JOURNAL OF BACTERIOLOGY, Feb. 2008, p. 1390–1400 Vol. 190, No. 4
0021-9193/08/$08.00+0 doi:10.1128/JB.01412-07

Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*▽

Hélène Deveau,[1] Rodolphe Barrangou,[2] Josiane E. Garneau,[1] Jessica Labonté,[1] Christophe Fremaux,[3] Patrick Boyaval,[3] Dennis A. Romero,[2] Philippe Horvath,[3] and Sylvain Moineau[1]*

Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire, Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Quebec City, Quebec G1V 0A6, Canada[1]; Danisco, USA, Inc., 3329 Agriculture Drive, Madison, Wisconsin 53716[2]; and Danisco France SAS, BP10, F-86220 Dangé-Saint-Romain, France[3]

Received 31 August 2007/Accepted 21 November 2007

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated genes are linked to a mechanism of acquired resistance against bacteriophages. Bacteria can integrate short stretches of phage-derived sequences (spacers) within CRISPR loci to become phage resistant. In this study, we further characterized the efficiency of CRISPR1 as a phage resistance mechanism in *Streptococcus thermophilus*. First, we show that CRISPR1 is distinct from previously known phage defense systems and is effective against the two main groups of *S. thermophilus* phages. Analyses of 30 bacteriophage-insensitive mutants of *S. thermophilus* indicate that the addition of one new spacer in CRISPR1 is the most frequent outcome of a phage challenge and that the iterative addition of spacers increases the overall phage resistance of the host. The added new spacers have a size of between 29 to 31 nucleotides, with 30 being by far the most frequent. Comparative analysis of 39 newly acquired spacers with the complete genomic sequences of the wild-type phages 2972, 858, and DT1 demonstrated that the newly added spacer must be identical to a region (named proto-spacer) in the phage genome to confer a phage resistance phenotype. Moreover, we found a CRISPR1-specific sequence (NNAGAAW) located downstream of the proto-spacer region that is important for the phage resistance phenotype. Finally, we show through the analyses of 20 mutant phages that virulent phages are rapidly evolving through single nucleotide mutations as well as deletions, in response to CRISPR1.

Streptococcus thermophilus is one of the most industrially important lactic acid bacteria since it is widely used for the manufacture of yogurt and a number of cheeses. Several strains of this low G+C gram-positive species are used in large-scale milk fermentations because each strain possesses its own distinctive properties that are suitable for the manufacture of an array of fermented products, each with unique organoleptic properties. The systematic use of the same *S. thermophilus* strains in dairy processes has been impaired by the ubiquitous presence of virulent phages. Consequently, *S. thermophilus* bacteriophages have been the subject of extensive research in recent years with the aim of preventing their multiplication (8, 12).

S. thermophilus phages, like their hosts, are rather homogenous since they all belong to one polythetic species containing both temperate and virulent phages (9, 13). *S. thermophilus* phages are morphologically similar to coliphage lambda and accordingly belong to the *Siphoviridae* family. *S. thermophilus* phages are currently classified into two groups based on their general DNA packaging scheme (*cos* or *pac*) and the composition of their structural proteome (27). Seven complete genome sequences of *S. thermophilus* phages are publicly available, including those of the *cos*-type phages DT1, Sfi19, Sfi21, and 7201 and the *pac*-type phages O1205, Sfi11, and 2972 (28).

* Corresponding author. Mailing address: Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire, Université Laval, Quebec City, Quebec G1V 0A6, Canada. Phone: (418) 656-3712. Fax: (418) 656-2861. E-mail: Sylvain.Moineau@bcm.ulaval.ca.

▽ Published ahead of print on 7 December 2007.

Information on phage-host interactions has increased appreciably in recent years. It is well known that bacteria have a plethora of mechanisms to fight a diverse phage population (10). Traditionally in lactic acid bacteria, natural phage defense systems are divided in four main groups, namely, the inhibition of phage adsorption, the inhibition of DNA ejection, restriction-modification systems, and abortive infection (Abi) systems (10, 24). Globally, these mechanisms have been extensively studied in *Lactococcus lactis*, as well as in *Escherichia coli* (10). Unfortunately, few of these natural phage resistance mechanisms have been found in *S. thermophilus* (35). To cope with virulent phages and the lack of known defense mechanisms, the dairy industry has developed protocols to rapidly isolate bacteriophage-insensitive *S. thermophilus* mutants (BIMs) (39). These BIMs are spontaneous, naturally occurring phage-resistant descendants that survive exposure to virulent phages. Up until recently, the mechanism responsible for this resistance was often attributed to mutations in the phage receptors (2, 15).

The complete genome sequence of three *S. thermophilus* host strains is now available (4, 11, 31). Comparative analyses of these closely related *S. thermophilus* strains have revealed that genetic polymorphisms primarily occur at a few hypervariable regions, including three CRISPR loci (4, 5, 22, 31). These CRISPR loci have been found in a wide range of bacterial genomes (19, 23, 30). They are composed of 21 to 48-bp direct DNA repeats interspersed with nonrepetitive spacers of similar length. The direct repeats are highly conserved, while the number and sequence of the spacers are diverse, even among strains of a same species. Sequence similarities between spac-

TABLE 1. Phage sensitivity of isolated BIMs derived from *S. thermophilus* strain DGCC7710

Strain	Phage used for challenge	Sensitivity to phage[a]:		Source or reference
		2972	858	
DGCC7710		+	+	28
DGCC7710$_{\phi858}^{+S1S2}$	858	+	–	2
DGCC7710$_{\phi858}^{+S3}$	858	–	–	2
DGCC7710$_{\phi2972}^{+S4}$	2972	–	+	2
DGCC7710$_{\phi2972}^{+S5}$	2972	–	+	2
DGCC7710$_{\phi2972}^{+S6}$	2972	–	–	2
DGCC7710$_{\phi2972}^{+S7}$	2972	–	–	2
DGCC7710$_{\phi2972}^{+S8}$	2972	–	+	2
DGCC7710$_{\phi858\phi2972}^{+S9S10S11S12}$	2972 + 858	–	–	2
DGCC7710$_{\phi858\phi2972}^{+S13S14}$	2972 + 858	–	–	2
DGCC7710$_{\phi2972}^{+S15}$	2972	–	–	This study
Derivatives of DGCC7710$_{\Phi858}^{+S1S2}$				
DGCC7710$_{\phi858}^{+S1S2}{}_{\phi2972}^{+S16S17}$	2972	–	–	This study
DGCC7710$_{\phi858}^{+S1S2}{}_{\phi2972}^{+S18S17}$	2972	–	–	This study
DGCC7710$_{\phi858}^{+S1S2}{}_{\phi2972}^{+S19}$	2972	–	–	This study
Derivatives of DGCC7710$_{\Phi2972}^{+S4}$				
DGCC7710$_{\phi2972}^{+S4}{}_{\phi858}^{+S30}$	858	–	–	This study
DGCC7710$_{\phi2972}^{+S4}{}_{\phi858}^{+S32}$	858	–	–	This study
DGCC7710$_{\phi2972}^{+S4}{}_{\phi858}^{+S33}$	858	–	–	This study
Derivatives of DGCC7710$_{\Phi2972}^{+S6}$				
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}$	2972.S6B	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S30}$	2972.S6B	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S31}$	2972.S6B	–	–	This study
Derivatives of DGCC7710$_{\Phi2972}^{+S6}{}_{\Phi2972.S6B}^{+S20}$				
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S21S22}$	2972.S20A	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S23}$	2972.S20A	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S24}$	2972.S20A	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S25S26}$	2972.S20A	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S27}$	2972.S20A	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S28}$	2972.S20A	–	–	This study
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}{}_{\phi2972.S20A}^{+S29}$	2972.S20A	–	–	This study

[a] +, Sensitive to phage; –, insensitive to phage (EOP $\leq 10^{-4}$).

ers and extrachromosomal elements such as phages and plasmids led to the hypothesis that the CRISPR loci as well as CRISPR-associated genes (*cas*) play a role in protecting cells from the invasion of foreign DNA (5, 20, 30, 36, 37). In fact, it was recently demonstrated that CRISPR1/*cas* provides resistance against virulent phages in *S. thermophilus* (2).

We recently used *S. thermophilus* strain DGCC7710 and the virulent *pac*-type phages 2972 and 858 to show that CRISPR plays a role in the development of BIMs (2). Specifically, we found that in response to challenges with phage 858 and/or 2972, *S. thermophilus* DGCC7710 integrates new spacers derived from the phage genomes, generating a phage-resistant phenotype. The specificity of the resistance was determined by the identity between spacer and phage sequences (2). While the insertion of new spacers provided significant phage resistance, a small population of phages was able to infect the BIMs. This suggested that both CRISPR locus and phage genomic regions are subject to rapid evolutionary changes (2).

