APPLIED AND ENVIRONMENTAL MICROBIOLOGY, July 2005, p. 4057–4068 Vol. 71, No. 7
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.7.4057–4068.2005

Genomic Organization and Molecular Analysis of Virulent Bacteriophage 2972 Infecting an Exopolysaccharide-Producing *Streptococcus thermophilus* Strain

Céline Lévesque,[1]† Martin Duplessis,[1,2] Jessica Labonté,[1,2] Steve Labrie,[1,2] Christophe Fremaux,[4] Denise Tremblay,[1,3] and Sylvain Moineau[1,2,3]*

Groupe de Recherche en Écologie Buccale (GREB), Faculté de Médecine Dentaire,[1] Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie,[2] and Centre de Référence pour Virus Bactériens Félix d'Hérelle,[3] Université Laval, Québec, Canada G1K 7P4, and Danisco, BP10, 86220 Dangé Saint Romain, France[4]

Received 4 November 2004/Accepted 1 February 2005

The *Streptococcus thermophilus* virulent *pac*-type phage 2972 was isolated from a yogurt made in France in 1999. It is a representative of several phages that have emerged with the industrial use of the exopolysaccharide-producing *S. thermophilus* strain RD534. The genome of phage 2972 has 34,704 bp with an overall G+C content of 40.15%, making it the shortest *S. thermophilus* phage genome analyzed so far. Forty-four open reading frames (ORFs) encoding putative proteins of 40 or more amino acids were identified, and bioinformatic analyses led to the assignment of putative functions to 23 ORFs. Comparative genomic analysis of phage 2972 with the six other sequenced *S. thermophilus* phage genomes confirmed that the replication module is conserved and that *cos*- and *pac*-type phages have distinct structural and packaging genes. Two group I introns were identified in the genome of 2972. They interrupted the genes coding for the putative endolysin and the terminase large subunit. Phage mRNA splicing was demonstrated for both introns, and the secondary structures were predicted. Eight structural proteins were also identified by N-terminal sequencing and/or matrix-assisted laser desorption ionization—time-of-flight mass spectrometry. Detailed analysis of the putative minor tail proteins ORF19 and ORF21 as well as the putative receptor-binding protein ORF20 showed the following interesting features: (i) ORF19 is a hybrid protein, because it displays significant identity with both *pac*- and *cos*-type phages; (ii) ORF20 is unique; and (iii) a protein similar to ORF21 of 2972 was also found in the structure of the *cos*-type phage DT1, indicating that this structural protein is present in both *S. thermophilus* phage groups. The implications of these findings for phage classification are discussed.

Streptococcus thermophilus is one of the most economically important lactic acid bacteria (LAB) used for the manufacture of yogurt and Swiss- or Italian-type hard cooked cheeses (19). This bacterium may also play a role as a probiotic, alleviating symptoms of lactose intolerance and other gastrointestinal disorders (28). Research on the physiology of *S. thermophilus* has generated significant insights into some of its properties, including sugar metabolism, protein utilization, and exopolysaccharide (EPS) production (7, 19). *S. thermophilus* belongs to the group of bacteria that are generally recognized as safe, which is an exception in the genus *Streptococcus*.

S. thermophilus bacteriophages have been a subject of ongoing interest, because they are ubiquitous in dairy environments and because their rapid lytic cycle can lead to significant bacterial lysis that results in milk fermentation delays (52). Many strategies have been employed by dairy factories to curtail phage infections. One extensively used tactic is the rotation of several LAB strains to prevent the proliferation of specific phage populations. Additionally, carefully selected so-called phage-insensitive *S. thermophilus* strains are introduced into dairy processes with the hope of limiting phage infections (49). However, despite these efforts, new *S. thermophilus* phages are still emerging. It is expected that the characterization of an increasing number of streptococcal phage genomes should lead to a better understanding of phage evolution, which is required for the development of long-term phage-resistant LAB strains.

S. thermophilus phages are a relatively homogenous group with the same morphology (B1 morphotype, *Siphoviridae* family) (1). They have an isometric capsid (diameter, 45 to 60 nm) and a long, noncontractile tail of various lengths (240 to 270 nm) and thicknesses (9 to 13 nm) (8). They are divided into two groups based on the packaging mechanism of their double-stranded DNA (*cos* and *pac* types) and the number of major structural proteins (37). Six complete genome sequences of *S. thermophilus* phages are currently available. They include the *cos*-type phages DT1 (66), Sfi19 (42), Sfi21 (12), and 7201 (62), as well as the *pac*-type phages O1205 (61) and Sfi11 (39, 40). DT1, Sfi19, and Sfi11 are virulent phages, while the others are temperate.

Comparative genomic analyses of these six genomes demonstrated that *S. thermophilus* phages share extensive DNA sequence similarity in the replication module and lysis cassette. Significant differences have been reported in the genes coding for structural proteins, which is in agreement with the classification scheme (20, 37). An interesting feature is the close genetic relationship between virulent and temperate *S. thermophilus* phages. It has even been proposed that virulent *S.*

* Corresponding author. Mailing address: GREB, Faculté de Médecine Dentaire, Université Laval, Québec, Canada G1K 7P4. Phone: (418) 656-3712. Fax: (418) 656-2861. E-mail: Sylvain.Moineau@bcm.ulaval.ca.

† Present address: Oral Microbiology, Faculty of Dentistry, University of Toronto, 124 Edward St., Toronto, Ontario, Canada M5G 1G6.

thermophilus phages arose from temperate phages through a combination of rearrangement and deletion events within the lysogeny module (11, 41).

One of the most significant contributions of the streptococcal phage genomic analyses has been in the field of phage taxonomy. These comparative analyses revealed the presence of related phages in other species and genera of low-G+C-content gram-positive bacteria (9). Another benefit of these genomic studies has been the use of some phage genetic elements to construct antiphage systems. These elements include the phage origin of replication (26, 62, 63), the CI-like repressor (14), the immunity gene (13), and the antisense RNA technology targeting the putative helicase and primase genes of *S. thermophilus* phages (63, 64).

In the present work, we report the complete nucleotide sequence and molecular characterization of 2972, a virulent *pac*-type phage that infects the exopolysaccharide-producing strain *S. thermophilus* RD534, which is used for the production of yogurt worldwide.

MATERIALS AND METHODS

Phage preparation and purification. The virulent *S. thermophilus* phages infecting strain *S. thermophilus* RD534 were provided by Danisco (France). For phage propagation, *S. thermophilus* RD534 was grown at 42°C without agitation in M17 broth (Québec, Québec, Canada) supplemented with 0.5% (wt/vol) lactose and 10 mM $CaCl_2$. When the optical density at 600 nm reached 0.2, approximately 10^7 PFU/ml of phage was added and the culture was incubated overnight at 42°C. The lysate was clarified by centrifugation and passed through a 0.45-μm-pore-size filter. Phages were purified by ultracentrifugation using a discontinuous CsCl gradient (56). Phage morphology was observed as described previously (50) with a Philips 420 transmission electron microscope operating at 80 kV.

Purification of phage DNA and DNA sequencing. Phage DNA was isolated using the QIAGEN lambda Maxi kit as described previously (31). DNA restriction profiles were analyzed using Molecular Analyst Fingerprinting Plus software (Bio-Rad Laboratories) and compared using the UPGMA (unweighted-pair group method using average linkages) clustering method. Phage 2972 DNA was sequenced from shotgun subclone libraries of the genome (Integrated Genomics, Inc., Chicago, IL). Then the gap between contigs was closed by sequencing gap-specific PCR products generated by using phage 2972 genomic DNA as a template; this procedure was performed by the DNA sequencing service of Université Laval. Computer-assisted DNA and protein analyses were performed using the Genetics Computer Group Sequence Analysis software package, version 10.3 (22). The genome sequence was analyzed using the open reading frame (ORF) finder graphical analysis tool (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to define potential coding regions. The PROSITE and Pfam databases were employed (http://hits.isb-sib.ch/cgi-bin/PFSCAN) to locate putative functional motifs. PSI-BLAST and Advanced BLAST Search 2.1 (http://www.ncbi.nlm.nih.gov/BLAST) were also used for sequence comparisons with databases (2).

RNA methods. Total RNA was extracted from phage 2972-infected *S. thermophilus* cells (17 min after infection) to study mRNA splicing. Transcription was stopped by adding rifampin at 150 μg per ml. Cells were collected by centrifugation and frozen in dry-ice–ethanol. The frozen cell pellets were resuspended in 1 ml of TRIZOL reagent (Invitrogen) and transferred to a 2-ml tube containing 0.7 g of glass beads (106 μm; Sigma-Aldrich). The mixture was vortexed with a Mini-Beadbeater-8 cell (BioSpec Products) three times, for 1 min each time (67). Between treatments, the cell suspensions were chilled on ice for 1 min. The supernatant was extracted twice with TRIZOL-chloroform. Nucleic acids were precipitated with isopropanol and resuspended in diethyl pyrocarbonate (DEPC)-treated water. Samples were treated with DNase I (10 U) for 30 min at 37°C with 80 U of RNaseOUT recombinant RNase inhibitor (Invitrogen).

