38	36
11	55	Tail protein	18	24
12	66	Baseplate	18	19
13	45	-	14	17

Fig. 4. LC–MS/MS analysis of structural proteins of phage KSY1. (A) Coomassie blue staining of a 12% SDS–polyacrylamide gel showing structural proteins from phage KSY1. Numbers on the right indicate bands cut out from the gel and identified by LC–MS/MS. The sizes (in kilo-Dalton) of the different proteins from the broad-range molecular mass standard (M) are indicated on the left. (B) Identification of KSY1 viral proteins from corresponding bands shown in panel A.

aureus. No plausible holin gene is present nearby but a reading frame embedded out of frame is present within the 5′ region of *orf53*. The embedded gene is preceded by a good translation initiation region and possesses a dual-start motif. It encodes a 64- or 62-residue protein that possesses a transmembrane domain in its N-terminal sequence but lacks the highly charged C-terminus characteristic of most holins. All these features are identical to that of the holin gene *hol187* embedded within the endolysin gene of *S. aureus* phage 187 (Loessner et al., 1999). Thereafter, these putative lysin and holin genes are denoted *orf53a* and *orf53b*.

Gp73 is strongly homologous to endolysin amidase of many firmicute phages, including the lactococcal phage r1t from the P335 group (Supplementary Fig. 1). The two genes immediately upstream, *orf71* and *orf72*, encode proteins that present all the characteristic features of holins. Both are small (i.e. <150 residues), have a clearly discernable transmembrane domain with no net predicted charge, and have a hydrophilic highly charged C-terminal sequence (Wang et al., 2000). *Orf69*- and *orf70*-encoded proteins also possess two and one putative transmembrane domain, respectively. Their genomic location is conserved relative to the endolysin gene in lactococcal prophage genome bIL285 (P335 group) and could also participate in the formation of a pore for the amidase to get access to the cell wall. The presence of two lytic systems in a phage genome has already been observed in several double-stranded DNA phages (Wang et al., 2000).

Discussion

Phage KSY1 is a member of an under-represented group of lactococcal phages and its characterization was carried out to assess the natural diversity of the lactococcal phage community. KSY1 was found to be very different from all lactococcal phages described so far, by its morphology, genome size and gene arrangement.

KSY1 belongs to the *Podoviridae* family of the Caudovirales order and has an unusually large elongated capsid (C3 morphotype). This is the first phage of this morphotype to be characterized at the genomic level. Its genome is much larger than that of previously characterized lactococcal phages and does not present their λ-type gene arrangement. KSY1 proteins show similarity to proteins from a wide range of diverse phages or bacteria, when lactococcal phages described so far shared protein homology with related phages only.

The KSY1 genomic region encoding the components of the distal part of its tail, is made of short gene segments that are over 80% identical to gene segments of several lactococcal phages of the polythetic P335 group. Some of these segments are identical to a TP901-1 gene encoding the receptor-binding protein (RBP), which is involved in the recognition of the host cell (Vegge et al., 2006), and has a modular structure (Spinelli et al., 2006a,b). The mosaic structure of this region, together with the high level of DNA sequence conservation, suggest that it has been acquired in recent genetic shuffling with phages of the P335 group. This might have occurred during infection of lactococcal cells, which are usually lysogenic for P335 phages, or during co-infection of a host with an ancestor of KSY1 and a virulent P335-like phage (Bouchard and Moineau, 2000; Huggins and Sandine, 1977; Labrie and Moineau, 2007). At least three of the shuffled proteins, found among the KSY1 structural proteins, are functional. It is known that phage genome regions involved in host adaptation are highly plastic (Mann et al., 2005; Miller, 2003; Tétart et al., 1998). The exchange of parts of their tail-fiber genes to evolve new host range has been largely documented in *E. coli* phages (reviewed

in Sandmeier, 1994) and also demonstrated in lactococcal phages (Dupont et al., 2004; Stuer-Lauridsen et al., 2003; Vegge et al., 2006). KSY1 represents an additional example of such an adaptation, by gain of genes from an unrelated phage group.