In the present study, we investigated the role of one of these CRISPR loci (CRISPR1) in phage-host interactions in *S. thermophilus* in greater detail. First, we demonstrated that this phage resistance mechanism is unique since it does not correspond to any known natural prokaryote antiviral barrier. Moreover, we analyzed BIMs derived from another *S. thermophilus* strain, namely, SMQ-301 (40), and show that its CRISPR1 locus can provide resistance against *cos*-type *S. thermophilus* phages. Finally, the homologous spacer region in the phage genome, which we propose to name proto-spacer, was analyzed for phage mutants infecting BIMs. Some of these phages showed a direct response to CRISPR1 by either a single nucleotide mutation or a deletion, in the proto-spacer region. Interestingly, in other phage mutants, a single nucleotide mutation was found in a short region (AGAA) that is located two nucleotides downstream of the proto-spacer sequence in the phage genome.

MATERIALS AND METHODS

Bacterial strains, phages, and microbiological assays. *S. thermophilus* host strains and their BIM derivatives (Tables 1 and 2) were grown in M17 broth supplemented with 0.5% lactose (LM17) at 42°C. Phages were propagated in LM17 supplemented with 10 mM calcium chloride. High-titer phage lysates were obtained as described elsewhere (28). The efficiency of plaquing (EOP) was determined by dividing the phage titer obtained by plating on a BIM by the titer obtained by plating the same phage on a sensitive host strain. Phage adsorption assays were performed as reported previously (16). Cell survival was assayed (3) by using a multiplicity of infection (MOI) of 5. For the efficiency of center of infection (ECOI) experiments, cells in exponential-phase (i.e., an optical density

TABLE 2. Phage sensitivity of isolated BIMs derived from *S. thermophilus* strain SMQ-301

Strain	Phage used for challenge	Sensitivity to phage DT1[a]	Source or reference
SMQ-301		+	40
SMQ-301$_{\phi DT1}^{+S34S35}$	DT1	–	This study
SMQ-301$_{\phi DT1}^{+S36S37}$	DT1	–	This study
SMQ-301$_{\phi DT1}^{+S38S36S37}$	DT1	–	This study
SMQ-301$_{\phi DT1}^{+S39}$	DT1	–	This study

[a] +, Sensitive to phage; –, insensitive to phage (EOP $\leq 10^{-4}$).

at 600 nm of 0.6) were infected by using an MOI of 5. Phages were first allowed to adsorb for 15 min, and then the unbound phages were removed by a quick centrifugation. The pellet of infected cells was washed twice with fresh LM17 broth. ECOI formation was calculated by dividing the phage titer obtained by plating resistant infected cells by the titer obtained with the sensitive infected strain. One-step growth curves (32) were performed by using an MOI of 0.2. The burst size was determined by dividing the average titer after the rise period by the average titer before the bacteria began to release virions. For each microbiological test, the mean value and the standard deviation were calculated from three independent experiments.

Isolation of BIMs. BIMs were obtained by challenging sensitive *S. thermophilus* strains with virulent phages 2972 (28), 858 (28), or DT1 (40) or with mutant phages (see below). BIMs were also obtained by challenging *S. thermophilus* strains with a mixture of phages 2972 and 858 at a ratio of 1:1. Briefly, 100 μl of an overnight culture of *S. thermophilus* was used to inoculate 10 ml of LM17, which was incubated at 42°C until the optical density at 600 nm reached 0.3. Phages and calcium chloride were then added at final concentrations of 10^7 PFU/ml and 10 mM, respectively. The phage-containing culture was incubated at 42°C for 24 h and monitored for lysis. A total of 100 μl of the lysed culture was then used to inoculate 10 ml of fresh LM17. The remaining lysate was centrifuged, and the pellet was transferred into another tube containing 10 ml of LM17 broth. These two cultures were incubated at 42°C for 16 h, serially diluted (1/10), and plated on LM17. The phage sensitivity of the isolated BIMs was first estimated by a spot test (33).

Isolation and characterization of mutant phages. All phage mutants were single-plaque purified three times and propagated as previously described (34). Twenty well-defined plaques from four BIMs were isolated and analyzed. The individual phages were propagated on the BIM used to isolate them and were designated by the name of the parental wild-type phage followed by the new spacer number in the host strain. A different letter was added to identify each distinct phage mutant. Phage DNA was isolated as described previously from 1 ml of phage lysate (34). Restriction endonucleases (Roche Diagnostics) were used as recommended by the manufacturer. When necessary, restricted phage DNA samples were heated for 10 min at 70°C to prevent cohesive end ligation. The DNA fragments were separated on 0.8% agarose gels in 1× TAE buffer and UV visualized after staining with ethidium bromide.

DNA sequencing of the 858 phage genome. Genomic DNA of virulent phage 858 was isolated by using Qiagen Lambda Maxi kit with previously described modifications (14). The primers used to sequence the genome of virulent phage 2972 (28) were used to begin sequencing the genome of the closely related phage 858, using isolated phage DNA as a template. New primers were designed from the nucleotide sequence, and primer walking on the two DNA strands was used to complete the sequencing of the genome. An ABI Prism 3700 at the genomic platform of the Centre Hospitalier de l'Université Laval was used for the sequencing. ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi) were used for DNA sequence searches (29). PSI-BLAST and Advanced BLAST Search 2.1 were also used for sequence comparisons (http://www.ncbi.nlm.nih.gov/BLAST) (1). Only the best match is presented.

Sequencing the CRISPR1 locus in *S. thermophilus* strains and the proto-spacer region in the phage genomes. The CRISPR1 locus of the BIMs (including repeats and spacers) and the proto-spacer region in the phage genomes were sequenced from PCR products. Total bacterial DNA was prepared as described previously (21). Computer-assisted DNA analyses of the CRISPR loci were performed by using the Staden Package (38) (http://staden.sourceforge.net/), CLUSTALW (http://www.ebi.ac.uk/clustalw/), and DNA Display (http://www.mrgtech.ca/DNA).

Nucleotide sequence accession numbers. The complete genomic sequences of the wild-type phages analyzed here are available under the indicated GenBank accession numbers: DT1 (AF085222), 2972 (AY699705), and 858 (EF529515).

RESULTS

Characterization of the phage resistance system. Phage resistance mechanisms are typically characterized by using a series of microbiological assays to determine their general mode of action. Wild-type virulent phage 2972, phage-sensitive strain *S. thermophilus* DGCC7710, and previously isolated *S. thermophilus* BIMs DGCC7710$_{\phi 858}^{+S3}$, DGCC 7710$_{\phi 2972}^{+S4}$, and DGCC7710$_{\phi 2972}^{+S4}$ϕ858^{+S32} (2) were selected for these assays (Table 1 and Fig. 1). Strains DGCC7710$_{\phi 858}^{+S3}$ and DGCC7710$_{\phi 2972}^{+S4}$ are BIM derivatives of DGCC7710 that have acquired a single new spacer in their respective CRISPR1 locus after a phage challenge, while DGCC7710$_{\phi 2972}^{+S4\phi 858+S32}$ is a second-generation BIM derived from the first-generation BIM DGCC7710$_{\phi 2972}^{+S4}$, and thus it has acquired a second spacer after a second phage challenge.

Phage adsorption assays showed that phage 2972 adsorbed at the same level (90%) to phage-sensitive and phage-resistant strains (Table 3), indicating that the addition of a new spacer in CRISPR1 did not prevent phage adsorption. The isolation of second generation of CRISPR BIMs through the addition of new spacers (originating from the phage genome) precludes that this defense mechanism prevents the ejection of the phage DNA into the cell. Ruling out restriction-modification systems was the isolation of two mutants of phage 2972 from DGCC7710$_{\phi 858}^{+S3}$ or DGCC7710$_{\phi 2972}^{+S4}$ (at a frequency of 10^{-5}) that propagate on their respective BIM hosts and wild-type DGCC7710 with equal efficiency (EOP = 1.0), even after passage through DGCC7710 (data not shown). One characteristic feature of Abi mechanisms is the high cell mortality following the abortion of phage infection (10). Cell survival assays showed that between 64 to 73% of the phage-resistant *S. thermophilus* strains survive the phage infection (Table 3), indicating that CRISPR1 is not an Abi mechanism (10). Interestingly, some infected cells still released virions (Table 3). The burst size, however, was significantly reduced, decreasing from 190 new phages per sensitive host cell (DGCC7710) to between 6 and 28 new virions, depending on the infected BIM strain (Table 3). Of note, the phage latency period was similar in the sensitive and first-generation BIMs (between 39 to 44 min at 42°C) but longer in the second-generation BIM (Table 3). Collectively, these results indicate that CRISPR1 is indeed a novel phage resistance system.

Iterative addition of spacers. It has been shown that the addition of a new spacer in the CRISPR1 locus can increase the phage resistance of a particular strain (2). However, the replication of a small number of phages still occurs (Table 3). It has also been reported that the random acquisition of multiple spacers (up to 4) after only one phage challenge can lead to increased phage resistance (2). We wanted to determine whether successive phage challenges and the subsequent iterative addition of spacers could lead to even greater phage protection.