DNA-free RNA samples were subjected to reverse transcription (RT) as follows. Ten micrograms of purified RNA and 6 μg of oligonucleotides (random hexamers; Invitrogen) were added to DEPC-treated water to obtain a final volume of 18.5 μl. The mixture was heated at 70°C for 10 min and snap-frozen in dry-ice–ethanol for 30 s. Then 400 U of SuperScript II RNase H^- reverse transcriptase (Invitrogen), 6 μl of 5× First Strand buffer (Invitrogen), 3 μl of 0.1 M dithiothreitol (DTT) (Invitrogen), and 0.75 mM dATP, dCTP, dGTP, and dTTP were added to the mixture, and the RT reaction was performed at 42°C for 16 h. The reaction was terminated by heating at 75°C for 15 min. The cDNA was amplified by PCR as described previously (23). The cDNA was heated at 94°C for 4 min, followed by 35 cycles of the following temperature-time profile: 94°C for 45 s, 57°C for 45 s, and 73°C for 1 min. After the final cycle, the mixtures were kept at 73°C for an extra 10 min. The primers used for amplification of the gene encoding the large subunit of the terminase were 5′-CTATCAAAGCAGCTAC GCCC-3′ (forward) and 5′-CCTTCACCGACTACCACGATA-3′ (reverse), and the primers used for the endolysin-encoding gene were 5′-GAAGTCAAATAT GTTAACGG-3′ (forward) and 5′-CTTCAGACTTGCCATCTGGA-3′ (reverse). PCR products were separated by electrophoresis on agarose gels (2%), stained with ethidium bromide, and visualized by UV. PCR products were also purified using QIAquick PCR purification columns (QIAGEN) and sequenced with the same primers used for the PCR amplification.

Phage structural protein analysis. Phage 2972 structural proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with a 15% polyacrylamide separating gel and a 4.5% polyacrylamide stacking gel (56). The proteins were then transferred to a polyvinylidene difluoride Immobilon-P^{SQ} membrane (Millipore). After staining with 0.1% (wt/vol) Coomassie blue in 40% (vol/vol) methanol and 1% (vol/vol) acetic acid, the protein bands of interest were excised. N-terminal sequencing was performed by Edman degradation using an Applied Biosystems model 473A pulsed liquid protein sequencer. For matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectrometry, Coomassie-stained proteins were extracted from the polyacrylamide gels and digested with trypsin using a MassPrep liquid handling station (Micromass Ltd.) according to the manufacturer's specifications. The resulting peptides were lyophilized and resuspended in 0.1% (vol/vol) trifluoroacetic acid (TFA). α-Cyano-4-hydroxycinnamic acid (1.7 mg/ml in 58% acetonitrile–0.1% TFA) was used as a matrix for the MALDI analysis. Equal volumes of peptide and matrix solution were mixed and spotted onto a stainless-steel MALDI sample plate. The sample-matrix solution was allowed to air dry at room temperature and was then washed with 0.1% TFA. MALDI-TOF spectra were acquired on a Voyager-DE PRO Biospectrometry Workstation (Applied Biosystems) and analyzed using DataExplorer software, version 4.0 (Applied Biosystems). The instrument was operated in the positive-ion reflector delayed-extraction mode. The PeptIdent tool (http://ca.expasy.org/tools/peptident.html) was used to search nonredundant Swiss-Prot/TrEMBL protein databases for matching peptide mass fingerprints and to identify proteins. Search criteria allowed a maximum of one missed cleavage by trypsin, complete carboxyamidomethylation of cysteine, partial methionine oxidation, and mass deviations under 60 ppm. The N-terminal sequencing and mass spectrometry analyses were both performed at the Eastern Québec Proteomic Center (Québec, Canada).

Nucleotide sequence accession number. The nucleotide sequence of the phage 2972 genome has been deposited in GenBank under accession no. AY699705.

RESULTS

Phages infecting *S. thermophilus* RD534. *S. thermophilus* RD534 is widely used for the commercial manufacture of yogurt because it produces an EPS that gives a viscous texture to the fermented dairy product. Compositional and structural analyses have revealed that the EPS produced by RD534 is composed primarily of D-glucose and D-galactose in a molar ratio of 1:1 (data not shown). The EPS of RD534 is thus similar to those produced by other *S. thermophilus* strains (38, 44). *S. thermophilus* RD534 is sensitive to a group of closely related virulent *S. thermophilus* phages (Fig. 1), all of which belong to the *S. thermophilus pac*-type group, as submolar fragments were still observed following the heating of restriction endonuclease digests of the phage DNA (data not shown). Phage 2972 was randomly selected as a representative of this group for further analysis. Electron microscopic analysis of the purified preparation of phage 2972 showed that it possesses a 55-nm-diameter isometric capsid and a 260-nm-long noncontractile tail (Fig. 2).

Determination of the complete nucleotide sequence. The genome of phage 2972 has 34,704 bp. It is the shortest *S. thermophilus* phage genome for which the complete nucleotide sequence is available. The genome of phage 2972 is 5.1 kb shorter than the genome of phage Sfi11 (39,807 bp), the other

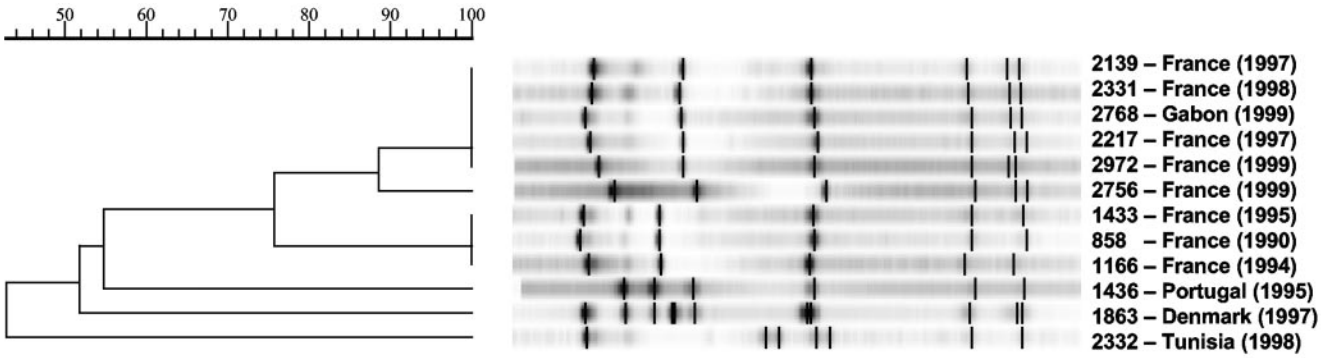


FIG. 1. Dendrogram of EcoRV restriction profiles of 12 virulent phages infecting *S. thermophilus* RD534. The phage names and the years and countries of isolation are given. Each phage was isolated from a different dairy factory.

virulent *pac*-type *S. thermophilus* phage for which the entire nucleotide sequence is available (39, 40). Phage 2972 DNA has an average G+C content of 40.15%, which is similar to the G+C content reported for the host genome (37.8 to 40%) (43). Forty-four ORFs of 40 codons or more were identified (Table 1). They were all located on the same strand and all started with either an ATG or a TTG initiation codon. Every ORF, except *orf4* and *orf37*, was preceded by a region that shares variable homologies with the Shine-Dalgarno sequence complementary to the 3′ end of the 16S rRNA of *S. thermophilus* (AAAGGAGGTGA). Of the 44 ORFs, 23 could be assigned a putative function based on their similarity to proteins with known functions or on conserved motifs (Table 1). Three putative promoters (P1, P2, and P3) were identified by their similarity to the consensus −35 and −10 sequences. The P1 promoter (gTGAtA–N_{16}–TATAAT; lowercase letters indicate a difference from the consensus sequence) was located 358 bp upstream of the ATG start codon of *orf1*, the P2 promoter (TTGACA–N_{17}–TAaAAT) was located 30 bp upstream of the *orf31* start codon, and the P3 promoter (TTGACA–N_{20}–TAc ttT) was located 166 bp upstream of the *orf39* start codon. Three putative terminator-like structures (factor-independent terminators) were also identified. Terminator T1 was located 18 bp downstream of the stop codon of *orf1*, T2 was located 64 bp upstream of the putative P2 promoter, and T3 was located 165 bp downstream of the *orf38* stop codon. A 250-bp noncoding region was identified between *orf38* and *orf39*. It contains four inverted repeats (73% A+T content within 140 bp) that may correspond to the origin of replication (*ori*) of the phage genome (6, 26). In all the *S. thermophilus* phage genomes analyzed to date, the phage *ori* is located just upstream of the genes coding for proteins involved in DNA replication.

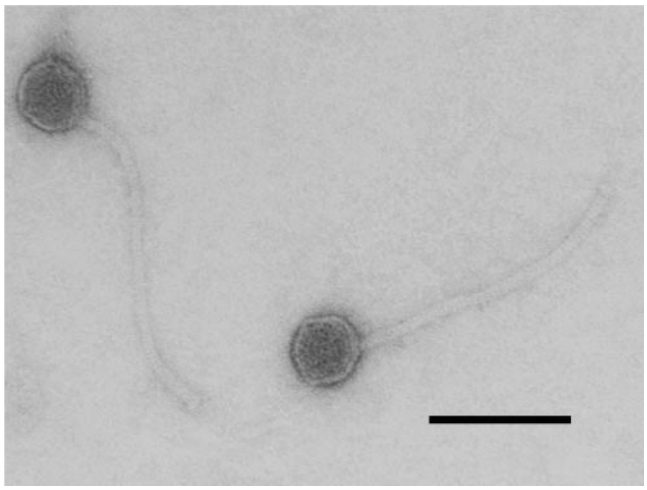


FIG. 2. Electron micrograph of *S. thermophilus* phage 2972 virions negatively stained with 2% phosphotungstic acid (pH 7.2). Bar, 100 nm.