A striking feature of KSY1 is the presence of a T7-like transcription system. This system involves a T7-like RNAP together with a series of conserved putative specific transcription promoters. These promoters show recognizable sequence similarity to those of phages in the T7 group. In addition, sequence degeneracy exhibited by some KSY1 promoters, depends on their location in the genome, suggesting a role in the temporal regulation of the KSY1 genes. The RNAP gene in KSY1 occupies the same relative position as its counterpart in T7-like phages. It is likely transcribed under the dependence of the apparently unique KSY1 promoter recognized by host RNAP. We also observed that late steps of KSY1 development are resistant to rifampicin, indicating that *L. lactis* RNAP is no longer necessary and that KSY1 possess a functional RNAP. These observations suggest that KSY1 uses a transcription strategy similar to that of phage T7. Up to now, single-subunit RNA polymerases have been found in T7-like and N4 phages and in eukaryotic organelles. Our finding extends this range to a phage active against a Gram-positive host.

Materials and methods

Propagation of the phage and DNA preparation

Phage KSY1 and its host were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca). The host, *Lactococcus lactis* IE-16, was grown at 30 °C in M17 broth supplemented with 0.5% glucose (GM17) (Terzaghi and Sandine, 1975). For phage amplification, phage and host were incubated at 22 °C for 24 h in the presence of 10 mM $CaCl_2$. Glycine (0.5%) was added to the top agar to increase plaque size and facilitate phage enumeration (Lillehaug, 1997). Phage lysates were concentrated with PEG and purified on a discontinuous CsCl gradient (Sambrook and Russell, 2001). Ultracentrifugation was performed using a Beckman SW41 Ti rotor at 35,000 rpm for 3 h. The DNA was isolated from CsCl-purified phages as reported elsewhere (Chibani Azaiez et al., 1998). Phage KSY1 was observed under a transmission electron microscope as described previously (Deveau et al., 2006).

SDS–PAGE analysis of structural proteins

Purified phages (10^{10} PFU/ml) were analyzed for structural proteins by standard Tris–glycine 12% SDS–polyacrylamide gel electrophoresis (PAGE) procedures (Laemmli, 1970). Samples were mixed with 4× sample loading buffer and boiled for 5 min before loading. Proteins were detected after Coomassie blue staining. Protein bands were cut out of the gel, digested with trypsin, and identified by liquid chromatography–tandem mass spectrometry (LC–MS/MS) (Centre Protéomique de l'Est du Québec, Centre de Recherche du CHUL, Université Laval).

DNA sequencing

Phage DNA was sheared by sonication to a size of about 1 kb. The DNA ends were repaired and ligated to a 20-mer *Not*I-linker, prepared by annealing primer 1 (5′-GGACTGG-CGGCCGCTCCAGG-3′) to primer 2 with reverse complementary sequence. The DNA fragments were PCR-amplified using primer 1, digested with *Not*I, and cloned into dephosphorylated, *Not*I-digested cloning vector (Rohwer et al., 2001). A pUC19-derivative, carrying a *Not*I site, was constructed and used as cloning vector. For sequencing, we used Big Dye chemistry, ABI PRISM 377A Sequencer (PE Applied Biosystems), and either primer 1 or 2. Approximately 1600 reads were assembled using GAP4, which is part of the Staden package software (Staden et al., 2000), resulting in a single open contig with a 8.8-fold coverage.

ORF prediction and annotation

Open reading frame prediction was performed using GeneMark.hmm (Lukashin and Borodovsky, 1998) and Heuristic GeneMark (Besemer and Borodovsky, 1999). The prediction was verified and further improved by visual inspection using criteria such as the presence of a ribosome-binding site, the possible existence of short Orfs and non-AUG start codons, as well as codon usage analysis. tRNA genes were identified through the use of tRNAscan-SE (Lowe and Eddy, 1997). Transcriptional terminators (ρ-independent) were identified by visual inspection of the sequence. Potential transcription promoters recognized by the host RNAP were searched within intergenic regions using the Neural Network Promoter Prediction Program (Reese, 2001). The translated ORF products were compared with known protein sequences using PSI Blast (Altschul et al., 1997) and CD-search (Marchler et al., 2005), using the non-redundant public GenBank database. Repeats and patterns were searched using MEME (Bailey and Elkan, 1994) and FindPatterns+ from the GCG package (GCG Version 11.0, Accelrys Inc., San Diego, CA).

Accession number

The sequence data reported in this paper has been deposited in EMBL/GenBank/DDBJ databases under accession number DQ535032.