For these experiments, we used *S. thermophilus* DGCC7710 derivatives that are still sensitive to virulent phage 2972

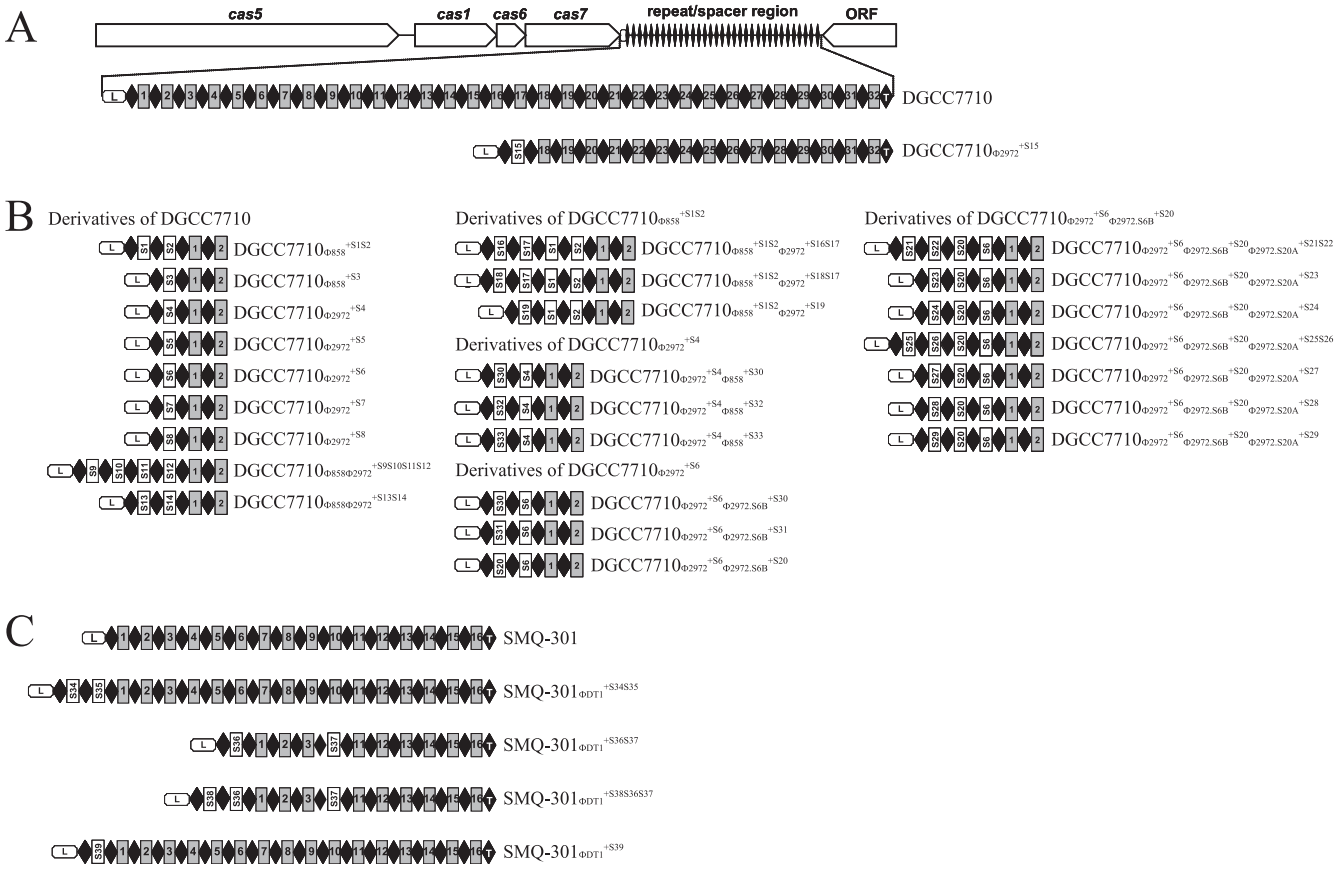

FIG. 1. *S. thermophilus* CRISPR1 locus overview and newly acquired spacers in various phage-resistant mutants. (A) Repeat/spacer region of strain DGCC7710 and the selected BIM named DGCC7710$_{\phi2972}{}^{+S15}$. Repeats are shown as black diamonds, spacers are numbered in gray boxes, and the leader (L) is shown as a white box. The terminal repeat of CRISPR1 locus is represented with a letter "T" inside the black diamond. (B) The spacer content at the 5′ end of the locus in various phage-resistant mutants is represented. The newly added spacers are indicated in white boxes with a designation containing the prefix S, followed by a number. The 3′ end of CRISPR1 in BIMs is identical to that of the wild-type strain. (C) Repeat/spacer region of *S. thermophilus* strain SMQ-301 and its derivatives.

or virulent phage 858. We selected *S. thermophilus* DGCC7710$_{\phi2972}{}^{+S4}$ obtained after a challenge with phage 2972 (but still sensitive to 858) and *S. thermophilus* DGCC7710$_{\phi858}{}^{+S1S2}$ obtained after a challenge with phage 858 (but still sensitive to 2972) (2). Both BIMs were infected with the appropriate virulent phage (DGCC7710$_{\phi2972}{}^{+S4}$ with 858 and DGCC7710$_{\phi858}{}^{+S1S2}$ with 2972) using the BIM isolation procedure previously described. For each challenge, several new BIMs were obtained, and three were selected for further characterization. In all cases, one or two new spacers were acquired at the leader end of the CRISPR1 locus (Table 1), which provided resistance to the second phage (Table 4). Interestingly, the EOP of the phage 2972 was decreased with the addition of a second spacer identical to the phage 2972 in the strain DGCC7710$_{\phi2972}{}^{+S4\phi858+S32}$, indicating that accumulating spacers can increase the overall phage resistance of the host (Table 4).

Moreover, *S. thermophilus* DGCC7710$_{\phi2972}{}^{+S6}$, a BIM resistant to both wild-type phages 2972 and 858 (Table 1), was challenged with a mutant phage (2972.S6B, Fig. 2) that is capable of bypassing the CRISPR1-mediated resistance due to a single nucleotide mutation in the S6 targeted region of its

TABLE 3. Microbiological effects of the new spacer presence on the bacteriophage 2972

Strain	Mean ± SD[a]					
	EOP	Adsorption (%)	ECOI (%)	Cell survival (%)	Burst size	Latent period (min)
DGCC7710	1	89.3 ± 2.6	100	<1	190 ± 33	40 ± 3
DGCC7710$_{\phi858}{}^{+S3}$	$(1.7 \pm 0.9) \times 10^{-5}$	94.3 ± 6.0	18.6 ± 1.9	72.5 ± 3.5	7 ± 2	39 ± 2
DGCC7710$_{\phi2972}{}^{+S4}$	$(1.8 \pm 2.0) \times 10^{-5}$	90.3 ± 5.3	19.7 ± 3.7	64.7 ± 5.2	28 ± 2	44 ± 4
DGCC7710$_{\phi2972}{}^{+S4}{}_{\phi858}{}^{+S32}$	$<10^{-7}$	91.9 ± 6.4	10.8 ± 8.1	66.1 ± 10.5	6 ± 4	55 ± 5

[a] Results are the means of three independent experiments.

TABLE 4. EOPs of the wild-type phages on various *S. thermophilus* strains

S. thermophilus strain	No. of spacers identical to phage:		Mean EOP ± SD[a] for phage:	
	2972	858	2972	858
DGCC7710	0	0	1.0	1.0
DGCC7710$_{\phi2972}^{+S4}$	1	0	$(1.8 \pm 2.0) \times 10^{-5}$	0.87 ± 0.04
DGCC7710$_{\phi2972}^{+S4}{}_{\phi858}^{+S32}$	2	1	$<10^{-7}$	$(2.4 \pm 1.3) \times 10^{-5}$
DGCC7710$_{\phi858}^{+S1S2}$	0	1	0.67 ± 0.45	$(9.3 \pm 9.3) \times 10^{-6}$
DGCC7710$_{\phi858}^{+S1S2}{}_{\phi2972}^{+S19}$	1	1	$(3.1 \pm 1.7) \times 10^{-4}$	$(1.6 \pm 1.3) \times 10^{-6}$

[a] Results are the means of three independent experiments.

genome (see below). Three new BIMs that were resistant to phage 2972.S6B were obtained and their CRISPR1 loci analyzed. These second-generation BIMs had also acquired a new and unique spacer at the leader end of their CRISPR1 (Table 1). Subsequently, a second-generation BIM (DGCC 7710$_{\phi2972}^{+S6\phi2972.S6B+S20}$) was challenged with another phage mutant (2972.S20A), resulting in a set of third-generation BIMs. Seven distinct third-generation BIMs were obtained, each of them had acquired one or two new spacers at the leader end of CRISPR1 (Table 1 and Fig. 1B). Taken altogether, these data clearly indicate that iterative addition of spacers is possible, resulting in increased phage resistance of these isogenic *S. thermophilus* strains. It also confirms that CRISPR loci, as well as phages, rapidly change in response to each other.