Analysis and organization of the genome. The genome of phage 2972 is organized into distinct modular regions commonly found in other phages of the *Siphoviridae* family (9, 20, 32). As can be seen in Fig. 3, two regions are highly conserved in the seven *S. thermophilus* phage genomes, including the segments containing the genes necessary for DNA replication and host cell lysis. One notable exception is the putative holin-encoding gene (*orf25*) in the lysis cassette of phage 2972. *orf25* codes for a 108-amino-acid (108-aa) protein with significant similarities (>50% identity) with the holins of a *Streptococcus pyogenes* 315.5 prophage (5), the *Streptococcus mitis* temperate phage SM1 (59), and the *Streptococcus pneumoniae* temperate phage MM1 (53) (Table 1). Topology prediction analyses identified three transmembrane domains in ORF25, a characteristic of type I holins (68). In most phages of the *Siphoviridae* family, the gene immediately downstream from the holin-encoding gene codes for the endolysin. In phage 2972, the products of *orf26* and *orf29* are both homologous (>76% identity) to the bacteriophage peptidoglycan hydrolases (amidases) of *S. thermophilus* phages DT1, 7201, Sfi19, Sfi21, Sfi11, and O1205 (Table 1). These two ORFs are separated by an intron. Another intron has been found in the phage 2972 genome. It is located between *orf3* and *orf4*. The *orf3* and *orf4* genes, encoding proteins of 223 aa and 195 aa, respectively, exhibit high identity with the large subunit of the terminase of *S. thermophilus* phage Sfi11 (Table 1). The introns of phage 2972 are described below.

Lysogeny module. In the genome of *S. thermophilus* temperate phages, the conserved region containing the DNA replication module is separated from the lysis cassette by the lysogeny module (Fig. 3). As in the virulent *S. thermophilus* phages DT1, Sfi11, and Sfi19, remnants of a lysogeny module are found in the genome of the virulent phage 2972. For example, *orf31* likely codes for a *cro*-like repressor. It is possible that phage 2972 picked up DNA by homologous recombination with a

TABLE 1. Features of phage 2972 ORFs and the putative functions of their products

ORF	Start	Stop	Size (aa)	Mol mass (kDa)	pI	Putative function and motif[a]	Best match(es) (% amino acid identity)
1	414	827	137	16.1	8.7	—	ORF137 *S. thermophilus* phage Sfi11 (135/137; 98%)
2	1009	1461	150	16.7	8.0	Small terminase	ORF25 *S. thermophilus* phage O1205 (74/147; 50%)
3	1448	2119	223	25.3	9.0	Large terminase	ORF411 phage Sfi11 (211/217; 97%)
4	2404	2991	195	22.6	4.8	Large terminase	ORF411 phage Sfi11 (183/192; 95%)
5	3000	4505	501	57.4	5.0	Portal protein	ORF27 phage O1205 (472/501; 94%)
6	4502	5395	297	34.3	8.8	Capsid protein	ORF28 phage O1205 (286/297; 96%)
7	5583	6164	193	21.2	4.8	Scaffold protein	ORF29 phage O1205 (184/193; 95%)
8	6184	6543	119	12.7	7.9	Capsid protein	ORF119 phage Sfi11 (110/119; 92%)
9	6562	7608	348	37.4	4.9	Capsid protein	ORF348 phage Sfi11 (326/348; 93%)
10	7620	7781	53	5.9	9.3	—	ORF32 phage O1205 (46/53; 86%)
11	7793	8134	113	13.0	4.6	—	ORF33 phage O1205 (108/112; 96%)
12	8131	8445	104	11.4	9.5	—	ORF34 phage O1205 (88/104; 84%)
13	8447	8785	112	12.4	8.9	—	ORF114 phage Sfi11 (94/112; 83%)
14	8787	9173	128	14.6	5.0	—	ORF128 phage Sfi11 (116/128; 90%)
15	9187	9696	169	18.5	4.9	Tail protein	ORF37 phage O1205 (157/169; 92%)
16	9772	10125	117	13.1	5.0	—	ORF117 phage Sfi11 (114/117; 97%)
17	10176	10493	105	12.6	9.9	—	ORF105 phage Sfi11 (103/105; 98%)
18	10483	15036	1517	153.5	9.5	TMP	ORF1510 phage Sfi11 (1191/1523; 78%)
19	15036	16571	511	57.7	5.2	Tail protein	ORF17 *S. thermophilus* phage DT1 (282/449; 62%), ORF512 phage Sfi11 (298/518; 57%), ORF41 phage O1205 (297/518; 57%), ORF515 *S. thermophilus* phage Sfi21 (264/448; 58%), ORF515 *S. thermophilus* phage Sfi19 (261/448; 58%), ORF34 *S. thermophilus* phage 7201 (186/455; 40%)
20	16571	21388	1605	177.3	5.3	Receptor-binding protein	ORF18 *S. thermophilus* phage MD2 (540/800; 67%)
21	21389	23410	673	74.2	6.1	Tail protein	ORF46 phage O1205 (462/674; 68%), ORF669 phage Sfi11 (463/674; 68%), ORF670 phage Sfi21 (392/676; 57%), ORF39 phage 7201 (391/676; 57%), ORF670 phage Sfi19 (391/676; 57%), ORF19 phage DT1 (345/663; 52%)
22	23427	23813	128	14.5	4.6	—	ORF149 phage Sfi11 (83/123; 67%), ORF21 phage DT1 (81/114; 71%), ORF131 phage Sfi19 (80/114; 70%), ORF47 phage O1205 (78/112; 69%), ORF117 phage Sfi21 (74/112; 66%), ORF40 phage 7201 (75/126; 59%)
23	23839	23982	47	5.4	6.6	—	*S. pyogenes* prophage 315.5 (24/40; 60%)
24	24009	24341	110	12.3	5.0	—	No hit
25	24366	24692	108	12.0	5.5	Holin	*S. pyogenes* prophage 315.5 (63/104; 60%), *S. mitis* phage SM1 (59/107; 55%), *S. pneumoniae* phage MM1 (58/106; 54%)
26	24689	25288	199	21.7	4.8	Endolysin	ORF25 phage DT1 (162/194; 83%), ORF44 phage 7201 (160/192; 83%), ORF288 phage Sfi19 (158/193; 81%), ORF288 phage Sfi21 (158/193; 81%), ORF288 phage Sfi11 (157/193; 81%), ORF51 phage O1205 (151/193; 78%),
27	25334	25456	40	4.5	8.8	—	No hit
28	25495	25680	61	6.8	9.4	Endonuclease	*S. thermophilus* phage S3b (54/54; 100%), *S. thermophilus* phage ST3 (54/54; 100%), ORF26 phage DT1 (60/61; 98%)
29	25747	25974	75	8.6	4.1	Endolysin	ORF44 phage 7201 (65/75; 86%), ORF51 phage O1205 (62/75; 82%), ORF288 phage Sfi11 (59/75; 78%), ORF288 phage Sfi21 (59/75; 78%), ORF288 phage Sfi19 (57/75; 76%)
30	26142	26273	43	5.2	8.9	—	No hit
31	26374	26583	69	7.8	7.9	*cro*-like repressor	ORF69 phage Sfi19 (67/69; 97%)
32	26600	26722	40	5.0	8.0	—	ORF40 phage Sfi11 (40/40; 100%)
33	26966	27439	157	18.0	6.2	—	ORF157 phage Sfi19 (157/157; 100%)
34	27436	28137	233	26.1	6.6	SSAP	ORF9 phage O1205 (89/235; 37%)
35	28094	29431	445	50.9	8.8	Helicase	ORF443 phage Sfi21 (383/441; 86%)
36	29438	29893	151	17.2	4.9	—	ORF151 phage Sfi19 (150/151; 99%)
37	29896	30711	271	30.4	5.8	Replication protein	ORF271 phage Sfi21 (264/271; 97%)
38	30698	32215	505	59.4	8.1	Primase	ORF504 phage Sfi11 (457/504; 90%)
39	32466	32786	106	12.1	9.9	—	ORF106 phage Sfi11 (103/105; 98%)
40	32770	33021	83	9.5	8.0	—	ORF89 phage Sfi19 (61/78; 78%)
41	33029	33184	51	6.3	5.6	—	ORF40 phage DT1 (45/51; 88%)
42	33185	33697	170	19.5	6.3	DNA binding	ORF42 phage DT1 (117/166; 70%)
43	33666	33992	108	12.1	9.2	—	ORF43 phage DT1 (55/84; 65%)
44	33996	34703	235	27.6	9.1	—	ORF235 phage Sfi19 (230/235; 97%)

[a] —, unknown function; TMP, tape measure protein; SSAP, single-strand annealing protein.