Acknowledgments

We thank Nicolas Pons for his help in the computer search of transcription promoters. We are grateful to Karidia Diallo and Hans-W. Ackermann for technical help at the beginning of the project. Hélène Deveau is the recipient of a graduate student scholarship from the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT). This study was funded, in part, by a strategic grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.M.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2007.03.044.

References

Ackermann, H.W., 1999. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. 51, 135–201.

Ackermann, H.W., 2001. Frequency of morphological phage descriptions in the year 2000. Arch. Virol. 146, 843–857.

Ackermann, H.W., Eisenstark, A., 1974. The present state of phage taxonomy. Intervirology 3, 201–219.

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

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, California, pp. 28–36.

Bamford, D.H., 2003. Do viruses form lineages across different domains of life? Res. Microbiol. 154, 231–236.

Besemer, J., Borodovsky, M., 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. 27, 3911–3920.

Bidnenko, E., Ehrlich, S.D., Chopin, M.C., 1995. Phage operon involved in sensitivity to the *Lactococcus lactis* abortive infection mechanism AbiD1. J. Bacteriol. 177, 3824–3829.

Bouchard, J.D., Moineau, S., 2000. Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. Virology 270, 65–75.

Brondsted, L., Ostergaard, S., Pedersen, M., Hammern, K., Vogensen, F.K., 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. Virology 283, 93–109.

Brüssow, H., Desiere, F., 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. 39, 213–222.

Brüssow, H., Hendrix, R.W., 2002. Phage genomics: small is beautiful. Cell 108, 13–16.

Casjens, S., Hatfull, G., Hendrix, R., 1992. Evolution of dsDNA-tailed bacteriophage genomes. Sem. Virol. 3, 383–397.

Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W., Cedergren, R., 1997. On the evolution of the single-subunit RNA polymerases. J. Mol. Evol. 45, 671–681.

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

Chopin, A., 1993. Organization and regulation of genes for amino acid biosynthesis in lactic acid bacteria. FEMS Microbiol. Rev. 12, 21–37.

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

Dunn, J.J., Studier, F.W., 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166, 477–535.

Dupont, K., Vogensen, F.K., Neve, H., Bresciani, J., Josephsen, J., 2004. Identification of the receptor-binding protein in 936-species lactococcal bacteriophages. Appl. Environ. Microbiol. 70, 5818–5824.

Fortier, L.-C., Bransi, A., Moineau, S., 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.

Huggins, A.R., Sandine, W.E., 1977. Incidence and properties of temperate bacteriophages induced from lactic streptococci. Appl. Environ. Microbiol. 33, 184–191.

Imburgio, D., Rong, M., Ma, K., McAllister, W.T., 2000. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. Biochemistry 39, 10419–10430.

Imburgio, D., Anikin, M., McAllister, W.T., 2002. Effects of substitutions in a conserved DX(2)GR sequence motif, found in many DNA-dependent nucleotide polymerases, on transcription by T7 RNA polymerase. J. Mol. Biol. 319, 37–51.

Jarvis, A.W., Fitzgerald, G.F., Mata, M., Mercenier, A., Neve, H., Powell, I.B., Ronda, C., Saxelin, M., Teuber, M., 1991. Species and type phages of lactococcal bacteriophages. Intervirology 32, 2–9.

Kenny, J.G., McGrath, S., Fitzgerald, G.F., van Sinderen, D., 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J. Bacteriol. 186, 3480–3491.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A., Hayward, R.S., 1993. The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. J. Mol. Biol. 232, 406–418.

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

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Lillehaug, D., 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. J. Appl. Microbiol. 83, 85–90.

Loessner, M.J., Gaeng, S., Scherer, S., 1999. Evidence for a holin-like protein gene fully embedded out of frame in the endolysin gene of *Staphylococcus aureus* bacteriophage 187. J. Bacteriol. 181, 4452–4460.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Lubbers, M.W., Waterfield, N.R., Beresford, T.P., Le Page, R.W., Jarvis, A.W., 1995. Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. Appl. Environ. Microbiol. 61, 4348–4356.

Lukashin, A.V., Borodovsky, M., 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26, 1107–1115.

Mahony, J., Deveau, H., McGrath, S., Ventura, M., Canchaya, C., Moineau, S., Fitzgerald, G.F., van Sinderen, D., 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.

Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., Krisch, H.M., 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. J. Bacteriol. 187, 3188–3200.

Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H., 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Res. 33, D192–D196.

McAllister, W.T., Raskin, C.A., 1993. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. Mol. Microbiol. 10, 1–6.

Miller, J.F., 2003. Bacteriophage and the evolution of epidemic cholera. Infect. Immun. 71, 2981–2982.

Moak, M., Molineux, I.J., 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol. Microbiol. 51, 1169–1183.

Moineau, S., Pandian, S., Klaenhammer, T.R., 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.

Pajunen, M.I., Elizondo, M.R., Skurnik, M., Kieleczawa, J., Molineux, I.J., 2002. Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3. J. Mol. Biol. 319, 1115–1132.

Pedersen, M., Ostergaard, S., Bresciani, J., Vogensen, F.K., 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology 276, 315–328.

Perler, F.B., Olsen, G.J., Adam, E., 1997. Compilation and analysis of intein sequences. Nucleic Acids Res. 25, 1087–1093.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., Desiere, F., Brussow, H., 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. 184, 6026–6036.

Reese, M.G., 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput. Chem. 26, 51–56.

Rohwer, F., 2003. Global phage diversity. Cell 113, 141.

Rohwer, F., Seguritan, V., Choi, D.H., Segall, A.M., Azam, F., 2001. Production of shotgun libraries using random amplification. Biotechniques 31, 108–118.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sandmeier, H., 1994. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers. Mol. Microbiol. 12, 343–350.

Saxelin, M.L., Nurmiaho, E.L., Korhola, M.P., Sundman, V., 1979. Partial characterization of a new C3-type capsule-dissolving phage of *Streptococcus cremoris*. Can. J. Microbiol. 25, 1182–1187.

Saxelin, M.L., Nurmiaho-Lassila, E.V., Merilaïnen, V.T., Forsen, R.I., 1986. Ultrastructure and host specificity of bacteriophages of *Streptococcus cremoris*, *Streptococcus lactis* subsp. *diacetylactis*, and *Leuconostoc cremoris* from Finnish fermented milk "Viili". Appl. Environ. Microbiol. 52, 771–777.

Schouler, C., Ehrlich, S.D., Chopin, M.C., 1994. Sequence and organization of the lactococcal prolate-headed bIL67 phage genome. Microbiology 140, 3061–3069.

Sousa, R., Chung, Y.J., Rose, J.P., Wang, B.C., 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. Nature 364, 593–599.

Spinelli, S., Campanacci, V., Blangy, S., Moineau, S., Tegoni, M., Cambillau, C., 2006a. Modular structure of the receptor binding proteins of *Lactococcus lactis* phages: the RBP structure of the temperate phage TP901-1. J. Biol. Chem. 281, 14256–14262.

Spinelli, S., Desmyter, A., Verrips, C.T., de Haard, H.J.W., Moineau, S., Cambillau, C., 2006b. Lactococcal bacteriophage p2 receptor binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat. Struct. Mol. Biol. 13, 85–89.

Staden, R., Beal, K.F., Bonfield, J.K., 2000. The Staden package, 1998. Methods Mol. Biol. 132, 115–130.

Stuer-Lauridsen, B., Janzen, T., Schnabl, J., Johansen, E., 2003. Identification of the host determinant of two prolate-headed phages infecting *Lactococcus lactis*. Virology 309, 10–17.

Terzaghi, B.E., Sandine, W.E., 1975. Improved medium for lactic streptococci and their bacteriophages. Appl. Microbiol. 29, 807–813.

Tétart, F., Desplats, C., Krisch, H.M., 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J. Mol. Biol. 282, 543–556.

Tock, M.R., Dryden, D.T., 2005. The biology of restriction and anti-restriction. Curr. Opin. Microbiol. 8, 466–472.

Vegge, C.S., Brondsted, L., Neve, H., Mc Grath, S., van Sinderen, D., Vogensen, F.K., 2005. Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. J. Bacteriol. 187, 4187–4197.

Vegge, C.S., Vogensen, F.K., Mc Grath, S., Neve, H., van Sinderen, D., Brondsted, L., 2006. Identification of the lower baseplate protein as the antireceptor of the temperate lactococcal bacteriophages TP901-1 and Tuc2009. J. Bacteriol. 188, 55–63.

Wang, I.N., Smith, D.L., Young, R., 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54, 799–825.