CRISPR1 analysis of another phage-host system in *S. thermophilus*. Until now (see above and reference 2), the study of *S. thermophilus* CRISPR1 was performed with a single host and two phages, i.e., *S. thermophilus* DGCC7710 and virulent *pac*-type phages 2972 and 858 (and their phage mutants). To determine whether the main conclusions described above apply to another *S. thermophilus* phage-host system, we isolated BIMs from *S. thermophilus* SMQ-301 challenged with the virulent *cos*-type phage DT1 (40), one of the best-characterized *cos*-type phages in *S. thermophilus* (15, 16, 17, 26). The analysis of the CRISPR1 locus of wild-type phage-sensitive strain *S. thermophilus* SMQ-301 revealed the presence of 16 unique spacers, half the number found in DGCC7710. It should be noted that the spacers of SMQ-301 are distinct from those in DGCC7710 (22). It was previously reported that BIMs from *S.*

	New spacer	Change in Codon	Amino acid
DGCC7710$_{\phi858}^{+S3}$	TTACGTTTGAAAAGAATATCAAATCAATGA		
Phage 2972	TTACGTTTGAAAAGAATATCAAATCAATGACGAGAAAGA		
Phage 2972.S3A	TTACGTTTGAAAAGAATATCAAATTAATGACGAGAAAGA	TCA - TTA	S - L
Phage 2972.S3B	TTACGTTTGAAAAGAATATCAAATCAACGACGAGAAAGA	ATG - ACG	M - T
Phage 2972.S3C	TTACGTTTGAAAAGAATATCAAATCAATGACGAGAGAGA	AAG - GAG	K - E
Phage 2972.S3D	TTACGTTTGAAAAGAATATCAAATCTATGACGAGAAAGA	TCA - TCT	S - S
Phage 2972.S3E	TTACGTTTGAAAAGAATATCAATTCAATGACGAGAAAGA	AAA - AAT	K - N
Phage 2972.S3F	TTACGTTTGAAAAGAATATCAAATTAATGGCGAGAAAGA	TCA - TTA	S - L
	TTACGTTTGAAAAGAATATCAAATTAATGGCGAGAAAGA	ACG - GCG	T - A
Phage 2972.S3G	TTACGTTTGAAAAGAACCATCAAATTAATGACGAGAAAGA	AAT - AAC	N - N
	TTACGTTTGAAAAGAACATCAAATTAATGACGAGAAAGA	TCA - TTA	S - L
DGCC7710$_{\phi2972}^{+S4}$	CTCAGTCGTTACTGGTGAACCAGTTTCAAT		
Phage 2972	CTCAGTCGTTACTGGTGAACCAGTTTCAATTGAGAAAAA		
Phage 2972.S4A	CTCAGTCGTTACTGGTGAACCAGTTTCAATTGAAAAAAA	GAG - GAA	E - E
Phage 2972.S4B	CTCAGTCGTTACTGGTGAACCAGTTTCGATTGAGAAAAA	TCA - TCG	S - S
Phage 2972.S4C	CTCAGTCGTTACTGGTGAACCAGTTTCAATTGAGAGAAA	AAA - AGA	K - R
Phage 2972.S4D	CTCAGTCGTTACTGGTGAACCGGTTTCAATTGAAAAAAA	CCA - CCG	P - P
	CTCAGTCGTTACTGGTGAACCGGTTTCAATTGAAAAAAA	GAG - GAA	E - E
DGCC7710$_{\phi2972}^{+S6}$	GCCCTTCTAATTGGATTACCTTCCGAGGTG		
Phage 2972	GCCCTTCTAATTGGATTACCTTCCGAGGTGTTAGAATTC	Minus strand	
Phage 2972.S6A	GCCCTTCTAATTGGATTACCTTCCGAGGTGTTAGAGTTC	ATT - ACT	I - T
Phage 2972.S6B	GCCCTTCTAATTGGATTACCTTCCGATGTGTTAGAATTC	CCT - CAT	P - H
Phage 2972.S6C	GCCCTTCTAATTGGATTACCTTCCGAGTTGTTAGAATTC	CCT - ACT	P - T
Phage 2972.S6D	GCCCTTCTAATTGGATTACCTTCCGA*GTGTTAGAATTC	Frameshift mutation	
DGCC7710$_{\phi2972}^{+S4}{}_{\phi858}^{+S32}$	ATTGTCTATTACGACAACATGGAAGATGAT		
Phage 858	ATTGTCTATTACGACAACATGGAAGATGATGTAGAAATT		
Phage 858.S32A	ATTGTCTATTACGACAACATGGAAGATGATGTATAAATT	GAA - TAA	E - Stop
Phage 858.S32B	ATTGTCTATTACGACAACATGGAAGATTATGTAGAAATT	GAT - TAT	D - Y
Phage 858.S32C	**********ATT	Start codon is deleted	
Phage 858.S32D	ATTGTCTATTACGACAACATGGAAGATGATGTAAAAATT	GAA - AAA	E - K
DGCC7710$_{\phi2972}^{+S6}{}_{\phi2972.S6B}^{+S20}$	TTATATCGAAGAACGACTGAAAGAGCTTGA		
Phage 2972	TTATATCGAAGAACGACTGAAAGAGCTTGAGAAGAAAAA		
Phage 2972.S20A	TTATATCGAAGAACGACTGAAAGAGCTTGAGAATAAAAA	AAG - AAT	E - N

FIG. 2. Nucleotide sequences in wild-type and mutant phages that correspond to the newly acquired spacers by the *S. thermophilus* strains. The AGAA motif is highlighted in gray. Each mutation is in boldface and underlined. *, Deletion.

thermophilus SMQ-301 are difficult to obtain (15); however, using the improved protocol described here, four BIMs resistant to DT1 were obtained and analyzed (Table 2 and Fig. 1C). In all cases, one to three additional spacers derived from the genome of DT1 were inserted into the CRISPR1 locus (Table 5 and Fig. 3). These findings confirm that CRISPR-mediated phage resistance can protect *S. thermophilus* against representatives of the two main groups of phages and that it operates through a general mechanism of action.

Analysis of the newly acquired spacers in the CRISPR1 locus of *S. thermophilus* BIMs. The spacer content of 26 BIMs derived from strain *S. thermophilus* DGCC7710 (generated in the present study and previously described [2]), as well as of the 4 BIMs from *S. thermophilus* SMQ-301, was analyzed (Table 1 and Fig. 1). Of the 30 analyzed BIMs, 21 had acquired a single new spacer, seven had acquired two new spacers, one had acquired three new spacers, and one had acquired four new spacers. Thus, the addition of a single new spacer in CRISPR1 appeared to be a common outcome of a phage challenge.

The original 32 spacers in DGCC7710 were conserved in all but one first-generation BIM, namely, *S. thermophilus* $DGCC7710_{\phi 2972}{}^{+S15}$. Interestingly, this BIM had acquired a new spacer at the leader end but lost the first 17 spacers present in wild-type strain *S. thermophilus* DGCC7710. Moreover, two of the four BIMs of SMQ-301 had also lost seven of the original spacers (spacers 4 through 10), suggesting that spacer deletion may occur concomitantly with the addition of new spacers (Table 2 and Fig. 1C).

All 30 analyzed BIMs acquired at least one new spacer at the leader end of CRISPR1. Surprisingly, in two BIMs of SMQ-301, a second new spacer was also added after the third original spacer. Of note, these two BIMs were also the ones that had lost seven of the original spacers. Thus, the addition of new spacers is clearly polarized toward one end of the CRISPR1 locus, and the acquisition of new spacers within CRISPR1 is also possible, albeit rare.

A total of 33 new spacers were acquired by the 26 BIMs derived from *S. thermophilus* DGCC7710 challenged with virulent phage 2972 or 858. In addition, six distinct spacers were acquired by the four BIMs of *S. thermophilus* SMQ-301 that was challenged with phage DT1. Analysis of these 39 new spacers showed that 32 of them were 30 nucleotides long (Table 5). Five spacers were 29 nucleotides long, while the remaining two spacers were 31 nucleotides long. Evidently, the addition of a 30-nucleotide-long spacer is the most frequent event, in agreement with the observation that the vast majority of CRISPR1 spacers have a 30-bp length (22).

Comparison of the 39 new spacers with three *S. thermophilus* phage genomes. To compare the newly acquired spacers with the genomic regions of the corresponding phages used in the challenge experiments, the complete genome of virulent phage 858 was sequenced (Table 6). The genomes of phages 2972 and DT1 were previously determined (28, 40). The annotation of phage 858 genome is presented in Table 6. As expected, it is highly related to other *pac*-type phages of *S. thermophilus* (28). The 858 and 2972 phage genomes share 90.9% nucleotide identity. Briefly, its linear double-stranded DNA contains 35,543 bp with an overall G+C content of 39.8%. Only 5 of the 46 predicted open reading frames (ORFs) of phage 858 did not have close homologs in other *S. thermophilus* phages. In fact, three of them (ORF38, ORF39, and ORF40) were closer to deduced ORFs from the genome of *Streptococcus suis* 89/1591.

Using the complete genomic sequences of phages 858, 2972, and DT1, we performed comparative analyses with the 39 newly added spacers found in the 30 analyzed *S. thermophilus* BIMs. The nucleotide sequence of 37 spacers out of 39 was 100% identical to a specific region found in the genome of at least one wild-type phage used in the challenge experiments (Table 5 and Fig. 3A). Spacers S2 and S26 had one mismatch with the proto-spacer in the phage genomes (Table 5). However, the BIMs containing these two mismatched spacers had also acquired other spacers that were identical to a phage genomic region.