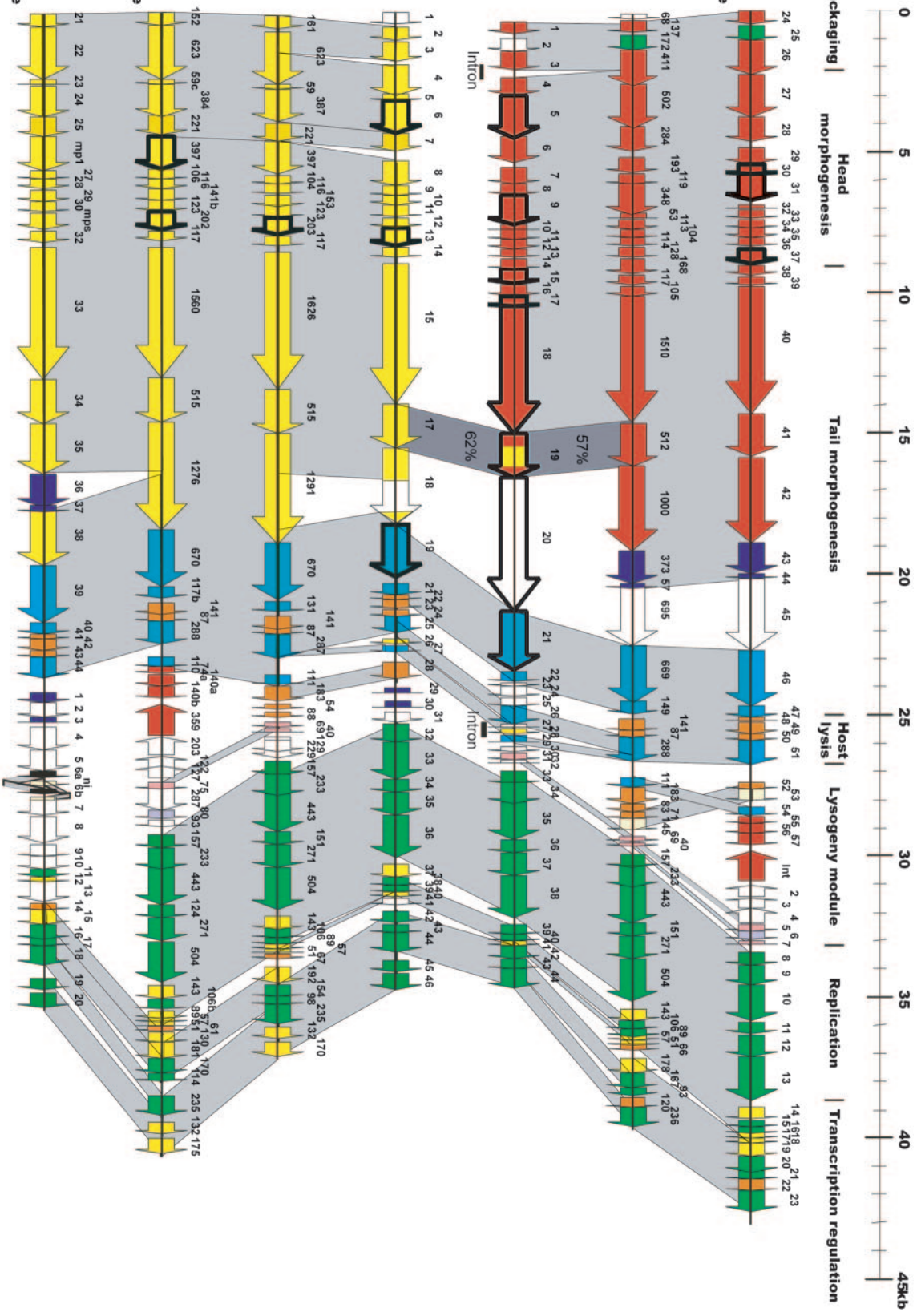


FIG. 3. Alignment of the genetic maps of all completely sequenced *S. thermophilus* phage genomes. The modular regions of the genomes coding for distinct functions are indicated above the maps. Deduced proteins sharing more than 50% amino acid identity are represented using the same color and are linked using grey shading whenever possible. ORFs with unique sequences are displayed in white. Genes coding for proteins identified by N-terminal sequencing or MALDI-TOF are identified by thick lines.

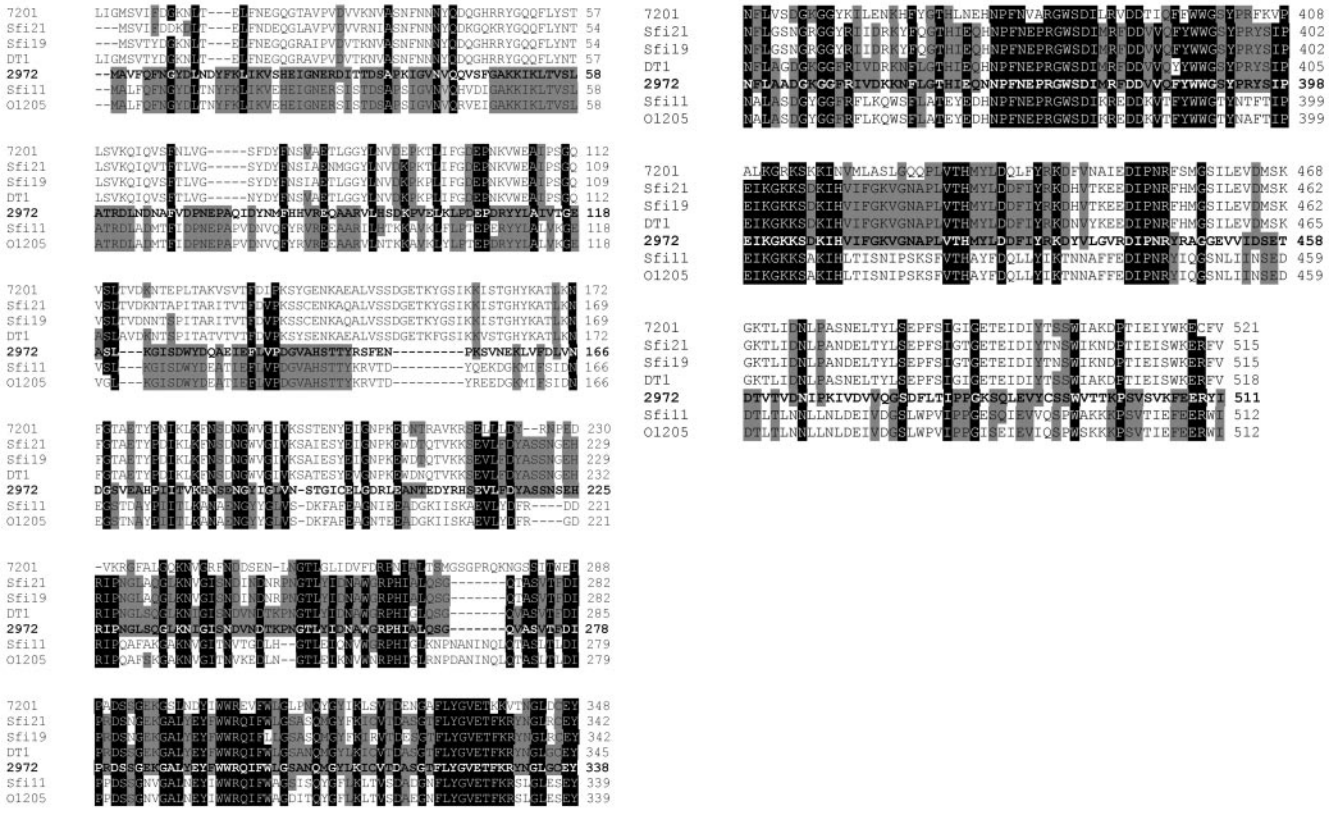


FIG. 4. Alignment of ORF19 of phage 2972 with similar proteins found in *S. thermophilus* phages 7201, Sfi21, Sfi19, DT1, Sfi11, and O1205. Amino acids conserved in six or seven aligned sequences are identified by black shading. Amino acids that are conserved in five or fewer sequences are identified by gray shading.

prophage in an *S. thermophilus* host, as demonstrated for other LAB phages (6, 24, 51). Alternatively, phage 2972 may have started out as a temperate phage that became virulent following deletion(s) and/or rearrangement(s) leading to a nonfunctional lysogeny module. It has been demonstrated that lytic phages can emerge after several passages of the temperate *S. thermophilus* phage Sfi21 on an indicator strain (11, 13).

Morphogenesis. Comparative genomic analysis has clearly demonstrated the presence of two clusters of morphogenesis genes in *S. thermophilus* phages (9, 20). These two clusters support the existence of two *S. thermophilus* phage groups, the *cos* and *pac* types. The morphogenesis genes of phage 2972 are in line with this grouping, as they are homologous with the morphogenesis genes of the *pac*-type phages Sfi11 and O1205 (Fig. 3). Because the morphogenesis module of *S. thermophilus* phages has already been extensively described elsewhere (9, 20), we will focus here on the divergences and novelties observed in phage 2972.

The putative tail protein ORF19 is one of the most intriguing gene products of the deduced proteome of phage 2972. As shown in Fig. 4, many conserved amino acids have been found in other *S. thermophilus* phage proteins that are similar to ORF19 of phage 2972. In general, the N- and C-terminal portions of these proteins are similar among members of the *cos*- and *pac*-type groups. However, the central region of proteins similar to ORF19 of phage 2972 (approximately aa 225 to aa 410) is conserved in both groups. ORF19 of phage 2972 shares 62% amino acid identity (282/449) with ORF17 of the *cos*-type phage DT1 and 57% identity (298/518) with ORF512 of the *pac*-type phage Sfi11 (Table 1). Overall, ORF19 of phage 2972 may thus be a hybrid structural protein that connects the two *S. thermophilus* phage groups.

Comparison of ORF20 with the deduced proteome of the other *pac*-type phages revealed that it was the most divergent structural protein. The function of ORF20 may be to recognize the specific phage receptor on the streptococcal surface. Indeed, ORF20 exhibited some degree of identity with the receptor-binding protein of the *S. thermophilus cos*-type phage MD2 (23). Receptor-binding proteins from phages infecting low-G+C-content gram-positive bacteria usually contain collagen-like repeat motifs at their C termini (60). This motif appears to be characteristic of collagen molecules, and its biological function is to provide elasticity and confer stability on the triple helix structure (4). Six collagen-like repeats were found in ORF20 of phage 2972 (Fig. 5). Three variable regions are found in ORF20, and two of them (VR1 and VR3) are flanked by collagen-like repeats as observed in other streptococcal phages (23). Interestingly, VR1 of ORF20 is also present in ORF38 of the *cos*-type phage 7201 and VR3 is also present in ORF45 of the *pac*-type phage O1205 (Fig. 5). These observations illustrate the modular organization of the putative receptor-binding proteins. As also shown in Fig. 5, the putative receptor-binding proteins are unique to each *cos*- and *pac*-type phage, except for the conserved C-terminal regions of

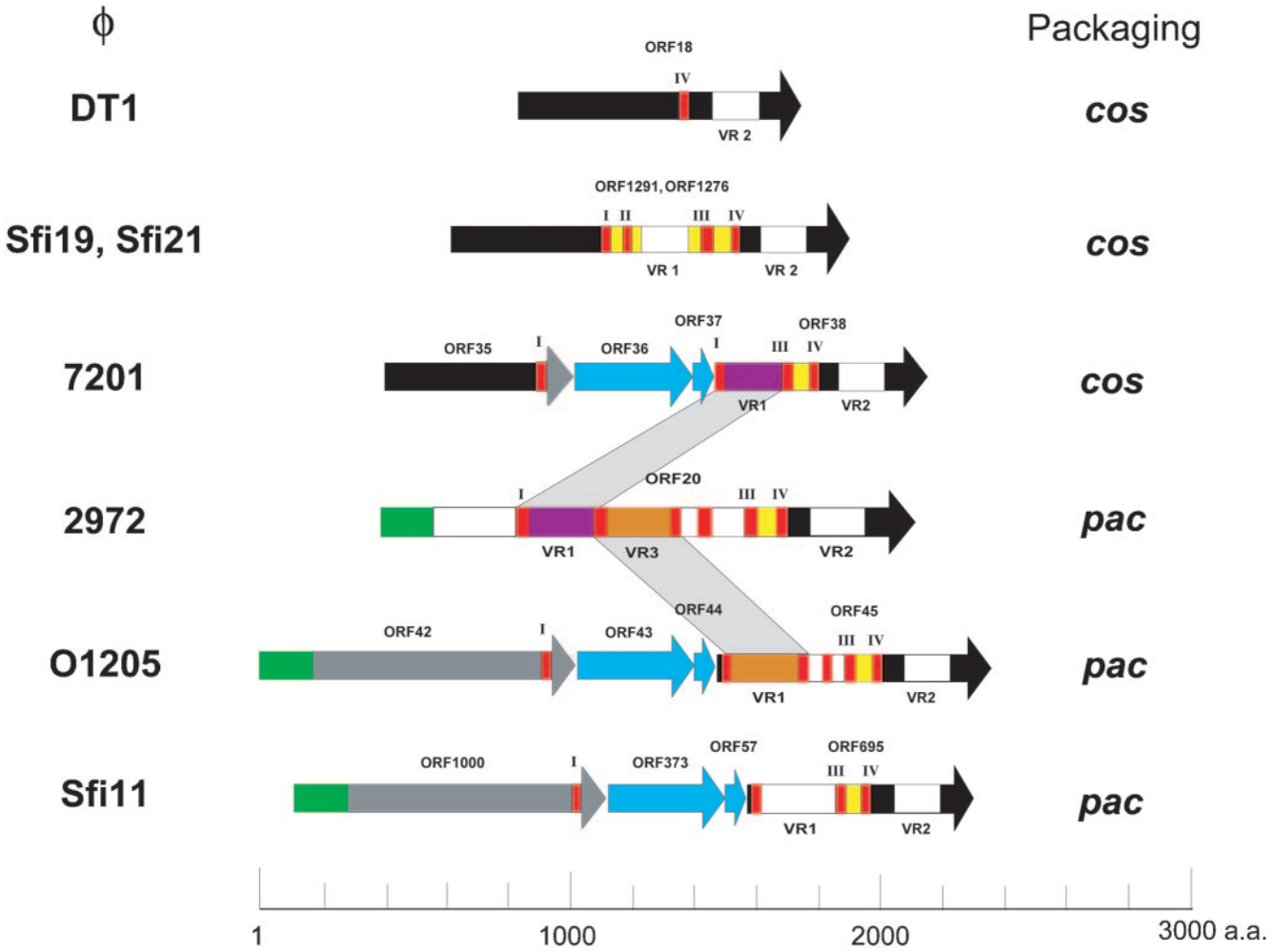
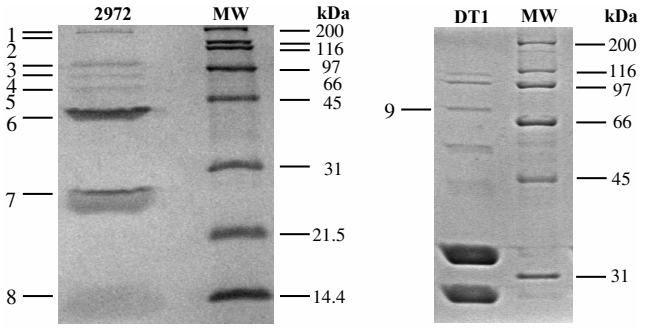


FIG. 5. Schematic illustration representing the alignment of the proteins possibly involved in host recognition of seven *S. thermophilus* phages. The same color indicates more than 80% similarity. Collagen-like repeats are shown in red. VR1, VR2, and VR3 indicate variable regions 1, 2, and 3, respectively. Regions with unique sequences are shown in white.

the deduced proteins. These findings are consistent with the fact that phages O1205, Sfi11, Sfi19, DT1, and 7201 cannot propagate on the host strain of phage 2972 (data not shown).

Protein composition of phage 2972. CsCl-purified phage particles were analyzed by SDS-PAGE in order to identify the proteins in the virion structure (Fig. 6). Eight structural proteins were identified by N-terminal sequencing and/or MALDI-TOF. They included the putative portal protein (ORF5), the major capsid protein (ORF9), the major tail proteins (ORF15 and ORF17), three putative minor tail proteins (ORF18, ORF19, and ORF21), and the putative receptor-binding protein (ORF20).

The identification of ORF21 in the structure of phage 2972 is interesting, as it indicates that this gene is part of the morphogenesis module. Previous genomic analyses were inconclusive in predicting its classification as a structural or a nonstructural protein (20). According to the comparative analysis presented in Table 1 and Fig. 3, this structural protein appears to be relatively conserved in *S. thermophilus cos*- and *pac*-type phages. For example, ORF21 of phage 2972 shares 52% identity (345/663 amino acids) with ORF19 of the *cos*-type phage



#	MW (kDa) Calculated	MW (kDa) Estimated	MALDI-TOF or N-terminal sequencing	Putative function
1	177	139	MALDI-TOF	ORF20, Anti-receptor
2	153	110	MALDI-TOF	ORF18, Tail protein
3	74	77	MALDI-TOF	ORF21, Tail protein
4	58	72	AVFQFNGYDLN	ORF19, Tail protein
5	57	68	MALDI-TOF	ORF5, Portal protein
6	37	40	GLIYDKVTASNIAGY	ORF9, Major capsid protein
7	19	29	ADTNKEALLG	ORF15, Major tail protein
8	10	13	MALDI-TOF	ORF17, Tail protein
9	76	78	VEFWSNND	ORF19, Tail protein

FIG. 6. Protein profiles of phages 2972 (*pac* type) and DT1 (*cos* type) as determined on SDS-PAGE gels stained with Coomassie blue. MW, molecular weight markers.

DT1 (Table 1). This is in contrast with the current view that these two groups of phages have different sets of morphogenesis genes. To confirm that ORF19 of phage DT1 is also present in the virion structure, CsCl-purified phage particles were analyzed by SDS-PAGE (Fig. 6). By using N-terminal sequencing, a protein of ~76 to 78 kDa was identified as ORF19, confirming that this protein is indeed present in the structure of this *S. thermophilus cos*-type phage.

Introns in the phage 2972 genome. As indicated above, sequence analysis suggested the presence of two introns in the genome of phage 2972. The first intron has been located between *orf3* and *orf4* within the gene coding for the putative terminase large subunit (*terL*-I), while the second intron has been found between *orf26* and *orf29*, interrupting the endolysin-encoding gene (*lys*-I). An intron interrupting the endolysin-encoding gene has already been characterized in other *S. thermophilus* phages (25), and a group I intron that interrupts the gene encoding the large subunit of the terminase of the virulent phage LL-H of *Lactobacillus delbrueckii* has also been reported (47).