Further analyses of the phage genomes indicated that all 39 new spacers analyzed in the present study correspond to a predicted coding region. Moreover, the spacer sequences covered all phage modules as well as both strands. However, the new spacers originated most often from the coding strand than the noncoding strand (28 of 39 spacers from the coding strand [71.7%]), and about half of them were localized in the early expressed region of the phage genome (22 of 39 spacers [56.4%]), although this latter region corresponded to only 27 to 31% of the phage genome (17) (Table 5 and Fig. 3A). Interestingly, some spacers (S30, S36, and S37) were independently acquired by two BIM strains (Table 1).

Comparative analyses of the regions flanking the proto-spacers in the phage genomes led to the identification of a specific sequence that was always located two nucleotides (NN) downstream from the proto-spacers (Table 5). This CRISPR1-specific sequence corresponds to the motif described recently (22). In fact, 34 of the 39 proto-spacers had the 3′-flanking AGAAW motif. The other five proto-spacers (corresponding to spacers S2, S11, S13, S35, and S38) had one mismatched nucleotide in the AGAAW motif. However, these five spacers were found in BIMs that had acquired multiple spacers after the phage challenge. To determine whether the strand and temporal expression biases noted above could be explained by the presence of the AGAAW motif, the distribution of this conserved sequence was analyzed in the genome of the three phages 2972, 858, and DT1 (Fig. 3B). The AGAAW motif was found almost three times more frequently on the coding strand than on the noncoding strand, with average values for the three *S. thermophilus* phages of 5.0 AGAAW/kb on the coding strand and 1.7 AGAAW/kb on the noncoding strand. These results suggest that spacer acquisition may not be random and that there may be a limited number of proto-spacers to be included in CRISPR1. On the other hand, between 36 and 40% of these motifs were found in the early expressed modules, while 56.4% of the acquired spacers correspond to this region. Thus, the proportion of the AGAAW motifs in the different transcription modules cannot totally explain the observed bias for the early expressed region.

Phage response to the acquisition of a new spacer. As indicated elsewhere (2), phage mutants capable of infecting newly generated BIMs can be isolated under laboratory conditions. The characterization of phage mutants obtained through the selective pressure of resistance systems is particularly useful, since it has previously led to a better understanding of novel phage defense mechanisms (6, 25, 41). Using a similar ap-

TABLE 5. List of new spacers found in CRISPR1 and the corresponding region in phages 2972, 858, and DT1

Spacer	Phage[a]	5′ position[b]	Spacer length (pb)	Proto-spacer sequence[c]	3′ flanking region[d]	Strand/module[e]	ORF/function in the genome of the phage used in the challenge
S1	858	31378	30	CAACACATTCAACAGATTAATGAAGAATAC	AA**AGAAA**AAA	(+)/E	ORF40/primase
S2	2972*	25432	30	TCCACTCACGTACAAATAGTGAG*C*GTACTC	CT*A*A**AAG**GAT	(–)/L	ORF27/unknown
S3	2972*	17202	30	TTACGTTTGAAAAGAATATCAAATCAATGA	CG**AGAAA**GAT	(+)/L	ORF20/receptor-binding protein
S4	2972	31582	30	CTCAGTCGTTACTGGTGAACCAGTTTCAAT	TG**AGAAA**AAA	(+)/E	ORF38/primase
S5	2972	22075	30	AGTTTCTTTGTCAGACTCTAACACAGCCGC	TC**AGAAA**GTT	(+)/L	ORF21/tail protein
S6	2972*	34521	30	GCCCTTCTAATTGGATTACCTTCCGAGGTG	TT**AGAAT**TCC	(–)/E	ORF44/unknown
S7	2972*	10299	30	AAGCAAGTTGATATATTTCTCTTTCTTTAT	TA**AGAAA**ACG	(–)/L	ORF17/unknown
S8	2972	30016	29	CGTTTTCAGTCATTGGTGGTTTGTCAGCG	AA**AGAAA**TAA	(–)/E	ORF37/replication
S9	2972*	7874	30	TTACTAGAGCGTGTCGTTAACCACTTTAAA	TC**AGAAT**ATG	(+)/M	ORF11/unknown
S10	2972*	20650	30	TTCGTTAAAGTCACCTCGTGCTAGCGTTGC	AT**AGAAA**GTT	(–)/L	ORF20/receptor-binding protein
S11	2972*	8360	30	ATAACGGTAGCAAATATAAACCTGTTACTG	TC**AGAA***G*CTA	(+)/M	ORF12/unknown
S12	2972[a]	18998	30	GAAGTAGCCATACAAGAAGATGGATCAGCA	CC**AGAAA**TTG	(+)/L	ORF20/receptor-binding protein
S13	2972*	33602	30	GATGTCACTGAGTGTCTAAGCATTGCGTAC	GA*G***GAAA**TCA	(+)/E	ORF42/DNA binding
S14	2972*	4830	30	TGAATAAGCAGTTCTTGACGACCAACCGAC	AT**AGAAA**AGT	(–)/M	ORF6/capsid protein
S15	2972*	34444	29	CAATTAACACAGCAATTAACACAGTATAT	AC**AGAAA**TTG	(+)/E	ORF44/unknown
S16	2972*	6799	30	ATGCCATTCTTTAAAGAGGCTTTACTCGTT	AA**AGAAA**ACG	(+)/M	ORF9/capsid protein
S17	2972	30547	30	GTTGGCGGACTACTCCTTCGAGGGGTTGAT	CC**AGAAA**TTA	(+)/E	ORF37/replication
S18	2972	30370	29	GAAGCACCTCTTGCGTTGATAAAAGTATT	GC**AGAAA**ATG	(+)/E	ORF37/replication
S19	2972	31709	29	ACATATCGACGTATCGTGATTATCCCATT	CA**AGAAA**ACA	(+)/E	ORF38/primase
S20	2972*	1113	30	TTATATCGAAGAACGACTGAAAGAGCTTGA	GA**AGAAA**AAA	(+)/M	ORF2/small terminase
S21	2972*	19188	30	AAATCAACGTACATCCCGATATAGGCACGA	TT**AGAAT**CAG	(–)/L	ORF20/receptor-binding protein
S22	2972	31708	30	GACATATCGACGTATCGTGATTATCCCATT	CA**AGAAA**ACA	(+)/E	ORF38/primase
S23	2972	26529	31	TGAAGTATTAGGTCTCTCAAAAGATGATATT	GT**AGAAT**ACT	(+)/E	ORF31/Cro-like repressor
S24	2972	29923	30	AGTTGATTGCGTAATCAACCATCTCCATAA	TT**AGAAT**GGA	(–)/E	ORF37/replication
S25	2972*	441	30	GCAACACTCAAACGTTGCAAACGCAAGCTT	CG**AGAAT**ATC	(+)/E	ORF1/unknown
S26	2972	31606	31	CTCAGTCGTTACTGGTGAACCAGTT***TCAAT	TG**AGAAA**AAA	(+)/E	ORF38/primase
S27	2972*	27032	30	TTTCATCGTCAATTTCCATGTTATAAATCT	CT**AGAAA**CTG	(–)/E	ORF33/unknown
S28	2972	26530	30	GAAGTATTAGGTCTCTCAAAAGATGATATT	GT**AGAAT**ACT	(+)/E	ORF31/cro-like repressor
S29	2972	32136	29	ATTGGCATGATTTCAATTTTAATTGGGAT	GT**AGAAA**AAG	(+)/E	ORF38/primase
S30	2972*	33968	30	TCCAAGTTATTTGAGGAGTTATTAAGACAT	GA**AGAAA**TAT	(+)/E	ORF43/unknown
S31	2972	30803	30	TACCGAAACGACTGGTTTGAAAAATTCAAG	GA**AGAAA**ATC	(+)/E	ORF38/primase
S32	2972*	33044	30	ATTGTCTATTACGACAACATGGAAGATGAT	GT**AGAAA**TTT	(+)/E	ORF41/unknown
S33	858	30335	30	CTTCAAATGTACTGCAAGGCTGCAAAAGTA	CC**AGAAA**ATA	(+)/E	ORF38/unknown
S34	DT1	14535	30	GCTACTGAAAGCTACGAGGTTGGTAATCCT	AA**AGAAT**GGG	(+)/L	ORF17/tail protein
S35	DT1	13255	30	GTAGTTAGAGCGCTTGAAGCTAACGGTATA	GA*A**CC***AA**ACA	(+)/L	ORF15/tail protein
S36	DT1	29132	30	TTAGATCTCATGAGTGGCGACAGTGAGCTT	GT**AGAAT**TAC	(+)/E	ORF36/primase
S37	DT1	20837	30	AACGATGAGGAACTCTTGGCAAAACTTACA	CA**AGAAT**AGC	(+)/L	ORF22/unknown
S38	DT1	9893	30	GCATTCATGGTTTGTTGGTATTTAACGTAT	TC*G***GAA***C*TGG	(–)/L	ORF15/tail protein
S39	DT1	2603	30	TATTTTATCAGTCATCATGGCGTCATAGCC	GA**AGAAA**ACG	(–)/M	ORF4/Large terminase

[a] *, DNA regions that are 100% identical between phages 858 and 2972.
[b] That is, the 5′ position of the proto-spacer in the phage genome.
[c] Underlined and italicized nucleotides indicate a mismatch between the phage and the spacer. An asterisk indicates a deletion.
[d] That is, the 3′ flanking sequence in the phage genome. A mismatch in the AGAAW motif is boldfaced.
[e] Transcription module: E, early expressed genes; M, middle expressed genes; L, late expressed genes.