To test for in vivo splicing of RNA transcripts, RT-PCR experiments were performed. By using specific primers located in *orf26* and *orf29* of phage 2972, a PCR product of 590 bp was amplified from the phage genomic DNA, while a 148-bp amplicon was obtained using the cDNA as a template (data not shown). Sequence analysis of the PCR products revealed that the splicing occurred after a uridine residue (coordinate 25280 within *orf26*) as well as after a guanosine residue (coordinate 25724 upstream of *orf29*), resulting in the excision of a 442-bp intron. The endolysin-encoding gene is thus 843 bp long, and it is believed to code for a 281-aa protein that possesses 79% identity (219/275) with the endolysin of *S. thermophilus* phage S3b. The splicing occurred at exactly the same site as that observed for the intron of phage S3b (25). The secondary structure of this intron was relatively similar to that of phage S3b (25), except that the P3.1 and P3.2 stems were included in the prediction (Fig. 7A). The P7.2 stem folding retained was similar to that of phage SPO1. The main nucleotide discrepancies were located within the P8 looped-out region. One other notable difference was in the P7.2 stem, where a guanine was present at coordinate 25438, compared to an adenine in phage SB3. An additional adenosine was also found in the looped-out region of P3.2, creating a short ORF (*orf27*, coding for 40 amino acids).

For the second intron, the in vivo splicing of the mRNA was demonstrated using specific primers located in *orf3* and *orf4*. A 506-bp DNA fragment was amplified from the phage genomic DNA, while a 199-bp PCR product was obtained using cDNA as a template (Fig. 8A). Sequence analysis of the PCR products revealed that the splicing occurred after a uridine and a guanosine residue (coordinates 2098 and 2406, respectively). The 307-bp intron had a G+C content similar to the rest of the phage genome, did not contain an ORF, and had 83% identity (55/66) with a group IA1 intron found in the chloroplastic gene coding for the rRNA large subunit of a *Chlamydomonas* sp. (GenBank accession no. L43539). After the mRNA splicing, the *terL* gene of phage 2972 was 1,236 bp long and coded for a 411-aa protein (Fig. 8B) that possessed 96% identity (394/410) with the intron-free ORF411 of *S. thermophilus* phage Sfi11 and 95% identity (393/410) with ORF26 of phage O1205. Analysis of the DNA region flanking the integration site of this intron in phage 2972 with the corresponding region in phages Sfi11 and O1205 revealed sequence variations close to the integration site that may explain the absence of an intron in these two phages (Fig. 8C). The secondary-structure prediction (15, 18, 46) of the terminase intron possessed all the canonical group I intron features (P1 to P9) that are required to form the catalytic core of the intron and that are essential for self-splicing activity (Fig. 7B). The P7.1 and P7.1a stem-loops between P3 and P7 are characteristics of a subgroup IA1 intron. An internal guide sequence that could bring P1 and P10 into close proximity to facilitate the splicing process (16, 17) was also recognized.

Lastly, structural differences were noted between the two introns found in the genome of phage 2972. *terL*-I possessed the P2, P5a, and P9.1 stems, while *lys*-I had the P3.1, P3.2, and P7.2 stems as well as an ORF (Fig. 7). The two introns thus belong to different subgroups (IA1/terminase, IA2/endolysin) and are probably from different sources.

DISCUSSION

Phage sensitivity and exopolysaccharide production. We demonstrated here that several phages can infect an EPS-producing *S. thermophilus* strain such as RD534. Some studies have previously suggested that cell surface phage receptors can be blocked by the loosely bound EPS produced by some bacterial strains, thus protecting the cells against phage infections (7, 29, 30, 65). Brüssow et al. (10) previously reported the isolation of phages (including Sfi11, Sfi19, and Sfi21) from ropy strains of *S. thermophilus*. We also recently reported that the EPS-producing *S. thermophilus* strain MR-1C and its EPS-negative derivative were both sensitive to the same three *pac*-type phages (7). Several phages can also infect EPS-producing *Lactococcus lactis* strains (21). Taken together, these results clearly indicate that the production of EPS does not confer potent protection against phage infections.

Genome of phage 2972. We presented the seventh complete genome of an *S. thermophilus* phage, and the third from the *pac*-type group. With its 34,704 bp, phage 2972 possesses the shortest *S. thermophilus* phage genome analyzed so far. As previously reported, the lysogeny region may be a recombination hot spot in *S. thermophilus* phages (40). The presence of a *cro*-like repressor gene suggests that virulent phage 2972 is derived from a temperate phage. It is noteworthy that a *cro*-like repressor gene is present in many virulent *S. thermophilus* phage genomes (40), and one might wonder whether the *cro*-like repressor still plays a role in the lytic cycle of virulent phages, particularly in lysogenic hosts.

Another noteworthy size variation among the three *pac*-type phages was noted in the genome area coding for the tail proteins. Because of its position in the genome and its similarity to ORF18 of phage DT1, which was experimentally shown to be involved in the recognition of *S. thermophilus* hosts (23), ORF20 is most likely involved in host recognition.

Introns of phage 2972. Two group I introns have been found in the genome of phage 2972. They were located in the genes coding for the terminase large subunit and the endolysin, where introns have been found in other phage genomes (25, 47). Phage introns seem to target crucial genes in the phage

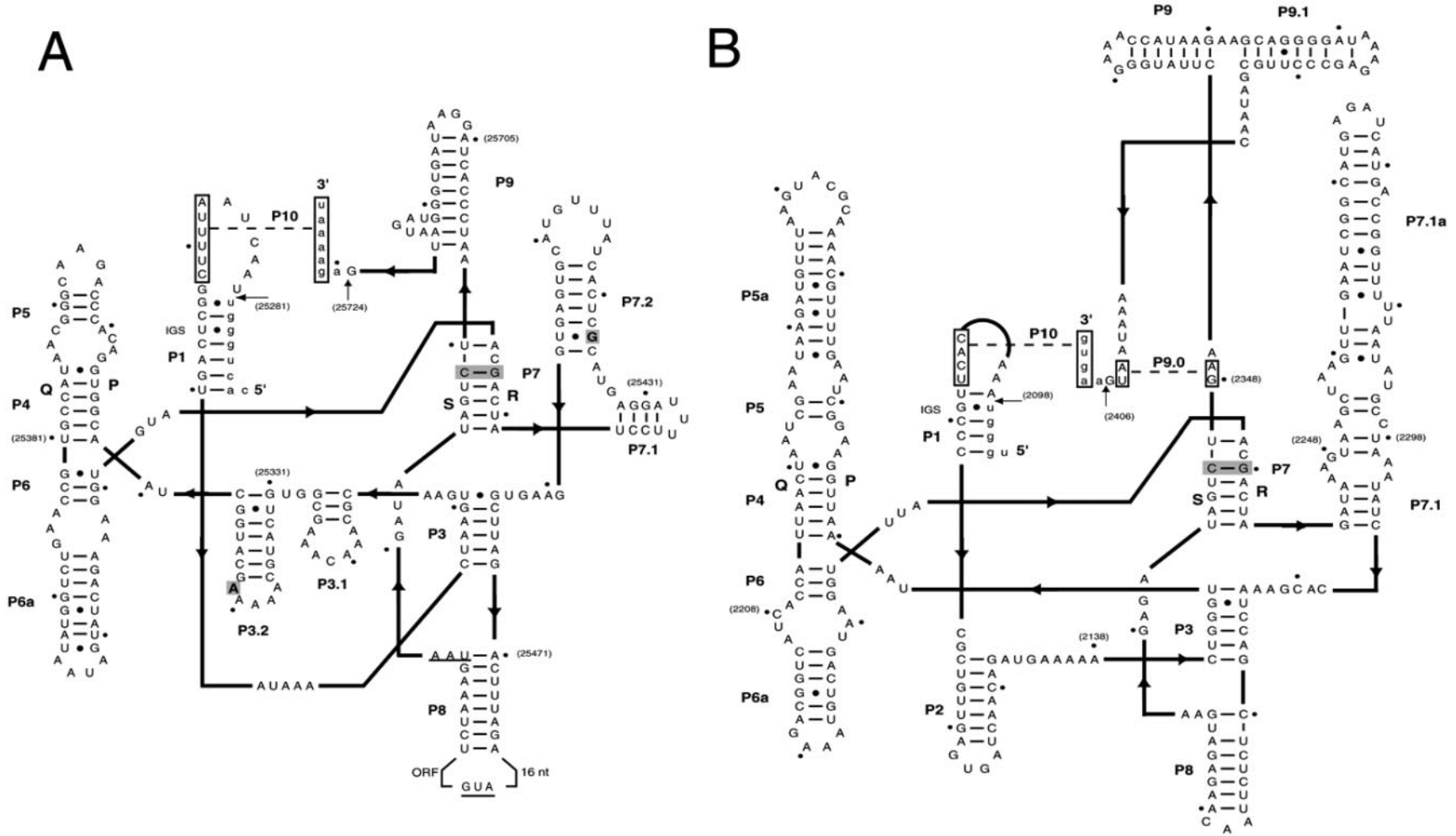


FIG. 7. Secondary-structure predictions of the two introns in the genome of phage 2972. The secondary-structure representation was made using a two-dimensional structural diagram (15, 18, 46). Arrows indicate the 5′ and 3′ splicing sites. Lower- and uppercase letters denote the exon and intron sequences, respectively. Boxed sequences indicate the regions that anneal to form P9.0 and P10. The shaded nucleotides in P7 represent the putative guanosine-binding site. Bold lines show connections between intron structure domains, with pointers indicating the 5′-to-3′ direction. IGS, internal guide sequence. Numbers in parentheses represent the nucleotide position on the phage 2972 genome. (A) Intron in the gene coding for the endolysin. The start and stop codons of ORF28 are underlined, and the two nucleotides that differ from the S3b intron are boldfaced. (B) Intron in the gene coding for the large subunit of the terminase.