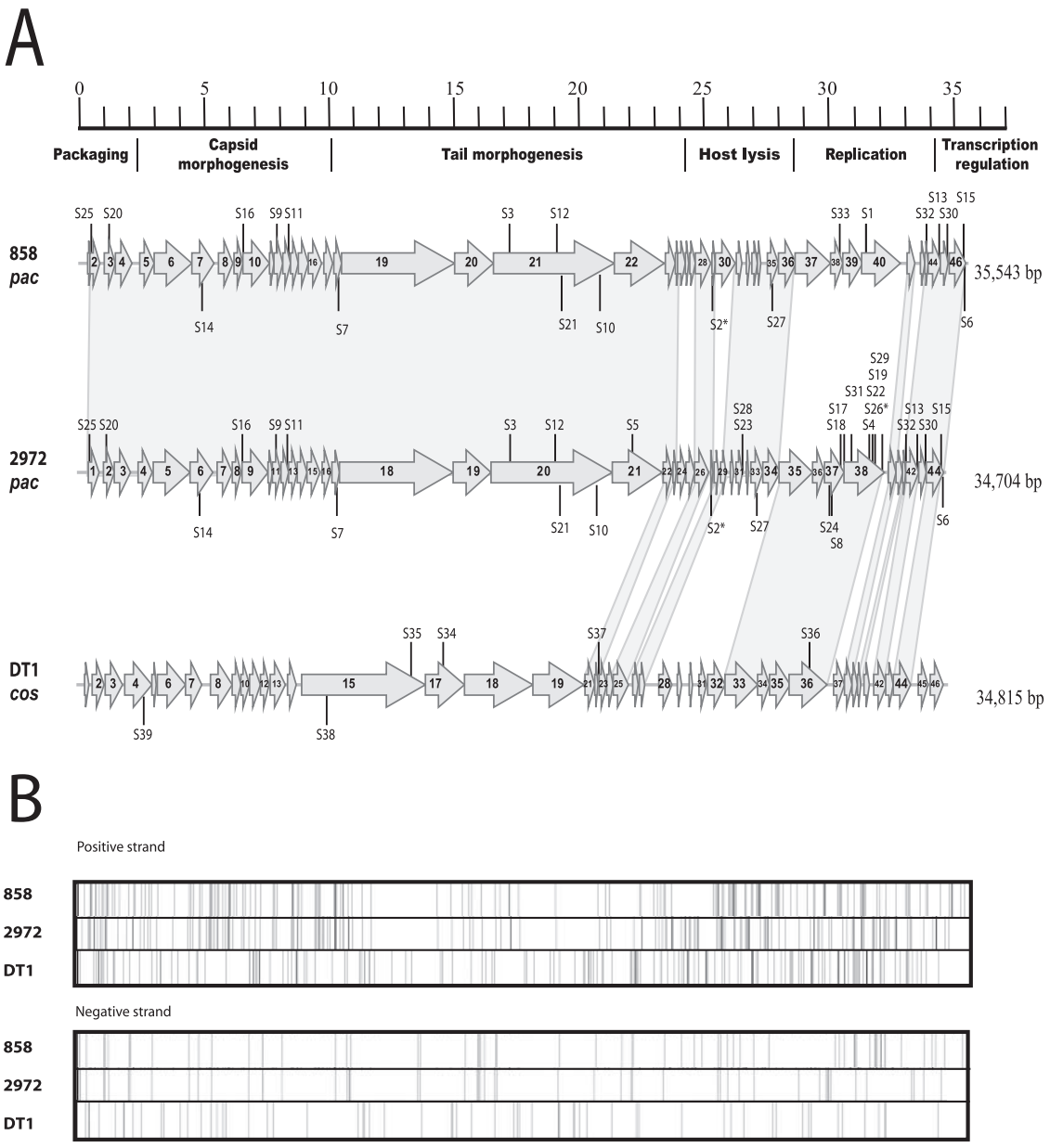

FIG. 3. Schematic representation of the *S. thermophilus* bacteriophage genomes used in the present study. (A) Distribution of the sequences corresponding to the new spacers in the three phage genomes. The spacers above the genome correspond to the positive strand, while those indicated in lower part correspond to the negative strand. Spacers indicated by an asterisk contain one mismatch with the phage sequence. ORFs connected by a gray box possess more than 70% identity at the amino acid level. (B) Distribution of the AGAAW motif on both strands for the three phages.

proach, 20 phage mutants that infect *S. thermophilus* BIMs DGCC7710$_{\phi858}^{+S3}$ (7 mutants of 2972), DGCC7710$_{\phi2972}^{+S4}$ (4 mutants of 2972), DGCC7710$_{\phi2972}^{+S6}$ (4 mutants of 2972), DGCC7710$_{\phi2972}^{+S4\phi858+S32}$ (4 mutants of 858), and DGCC7710$_{\phi2972}^{+S6\phi2972.S6B+S20}$ (1 mutant of 2972) were further characterized (Fig. 2). All of the phage mutants had the same restriction profiles as the wild-type phages (data not shown). The proto-spacers, as well as their flanking regions (approximately 100 pb upstream and downstream), were sequenced in the mutants. Four distinct types of mutations were observed in these mutants: (i) a single-nucleotide mutation directly within the proto-spacer (8 of 20 mutant phages), (ii) a two-nucleotide mutation directly in the proto-spacer (3 of 20 mutant phages), (iii) a single nucleotide mutation in the AGAA flanking sequence (7 of 20 mutant phages), and (iv) a deletion in the proto-spacer (2 of 20 mutant phages).

In 14 cases where a nucleotide mutation occurred, the deduced amino acid was changed. These mutations had apparently no effect (besides enabling infection of the BIM) on the completion of the phage lytic cycle. In 6 other cases, the amino acid was not changed, but this silent mutation generated a change of codon. Again, these mutations did not prevent the

TABLE 6. ORFs deduced from the genome of *S. thermophilus* phage 858 and their predicted functions

ORF	Start	Stop	Size (aa)[a]	Mass (kDa)	pI	Putative ribosome-binding site and start codon (AAAGGAGGTGA)	Best matches[b] (% amino acid identity)	Size[c] (aa)	Putative function
1	319	525	68	8.1	10.5	Not identified **ATG**	gp68 phage Sfi11 67/68 (98)	68	
2	414	827	137	16.2	8.7	CAGAGAGGTTAgtaca **ATG**	ORF1 phage 2972 137/137 (100)	137	
3	1009	1461	150	16.8	6.7	AAAATATTTGCggatgaatac **TTG**	ORF2 phage 2972 145/150 (96)	150	Terminase small subunit
4	1448	2119	223	25.3	9.0	AAAGGAGCTGTaagcg **ATG**	ORF3 phage 2972 223/223 (100)	223	Terminase large subunit
5	2412	2999	195	22.7	4.8	AGAGCCCTTGCgataacaaata **ATG**	ORF4 phage 2972 194/195 (99)	195	Terminase large subunit
6	3008	4513	501	57.6	5.0	GTAGGAGGAATg **ATG**	ORF5 phage 2972 501/501 (100)	501	Portal protein
7	4510	5403	297	34.4	8.8	TTGAGAGGGAAtatga **ATG**	ORF6 phage 2972 294/297 (98)	297	Capsid protein
8	5591	6172	193	21.3	4.8	TAGATAGGAGAaata **ATG**	ORF7 phage 2972 193/193 (100)	193	Scaffold protein
9	6192	6551	119	12.8	8.8	AAAGGAAATTTtaa **ATG**	ORF8 phage 2972 115/119 (96)	119	Capsid protein
10	6570	7616	348	37.5	4.9	AGAGGAGGAACattaaaac **ATG**	ORF9 phage 2972 346/348 (99)	348	Capsid protein
11	7628	7789	53	6.0	9.3	TTAAGAGGTACtgat **ATG**	ORF10 phage 2972 52/53 (98)	53	
12	7801	8142	113	13.0	4.6	TAGTGAGGTATggc **ATG**	ORF11 phage 2972 113/113 (100)	113	
13	8139	8453	104	11.5	9.5	AAAGAGGGAGAggtgttatttct **ATG**	ORF12 phage 2972 104/104 (100)	104	
14	8455	8793	112	12.5	8.9	CAAGGTGGTGAaataac **ATG**	ORF13 phage 2972 112/112 (100)	112	
15	8795	9181	128	14.6	5.0	AATGGCTAAGTgggaataag **ATG**	ORF14 phage 2972 127/128 (99)	128	
16	9195	9704	169	18.5	4.8	TCAGGAGGAAAaa **ATG**	ORF15 phage 2972 168/169 (99)	168	Tail protein
17	9782	10135	117	13.2	4.7	ATAGGAGTTAAaaaca **ATG**	gp117 phage Sfi11 101/117 (86)	117	
18	10186	10503	105	12.6	9.9	CGAGGAATTAAtcactaatgct **ATG**	ORF17 phage 2972 104/105 (99)	105	
19	10493	15046	1517	153.4	9.6	AGAGGGGCTTGctag **ATG**	ORF18 phage 2972 1511/1517(99)	1517	Tail protein
20	15046	16581	511	57.7	5.3	ATGAGAGGTATtaaata **ATG**	ORF19 phage 2972 508/511 (99)	511	
21	16581	21434	1617	178.5	5.3	ATTTGAGGAGAgatatatata **ATG**	ORF20 phage 2972 1599/1617(98)	1605	Receptor-binding protein
22	21435	23456	673	74.2	5.7	GTAGGAGGTTTtaa **TTG**	ORF21 phage 2972 659/673 (97)	673	Structural protein
23	23473	23859	128	14.6	4.7	AAGAAAGGAAAaatat **ATG**	ORF22 phage 2972 126/128 (98)	128	
24	23885	24028	47	5.5	6.6	TTAGGAGGAAGaaca **ATG**	ORF23 phage 2972 47/47 (100)	47	
25	23991	24278	95	11.0	8.7	CAAGAGCTTGTaggcttgtctc **ATG**	*T. parva* 22/87 (25)	497	
26	24275	24409	44	5.0	10.7	AAGCGAGGTTGataa **ATG**	*S. pyogenes* MGAS5005 19/35 (54)	61	
27	24437	24691	84	9.3	9.5	GAAACTTGAGAgg **ATG**	ORF87 phage Sfi21 68/79 (86)	87	Holin
28	24693	25295	200	21.8	4.5	TAAGGAAGGAAaatagt **ATG**	ORF25 phage DT1 180/200 (90)	200	Endolysin
29	25341	25463	40	4.6	8.8	AAACAAAGCGGtgtc **ATG**	ORF27 phage 2972 40/40 (100)	40	
30	25502	26263	253	30.0	9.7	AAAGATGGTGTcataag **ATG**	Phage J1 248/253 (98)	253	Endonuclease
31	26325	26552	75	8.6	3.9	AATCCCGGTTAca **ATG**	Phage J1 70/75 (93)	186	Endolysin
32	26720	26851	43	5.3	8.9	AAACGAGGTGAaaaca **ATG**	ORF30 phage 2972 43/43 (100)	43	
33	26952	27161	69	7.9	7.9	AAAGGAGAACTta **ATG**	ORF31 phage 2972 68/69 (98)	69	Cro-like repressor
34	27178	27300	40	5.0	8.0	AAAGGTATTTAaa **ATG**	ORF32 phage 2972 40/40 (100)	40	
35	27557	28030	157	18.1	6.2	TAGGGAGGGTAggaattaaat **ATG**	ORF33 phage 2972 157/157 (100)	157	
36	28027	28728	233	26.2	6.6	AAAGGAGAAACcttaacataag **ATG**	ORF34 phage 2972 233/233 (100)	233	
37	28685	30097	470	54.2	8.6	AAAGGGGTGTAaggtag **ATG**	ORF35 phage 2972 309/473 (65)	445	Helicase
38	30104	30577	157	18.2	4.8	AGATTTGGAGAtaaaaaaac **ATG**	*S. suis* 89/1591 110/157 (70)	156	
39	30582	31397	271	30.7	7.8	Not identified **ATG**	*S. suis* 89/1591 180/263 (68)	268	
40	31366	32910	514	60.0	8.0	AAAGGAGTTAGatactaaac **ATG**	*S. suis* 89/1591 363/508 (71)	517	Primase
41	33167	33487	106	12.1	9.9	AAAGAAAGGCAactttcaa **GTG**	ORF39 phage 2972 102/106 (96)	106	
42	33698	33871	57	6.6	8.1	GAAAGAGATGAtagaact **ATG**	ORF17 phage O1205 44/57 (77)	57	
43	33868	34023	51	6.3	5.6	GTAGGAGATTAgtagagtt **ATG**	ORF41 phage 2972 51/51 (100)	51	
44	34024	34536	170	19.6	6.3	TGAGGTGGAATag **ATG**	ORF42 phage 2972 170/170 (100)	170	DNA-binding protein
45	34505	34831	108	12.1	9.2	ATAGGAAAGGAaagatggtaa **ATG**	ORF43 phage 2972 108/108 (100)	108	
46	34835	35542	235	27.7	9.1	TGAGGAGTTATtaagac **ATG**	ORF44 phage 2972 234/235 (99)	235	

[a] That is, the number of amino acids (aa) of the predicted protein.
[b] If the host species is not mentioned, the phage infects a *S. thermophilus* strain.
[c] That is, the size in amino acids (aa) of the best-matched protein.

phage to complete the infection process. In mutant phages 858.S32C and 2972.S6D, 75-nucleotide and 1-nucleotide deletions, respectively, occurred. The 75-nucleotide deletion in phage 858.S32C targeted the end of ORF42 and the beginning of ORF43 of phage 858 (Table 6). No putative function could be assigned to either ORF42 (57 amino acids) or ORF43 (51 amino acids) (Table 6). Interestingly, the deletion led to the formation of an ORF42-ORF43 fusion product, but no function could be assigned to the deduced fusion protein (83 amino acids). Phage mutant 2972.S6D had a one-base deletion, which led to a frameshift and, consequently, the presence of several stop codons in the ORF44 sequence for which no putative function could been assigned (Table 6).

Taken altogether, these data confirm that a newly added spacer must be identical to the proto-spacer to be fully effective and that the CRISPR1-specific sequence (NNAGAAW) is also important for the phage resistance phenotype.

DISCUSSION

The remarkable diversity and metabolic capabilities of bacteria allow them to grow and prosper in every ecosystem where life forms have been found (18). Similarly, bacteriophages are present in these same ecosystems, including manufactured ecological niches such as food fermentation vats. It is now believed that phages represent the most abundant biological entities on the planet (7). Thus, it is not surprising to observe that bacteria have devised a number of strategies to defend against these prolific invaders. CRISPRs and their associated *cas* genes constitute the latest defense mechanism unveiled in prokaryotes (2).

In the present study, we show that CRISPR-mediated phage resistance is indeed a novel antiphage system since its general mode of action is distinct from the previously known systems. Our results also demonstrate that CRISPR-mediated phage resistance protects *S. thermophilus* against the two main groups of phage known to infect this bacterial species. Thus, this antiphage system is exceptionally broad and effective. This wide-ranging efficacy against phages is in agreement with the fact that CRISPRs have been found in a wide range of bacterial genomes (19, 22, 30).

The isolation and characterization of BIMs obtained through iterative phage challenges have revealed that one spacer will typically be added to the CRISPR1 locus. However, multiple spacers can also be acquired by CRISPR1, providing enhanced resistance to phages. The iterative addition of spacers is particularly interesting and separates the CRISPR-mediated phage resistance from other natural antiphage defense systems. With the other four systems (adsorption inhibition, DNA ejection inhibition, restriction-modification systems, and Abi), it is not possible to generate new phage-resistant derivatives (when phage mutants have emerged) without any fitness cost to the host at each generation. In contrast, CRISPR-mediated phage resistance allows the acquisition of a new spacer specific to the phage mutants without an obvious fitness cost associated with it. Thus, it is possible to create multiresistant *S. thermophilus* strains by successive challenges using different phages.

Because new spacers were almost always inserted at the leader end of the CRISPR1 locus, it is tempting to hypothesize that spacer position could serve as a memory of previous phage encounters by a strain. Although this may be true for many BIMs, spacer deletion did occur in some of the BIMs. Thus, this presumed historical perspective, albeit interesting, may be of limited value in some cases. Similarly, the reason for the prevalence of new spacer acquisitions at the leader end of the repeat arrays is unknown, although a putative role of the leader could explain this phenomenon. Nonetheless, it is worth mentioning that when spacer deletion was observed in a BIM, a new spacer always occurred in the vicinity of the deleted region. It is possible that the spacer deletion occurs by homologous recombination between CRISPR direct repeats.

A key feature in CRISPR-mediated phage resistance is that the newly acquired spacer (between 29 and 31 nucleotides in size) must be identical to the phage genomic sequence to provide resistance. Only 2 of the newly acquired spacers (of 39) described here were not identical (one mismatch) to a known phage sequence. However, these two spacers were found in BIMs that had acquired more than one spacer, and the other associated spacers had a perfect match to a phage genomic sequence. Moreover, most of the analyzed phage mutants that were able to infect these BIMs had a mutation in the proto-spacer. Interestingly, 7 of the 20 mutated phages analyzed had no mutation in the proto-spacer but had a mutation in the AGAA flanking sequence. This sequence appears to play a critical role in CRISPR1-mediated phage resistance in *S. thermophilus* because a mutation within this sequence allows the phage to escape the CRISPR1-mediated resistance. This strongly suggests that CRISPR1 and/or the *cas*-associated proteins may be involved in a nucleotide recognition mechanism. The importance of the NNAGAAW motif was recently confirmed by their presence in proto-spacer region corresponding to the CRISPR1 spacers from several different *S. thermophilus* strains (22).