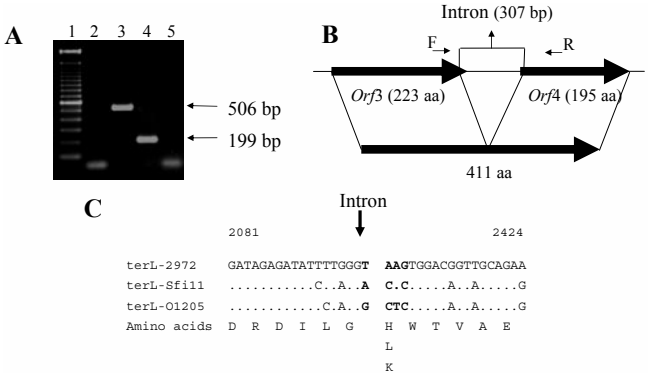


FIG. 8. Characterization of the intron between the genes coding for the terminase large subunit of phage 2972. (A) In vivo splicing of 2972 intron RNA (terminase large-subunit gene). Lane 1, 100-bp marker (Invitrogen); lane 2, negative control without DNA, cDNA, or RNA; lane 3, PCR product obtained with 2972 DNA as a template; lane 4, PCR product obtained with RNA isolated from 2972-infected *S. thermophilus* cells (cDNA); lane 5, PCR product (no reverse transcription) obtained with RNA isolated from 2972-infected cells. (B) Phage 2972 genomic region containing the intron and the two ORFs coding for the terminase large subunit. Large, thick arrows indicate open reading frames; the numbers of amino acids (aa) in the deduced proteins are given. Splicing of the 307-bp intron resulted in a 411-aa protein. Small arrows represent the primers used for PCR and sequencing. (C) Nucleotide sequence alignment of the regions flanking the splicing site with the corresponding intron-free regions in two other *pac*-type phages (Sfi11 [*orf411*] and O1205 [*orf26*]). Differences relative to the phage 2972 sequence are indicated. Vertical arrow indicates the intron insertion site. The nucleotide positions are based on the 2972 genomic sequence (GenBank accession no. AY699705).

genome (DNA polymerase, thymidylate synthase, ribonucleotide reductase, structural proteins, terminase large subunit, and endolysin), while introns in eubacteria are in tRNA genes and all belong to a different subgroup (IC3) (25, 36). The distribution and the homing of group I introns have been studied for T-even-like bacteriophages, and the results suggested that these introns share a recent common ancestor that has spread horizontally throughout the phage population, most likely via mixed infections (57). It is plausible that such mixed infections also account for the two distinct introns found in the genome of phage 2972. In this regard, *lys*-I has been found in numerous other *S. thermophilus* phages (25), and dairy environments are known to contain several distinct phages (8–10). In addition, phage intron homing or invasion appears to be a very successful mechanism among the phages of low-G+C-content gram-positive bacteria. It has been detected in phages of *Bacillus*, *Lactobacillus*, *Lactococcus*, *Staphylococcus*, and *Streptococcus* species. So far, in gram-negative bacteria, introns have been observed only in the T-even phage group (57).

The secondary structures revealed that these two introns belong to a distinct subgroup (IA1/terminase, IA2/endolysin). Previous studies have shown that almost all phage introns belong to the IA2 subgroup and possess a P7.2 stem (3, 27, 28, 35, 37, 48, 58). Despite structural similarities with the introns of the IA2 subgroup such as the *orf142*-I2 of the *Staphylococcus aureus* Twort phage (34) and the T4 *nrdB*, the *terL* intron structure of 2972 possesses the unusual P7.1 and P7.1a stems, which place it in the IA1 subgroup. This is one of the first phage introns of the IA1 subgroup to be characterized. Recently, a subgroup-IA1 intron was uncovered in the genome of the *Synechococcus* cyanophage S-PM-2 (48). It interrupted a gene (*psbA*) coding for a core component of the photosynthetic reaction center PSII (photosystem II) (48). Group I introns range in size from 200 to 3,000 bp, depending on the length of the peripheral sequence and on whether or not they contain ORFs (33). The *terL*-I is a relatively short, 307-bp intron that does not contain an ORF, in contrast to phage Twort *orf142*-I1, -I2, and -I3 (34) and phage S-PM2 *psdA*-I (48).

Structural phage proteins and phage classification. The identification of a conserved structural protein (ORF21) in both groups of *S. thermophilus* phages (*cos* and *pac* types) is interesting considering the fact that these two groups are regarded as two lineages of the family *Siphoviridae* (54). Moreover, the discovery of the structural hybrid protein ORF19 in phage 2972 suggests that recombination may have occurred within these two distinct structural gene clusters, possibly during mixed infections. While the functions of ORF19 and ORF21 remain to be determined, the position of their genes in the genome suggests that they are tail-related proteins. *orf20* possibly codes for the receptor-binding protein of phage 2972. It is not known whether these three proteins interact with each other, but it appears that this area of the genome has the flexibility and potential for domain shuffling. Such rearrangements may favor the formation of functional recombinant phages with a modified host range.

Despite the finding of a common structural protein, the current classification of *S. thermophilus* phages based on DNA packaging mechanisms (*cos* and *pac*) and structural protein composition remains valid. *S. thermophilus* phages were first classified into a single DNA homology group based on DNA-DNA hybridization data, a classification supported by their similar morphology (45). Since then, comparative analyses of a growing number of streptococcal phage sequences have confirmed these conserved genomic regions. The presence of hybrid and conserved structural proteins provides new evidence to support the hypothesis of a common ancestor. The subsequent sorting of these phages into two phage groups is also quite evident based on the overall makeup of their genomes and proteomes. A view toward practical applications is perhaps the most compelling reason for maintaining the current classification of *S. thermophilus* phages in two groups. In our experience, most phage-sensitive *S. thermophilus* strains are infected either by *cos*-type phages or by *pac*-type phages. A given phage-sensitive *S. thermophilus* strain is rarely infected by members of both phage groups. For example, *S. thermophilus* RD534 has been infected only by *pac*-type phages, such as 2972. Consequently, it has been possible to design or rotate starter cultures based on group sensitivity.

On a larger scale, a number of proposals have been put forward recently to modify the current approach of the International Committee on Taxonomy of Viruses (35, 54, 55). The study reported here supplies evidence that proposals based solely on structural gene modules may not be the answer. Lastly, it should be remembered that such proposals not only need to be scientifically sound but should also be useful from an applied perspective.

ACKNOWLEDGMENTS

We thank Diane Montpetit for assistance with electron microscopy, Jean-François Pombert and Christian Otis for assistance in secondary-structure analysis of the introns, Dennis Romero for valuable discussions, and Gene Bourgeau for editorial assistance.

This work was funded, in part, by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. **Ackermann, H.-W.** 1999. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. **51:**135–201.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Bechhofer, D. H., K. K. Hue, and D. A. Shub.** 1994. An intron in the thymidylate synthase gene of *Bacillus* bacteriophage β22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame. Proc. Natl. Acad. Sci. USA **91:**11669–11673.
4. **Beck, K., and B. Brodsky.** 1998. Supercoiled protein motifs: the collagen triple-helix and the α-helical coiled coil. J. Struct. Biol. **122:**17–29.
5. **Beres, S. B., G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, N. D. Mammarelle, M.-Y. Liu, J. C. Smoot, S. F. Porcella, L. D. Parkins, D. S. Campbell, T. M. Smith, J. K. McCormick, D. Y. M. Leung, P. M. Schlievert, and J. M. Musser.** 2002. Genome sequence of a serotype M3 strain of a group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. USA **99:**10078–10083.
6. **Bouchard, J. D., and S. Moineau.** 2000. Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. Virology **270:**65–75.
7. **Broadbent, J. R., D. J. McMahon, D. L. Welker, C. J. Oberg, and S. Moineau.** 2003. Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. J. Dairy Sci. **86:**407–423.
8. **Brüssow, H.** 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. **55:**283–303.
9. **Brüssow, H., and F. Desiere.** 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. **39:** 213–222.
10. **Brüssow, H., M. Fremont, A. Bruttin, J. Sidoti, A. Constable, and V. Fryder.** 1994. Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. Appl. Environ. Microbiol. **60:**4537–4543.
11. **Bruttin, A., and H. Brüssow.** 1996. Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. Virology **219:**96–104.
12. **Bruttin, A., F. Desiere, S. Lucchini, S. Foley, and H. Brüssow.** 1997. Characterization of the lysogeny DNA module from the temperate *Streptococcus thermophilus* bacteriophage Sfi21. Virology **233:**136–148.
13. **Bruttin, A., S. Foley, and H. Brüssow.** 1997. The site-specific integration system of the temperate *Streptococcus thermophilus* bacteriophage Sfi21. Virology **237:**148–158.
14. **Bruttin, A., S. Foley, and H. Brüssow.** 2002. DNA-binding activity of the *Streptococcus thermophilus* phage Sfi21 repressor. Virology **303:**100–109.
15. **Burke, J. M., M. Belfort, T. R. Cech, R. W. Davies, R. J. Schweyen, D. A. Shub, J. W. Szostak, and H. F. Tabak.** 1987. Structural conventions for group I introns. Nucleic Acids Res. **15:**7217–7221.
16. **Cech, T. R.** 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene **73:**259–271.
17. **Cech, T. R.** 1990. Self-splicing of group I introns. Annu. Rev. Biochem. **59:**543–568.
18. **Cech, T. R., S. H. Damberger, and R. R. Gutell.** 1994. Representation of the secondary and tertiary structure of group I introns. Nat. Struct. Biol. **1:**273–280.
19. **Delcour, J., T. Ferain, and P. Hols.** 2000. Advances in the genetics of thermophilic lactic acid bacteria. Curr. Opin. Biotechnol. **11:**497–504.
20. **Desiere, F., S. Lucchini, C. Canchaya, M. Ventura, and H. Brüssow.** 2002. Comparative genomics of phages and prophages in lactic acid bacteria. Antonie Leeuwenhoek **82:**73–91.
21. **Deveau, H., M.-R. van Calsteren, and S. Moineau.** 2002. The effect of exopolysaccharides on phage-host interactions in *Lactococcus lactis*. Appl. Environ. Microbiol. **68:**4364–4369.
22. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
23. **Duplessis, M., and S. Moineau.** 2001. Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. Mol. Microbiol. **41:**325–336.
24. **Durmaz, E., and T. R. Klaenhammer.** 2000. Genetic analysis of chromosomal regions of *Lactococcus lactis* acquired by recombinant lytic phages. Appl. Environ. Microbiol. **66:**895–903.
25. **Foley, S., A. Bruttin, and H. Brüssow.** 2000. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. J. Virol. **74:**611–618.
26. **Foley, S., S. Lucchini, M. C. Zwahlen, and H. Brüssow.** 1998. A short noncoding viral DNA element showing characteristics of a replication origin confers bacteriophage resistance to *Streptococcus thermophilus*. Virology **250:**377–387.
27. **Goodrich-Blair, H., V. Scarlato, J. M. Gott, M. Q. Xu, and D. A. Shub.** 1990. A self-splicing group I intron in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. Cell **63:**417–424.
28. **Hirayama, K., and J. Rafter.** 2000. The role of probiotic bacteria in cancer prevention. Microbes Infect. **2:**681–686.
29. **Hughes, K. A., I. W. Sutherland, and M. V. Jones.** 1998. Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase. Microbiology **144:**3039–3047.
30. **Kang, K. S., and I. W. Cottrell.** 1979. Polysaccharides, p. 417–481. *In* H. J. Peppler and D. Perlman (ed.), Microbial technology: microbial processes, 2nd ed., vol. 1. Academic Press, Inc., New York, N.Y.
31. **Labrie, S., and S. Moineau.** 2000. Multiplex PCR for detection and identification of lactococcal bacteriophages. Appl. Environ. Microbiol. **66:**987–994.
32. **Labrie, S., and S. Moineau.** 2002. Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. Virology **296:**308–320.
33. **Lambowitz, A. M., and M. Belfort.** 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. **62:**587–622.
34. **Landthaler, M., and D. A. Shub.** 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc. Natl. Acad. Sci. USA **96:**7005–7010.
35. **Lawrence, J. G., G. F. Hatfull, and R. W. Hendrix.** 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. **184:**4891–4905.
36. **Lazarevic, V., B. Soldo, A. Dusterhoft, H. Hilbert, C. Mauel, and D. Karamata.** 1998. Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPβ. Proc. Natl. Acad. Sci. USA **95:**1692–1697.
37. **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.
38. **Lemoine, J., F. Chirat, J.-M. Wieruszeski, G. Strecker, N. Favre, and J.-R. Neeser.** 1997. Structural characterization of the exocellular polysaccharides produced by *Streptococcus thermophilus* SFi39 and SFi12. Appl. Environ. Microbiol. **63:**3512–3518.
39. **Lucchini, S., F. Desiere, and H. Brüssow.** 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage ϕSfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. Virology **246:**63–73.
40. **Lucchini, S., F. Desiere, and H. Brüssow.** 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J. Virol. **73:**8647–8656.
41. **Lucchini, S., F. Desiere, and H. Brüssow.** 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria. Virology **263:**427–435.
42. **Lucchini, S., F. Desiere, and H. Brüssow.** 1999. The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison of *cos*-site phages Sfi19 and Sfi21. Virology **260:** 232–243.
43. **Lysenko, A. M., S. G. Botina, V. I. Ganina, and V. V. Sukhodolets.** 2001. DNA relatedness, divergence, and sibling species of the lactic acid bacterium *Streptococcus thermophilus*. Microbiology **70:**59–63.
44. **Marshall, V. M., A. P. Laws, Y. Gu, F. Levander, P. Radstrom, L. De Vuyst, B. Degeest, F. Vaningelgem, H. Dunn, and M. Elvin.** 2001. Exopolysaccharide-producing strains of thermophilic lactic acid bacteria cluster into groups according to their EPS structure. Lett. Appl. Microbiol. **32:**433–437.
45. **Mercenier, A., P. H. Pouwels, and B. M. Chassy.** 1994. Genetic engineering of lactobacilli, leuconostocs, and *Streptococcus thermophilus*, p. 253–293. *In* M. J. Gasson and W. M. DeVos (ed.), Genetics and biotechnology of lactic acid bacteria. Blackie Academic and Professional, Glasgow, United Kingdom.
46. **Michel, F., and E. Westhof.** 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. **216:**585–610.
47. **Mikkonen, M., and T. Alatossava.** 1995. A group I intron in the terminase gene of *Lactobacillus delbrueckii* subsp. *lactis* phage LL-H. Microbiology **141:**2183–2190.
48. **Millard, A., M. R. Clokie, D. A. Shub, and N. H. Mann.** 2004. Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. Proc. Natl. Acad. Sci. USA **101:**11007–11012.
49. **Moineau, S.** 1999. Applications of phage resistance in lactic acid bacteria. Antonie Leeuwenhoek **76:**377–382.

50. **Moineau, S., J. Fortier, H. W. Ackermann, and S. Pandian.** 1992. Characterization of lactococcal bacteriophages from Québec cheese plants. Can. J. Microbiol. **38:**875–882.
51. **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.
52. **Moineau, S., D. Tremblay, and S. Labrie.** 2002. Phages of lactic acid bacteria: from genomics to industrial applications. ASM News **68:**388–393.
53. **Obregon, V., J. L. Garcia, E. Garcia, R. Lopez, and P. Garcia.** 2003. Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. J. Bacteriol. **185:**2362–2368.
54. **Proux, C., D. van Sinderen, J. Suarez, P. Garcia, V. Ladero, G. F. Fitzgerald, F. Desiere, and H. Brüssow.** 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. **184:**6026–6036.
55. **Rohwer, F., and R. Edwards.** 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. **184:**4529–4535.
56. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
57. **Sandegren, L., and B. M. Sjoberg.** 2004. Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. J. Biol. Chem. **279:**22218–22227.
58. **Shub, D. A., J. M. Gott, M. Q. Xu, B. F. Lang, F. Michel, J. Tomaschewski, J. Pedersen-Lane, and M. Belfort.** 1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proc. Natl. Acad. Sci. USA **85:**1151–1155.
59. **Siboo, I. R., B. A. Bensing, and P. M. Sullam.** 2003. Genomic organization and molecular characterization of SM1, a temperate bacteriophage of *Streptococcus mitis*. J. Bacteriol. **185:**6968–6975.
60. **Smith, M. C., N. Burns, J. R. Sayers, J. A. Sorrell, S. R. Casjens, and R. W. Hendrix.** 1998. Bacteriophage collagen. Science **279:**1834.
61. **Stanley, E., G. F. Fitzgerald, C. Le Marrec, B. Fayard, and D. van Sinderen.** 1997. Sequence analysis and characterization of φO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. Microbiology **143:**3417–3429.
62. **Stanley, E., L. Walsh, A. van der Zwet, G. F. Fitzgerald, and D. van Sinderen.** 2000. Identification of four loci isolated from two *Streptococcus thermophilus* phage genomes responsible for mediating bacteriophage resistance. FEMS Microbiol. Lett. **182:**271–277.
63. **Sturino, J. M., and T. R. Klaenhammer.** 2002. Expression of antisense RNA targeted against *Streptococcus thermophilus* bacteriophages. Appl. Environ. Microbiol. **68:**588–596.
64. **Sturino, J. M., and T. R. Klaenhammer.** 2004. Antisense RNA targeting of primase interferes with bacteriophage replication in *Streptococcus thermophilus*. Appl. Environ. Microbiol. **70:**1735–1743.
65. **Sutherland, I. W., K. A. Hughes, L. C. Skillman, and K. Tait.** 2004. The interaction of phage and biofilms. FEMS Microbiol. Lett. **232:**1–6.
66. **Tremblay, D., and S. Moineau.** 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. Virology **255:**63–76.
67. **Walker, D. C., H. S. Girgis, and T. R. Klaenhammer.** 1999. The groESL chaperone operon of *Lactobacillus johnsonii*. Appl. Environ. Microbiol. **65:**3033–3041.
68. **Wang, I.-N., D. L. Smith, and R. Young.** 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. **54:**799–825.