All of the phage genomic sequences matching the acquired spacers in the *S. thermophilus* BIMs were found in a coding region, and the coding strand was three times more frequently targeted. We believe that the frequency of the NNAGAAW motif in the phage genome (2.9 on the coding strand for 1 on the noncoding strand) is responsible for this bias. In addition, early transcribed modules of the phage genome appear to be more frequently targeted for the acquisition of new CRISPR1 spacers. It has been previously hypothesized that CRISPR may play a role in an RNA interference system (30). Thus, it is tempting to speculate that the early transcribed phage mRNAs would be a preferential target for a mechanism such as RNA interference. Rapidly silencing the phage infection may allow the cell to recover thereby, increasing cell survival.

The acquisition of a spacer from a coding sequence also suggests that the targeted gene is important for phage development. However, this observation is debatable as most of the phage genome (89.6% for DT1 and 93.8% for 2972) is coding. Furthermore, phage mutants were still able to propagate efficiently despite the apparent gene inactivation. As indicated above, it is possible that the CRISPR-mediated resistance somehow targets the mRNA. Knowing that many *S. thermophilus* phage genes are transcribed as part of a polycistronic mRNA (26, 42), inactivating larger transcripts may prevent the translation of essential phage proteins.

In conclusion, the CRISPR/*cas* system clearly represents a

novel and interesting avenue for the development of phage-resistant bacterial strains for fermentation and biotechnological processes. Moreover, because of the widespread distribution of phages in various ecosystems, CRISPRs likely play a significant role in prokaryotic evolution and ecology (2). The identification of a nucleotide motif in the phage genome that is important for the phage resistance phenotype is another clue toward the elucidation of the molecular mode of action of the CRISPR1 mechanism.

ACKNOWLEDGMENTS

We thank Denise Tremblay, Simon Labrie, and Julie Samson for stimulating discussions; Yanick Bourbeau for the DNA display conception; and Gene Bourgeau for editorial assistance.

S.M. acknowledges the support of the Natural Sciences and Engineering Research Council of Canada through its Discovery program.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schäffer, 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.
2. **Barrangou, R., C. Fremaux, P. Boyaval, M. Richards, H. Deveau, S. Moineau, D. A. Romero, and P. Horvath.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709–1712.
3. **Behnke, D., and H. Malke.** 1978. Bacteriophage interference in *Streptococcus pyogenes*. I. Characterization of prophage-host systems interfering with the virulent phage A25. Virology **85:**118–128.
4. **Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyrpides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols.** 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat. Biotechnol. **22:**1554–1558.
5. **Bolotin, A., B. Quinquis, A. Sorokin, and S. D. Ehrlich.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151:**2551–2561.
6. **Bouchard, J. D., and S. Moineau.** 2004. Lactococcal phage genes involved in AbiK sensitivity and their relation to single-strand annealing proteins. J. Bacteriol. **186:**3649–3652.
7. **Breitbart, M., and F. Rohwer.** 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. **13:**278–284.
8. **Brüssow, H.** 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. **55:**283–303.
9. **Brüssow, H., and F. Desière.** 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. **39:**213–222.
10. **Chopin, M.-C., A. Chopin, and E. Bidnenko.** 2005. Phage abortive infection in lactococci: variations on a theme. Curr. Opin. Microbiol. **8:**473–479.
11. **Delcour, J., T. Ferain, and P. Hols.** 2000. Advances in the genetics of thermophilic lactic acid bacteria. Curr. Opin. Biotechnol. **11:**497–504.
12. **Desiere, F., S. Lucchini, C. Canchaya, M. Ventura, and H. Brüssow.** 2002. Comparative genomics of phages and prophages in lactic acid bacteria. Antonie van Leeuwenhoek **82:**73–91.
13. **Deveau, H., S. J. Labrie, M. C. Chopin, and S. Moineau.** 2006. Biodiversity and classification of lactococcal phages. Appl. Environ. Microbiol. **72:**4338–4346.
14. **Deveau, H., M. R. Van Calsteren, and S. Moineau.** 2002. Effect of exopolysaccharides on phage-host interactions in *Lactococcus lactis*. Appl. Environ. Microbiol. **68:**4364–4369.
15. **Duplessis, M., C. M. Lévesque, and S. Moineau.** 2006. Characterization of *Streptococcus thermophilus* host-range phage mutants. Appl. Environ. Microbiol. **72:**3036–3041.
16. **Duplessis, M., and S. Moineau.** 2001. Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. Mol. Microbiol. **41:**325–336.
17. **Duplessis, M., W. M. Russell, D. A. Romero, and S. Moineau.** 2005. Global gene expression analysis of two *Streptococcus thermophilus* bacteriophages using DNA microarray. Virology **340:**192–208.
18. **Emond, E., and S. Moineau.** 2007. Bacteriophages and food fermentations, p. 93–124. *In* S. McGrath and D. van Sinderen (ed.), Bacteriophage: genetics and molecular biology. Horizon Scientific Press/Caister Academic Press, New York, NY.
19. **Godde, J. S., and A. Bickerton.** 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol. **62:**718–729.
20. **Haft, D. H., J. Selengut, E. F. Mongodin, and K. E. Nelson.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. **1:**474–483.
21. **Hill, C., J. Massey, and T. R. Klaenhammer.** 1991. Rapid method to characterize lactococcal bacteriophage genomes. Appl. Environ. Microbiol. **57:**283–288.
22. **Horvath, P., D. A. Romero, A.-C. Coûté-Monvoisin, M. Richards, H. Deveau, S. Moineau, P. Boyaval, C. Fremaux, and R. Barrangou.** 2007. Diversity, activity and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. **190:**1401–1412.
23. **Jansen, R., J. D. Embden, W. Gaastra, and L. M. Schouls.** 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. **43:**1565–1575.
24. **Josephsen, J., and H. Neve.** 2004. Bacteriophage and antiphage mechanisms of lactic acid bacteria, p. 295–350. *In* S. Salminen, A. Von Wright, and A. Ouwehand (ed.), Lactic acid bacteria: microbiological and functional aspects, 3rd ed. Marcel Dekker, Inc., New York, NY.
25. **Labrie, S. J., and S. Moineau.** 2007. Abortive infection mechanisms and prophage sequences significantly influence the genetic make-up of emerging lytic lactococcal phages. J. Bacteriol. **189:**1482–1487.
26. **Lamothe, G., C. Lévesque, F. Bissonnette, A. Cochu, C. Vadeboncoeur, M. Frenette, M. Duplessis, D. Tremblay, and S. Moineau.** 2005. Characterization of the *cro-ori* region of the *Streptococcus thermophilus* virulent bacteriophage DT1. Appl. Environ. Microbiol. **71:**1237–1246.
27. **Le Marrec, C., D. van Sinderen, L. Walsh, E. Stanley, E. Vlegels, S. Moineau, P. Heinze, G. Fitzgerald, and B. Fayard.** 1997. Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. Appl. Environ. Microbiol. **63:**3246–3253.
28. **Lévesque, C., M. Duplessis, J. Labonté, S. Labrie, C. Fremaux, D. Tremblay, and S. Moineau.** 2005. Genomic organization and molecular analysis of the virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. Appl. Environ. Microbiol. **71:**4057–4068.
29. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115.
30. **Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin.** 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct. **1:**1–26.
31. **Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills.** 2006. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. USA **103:**15611–15616.
32. **Moineau, S., E. Durmaz, S. Pandian, and T. R. Klaenhammer.** 1993. Differentiation of two abortive mechanisms by using monoclonal antibodies directed toward lactococcal bacteriophage capsid proteins. Appl. Environ. Microbiol. **59:**208–212.
33. **Moineau, S., J. Fortier, H.-W. Ackermann, and S. Pandian.** 1992. Characterization of lactococcal bacteriophages from Quebec cheese plants. Can. J. Microbiol. **38:**875–882.
34. **Moineau, S., S. Pandian, and T. R. Klaenhammer.** 1994. Evolution of a lytic bacteriophage via DNA acquisition from the *Lactococcus lactis* chromosome. Appl. Environ. Microbiol. **60:**1832–1841.
35. **Moineau, S., D. Tremblay, and S. Labrie.** 2002. Phages of lactic acid bacteria: from genomics to industrial applications. ASM News **68:**388–393.
36. **Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez, and E. Soria.** 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. **60:**174–182.
37. **Pourcel, C., G. Salvignol, and G. Vergnaud.** 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology **151:**653–663.
38. **Staden, R.** 1996. The Staden sequence analysis package. Mol. Biotechnol. **5:**233–241.
39. **Sturino, J. M., and T. R. Klaenhammer.** 2006. Engineered bacteriophage-defense systems in bioprocessing. Nat. Rev. Microbiol. **4:**395–404.
40. **Tremblay, D. M., and S. Moineau.** 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. Virology **255:**63–76.
41. **Tremblay, D., M. Tegoni, S. Spinelli, V. Campanacci, S. Blangy, C. Huyghe, A. Desmyter, S. Labrie, S. Moineau, and C. Cambillau.** 2006. Receptor-binding protein of *Lactococcus lactis* phages: identification and characterization of the saccharide receptor-binding site. J. Bacteriol. **188:**2400–2410.
42. **Ventura, M., S. Foley, A. Bruttin, S. C. Chennoufi, C. Canchaya, and H. Brüssow.** 2002. Transcription mapping as a tool in phage genomics: the case of the temperate *Streptococcus thermophilus* phage Sfi21. Virology **296:**62–76.