

Comparative Genomics of *Streptococcus thermophilus* Phage Species Supports a Modular Evolution Theory

SACHA LUCCHINI, FRANK DESIERE, AND HARALD BRÜSSOW*

Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland

Received 14 April 1999/Accepted 18 June 1999

The comparative analysis of five completely sequenced *Streptococcus thermophilus* bacteriophage genomes demonstrated that their diversification was achieved by a combination of DNA recombination events and an accumulation of point mutations. The five phages included lytic and temperate phages, both *pac* site and *cos* site, from three distinct geographical areas. The units of genetic exchange were either large, comprising the entire morphogenesis gene cluster, excluding the putative tail fiber genes, or small, consisting of one or maximally two genes or even segments of a gene. Many indels were flanked by DNA repeats. Differences in a single putative tail fiber gene correlated with the host ranges of the phages. The predicted tail fiber protein consisted of highly conserved domains containing conspicuous glycine repeats interspersed with highly variable domains. As in the T-even coliphage adhesins, the glycine-containing domains were recombinational hot spots. Downstream of a highly conserved DNA replication region, all lytic phages showed a short duplication; in three isolates the origin of replication was repeated. The lytic phages could conceivably be derived from the temperate phages by deletion and multiple rearrangement events in the lysogeny module, giving rise to occasional selfish phages that defy the superinfection control systems of the corresponding temperate phages.

The most important mechanism for producing new types of RNA viruses is mutation, mostly in the form of point mutation. For RNA viruses in which the genome is present in several discrete genome segments, new viruses can also be created by reassortment of these segments. This strategy provides a rapid method for the production of viruses with totally new potentials and is thought to be the basis for the dramatic antigenic shifts in influenza virus (48). Recombination, in which a single polynucleotide strand contains sequences which have originated from two parental types, is only infrequently utilized among the RNA viruses. However, examples have been documented for picornaviruses (44), coronaviruses (33), alphaviruses (24, 47), arterivirus (28), and certain plant viruses (13). The two major forces acting upon DNA virus genomes to generate diversity are mutation and recombination. The role of recombination for the generation of diversity has been documented for several DNA viruses (e.g., polyomaviruses and adenoviruses [reviewed in reference 41]). Some DNA viruses exert control over recombination by encoding proteins that enhance recombination between viral genomes and thus speed up their own evolution. This strategy is most prominent in bacterial DNA viruses (reviewed in reference 41). Recombination between related viruses might produce a genome containing a *trans*-acting gene product trying unsuccessfully to interact with a *cis*-acting element derived from another virus. To avoid these difficulties, viral genomes are commonly organized with *cis*-acting elements located near the genes encoding the proteins that bind them. Genes which encode proteins that interact are also frequently located next to each other. Experimental data from lambdoid phages led to the formulation of the modular theory of phage evolution by Botstein (2). In fact, lambdoid phages are related to each other in a mosaic fashion, implying an evolutionary history with significant amounts of horizontal exchange of genetic material (14). The relationship

between lambdoid phages has mainly been developed by heteroduplex mapping techniques (for a recent review, see reference 26). Now that an increasing number of phage genomes have been sequenced, it is possible to ask about the similarity between different phage genomes at the nucleotide level (25). These comparisons should be especially revealing when they are done at the level of whole-genome comparisons and between related phages differing in important aspects of their phenotypes (e.g., lytic versus temperate life style, polypeptide composition, immunity groups, host range, DNA-packaging mechanism, and geographical and ecological origins). Only two such comparisons have been reported. The comparison between the lytic and temperate mycobacteriophages, D29 and L5 (order, *Caudovirales*; family, *Siphoviridae*; genus, “L5-like viruses” [34]), respectively, showed that their genomes differed by a large number of insertions, deletions, and substitutions of genes (21). The comparison of the lytic and temperate *cos* sites containing *Streptococcus thermophilus* phages Sfi19 and Sfi21, respectively, demonstrated that their genomes were similarly organized and differed by gene deletions and insertions and duplication and DNA rearrangement events, in addition to numerous point mutations (32). However, comparisons among more than two related phage genomes are necessary to decipher the processes that shaped the genomes of a given phage species. *S. thermophilus* phages (5, 7) could be suitable objects for such a study. Due to their industrial importance in milk fermentation (36), many *S. thermophilus* phages have been isolated, covering a substantial geographical diversity (reference 7 and references therein). A longitudinal factory survey has documented the ecological dynamics of phage infection (11). All phages characterized so far belong to the same morphological class (*Siphoviridae*, B1 morphotype) and the same DNA homology group (8, 9), but they could be split into two groups: *cos*-site and *pac*-site phages. The two DNA-packaging mechanisms correlated with two distinct structural polypeptide patterns (30) and, in yogurt phage isolates, with host range and serotype (8). Four complete *S. thermophilus* phage genomes have been reported: the temperate *pac*-site phage O1205 (40), the temperate *cos*-site phage Sfi21 (32), and two lytic *cos*-site

* Corresponding author. Mailing address: Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland. Phone: 41 21 785 8676. Fax: 41 21 785 8925. E-mail: harald.brussow@rdls.nestle.com.

TABLE 1. Characteristics of *S. thermophilus* phages^a

Phage	Life cycle	DNA packaging	Origin	Ecology	Host range	Serotype	Immunity ^b	Reference
O1205	Temperate	<i>pac</i>	Greece	Yogurt	Non-LG I and II	NT	NA	40
Sfi11	Lytic	<i>pac</i>	France	Yogurt	LG II	2	NA	31
Sfi21	Temperate	<i>cos</i>	France	Yogurt	LG I	1	+	12, 16–19, 32
Sfi19	Lytic	<i>cos</i>	France	Milk	LG I	1	–	32
Sfi18	Lytic	<i>cos</i>	France	Yogurt	LG I	1	+	This report
DT1	Lytic	<i>cos</i>	Canada	Cheese	NT	NT	NT	44

^a For the definition of host range and lytic group (LG) and of serotype, see references 8 and 31; NT, not tested.

^b The indicated phage is inhibited in its growth (+) on a lysogenic cell containing the Sfi21 prophage, or it is not inhibited on the lysogen (–); NA, nonapplicable. The O1205 lysogen shows no immunity to superinfection. The Sfi21-Sfi19 comparison has been reported previously (32).

phages, the French yogurt isolate Sfi19 (17, 32), and the Canadian cheese isolate DT1 (45). Here we report the complete sequence of the lytic *pac*-site phage Sfi11 (31) and half of the genome sequence of the lytic *cos*-site phage Sfi18. This creates the largest sequence database set for a group of closely related phages infecting the same host species. The phages were chosen to cover differences in life styles (lytic and temperate), DNA packaging mechanisms and structural polypeptide patterns (*pac* site and *cos* site), host ranges, serotypes, susceptibilities to Sfi21 prophage control, ecological environments, and geographical origins. We first used pairwise comparisons of phages differing in defined phenotypes in order to establish associations between the phenotypes and genotypes of the phages. We then used multiple alignments to decipher basic principles of genome diversification in *S. thermophilus* phages.

MATERIALS AND METHODS

Phages, strains, and media. The phages were propagated on their appropriate *S. thermophilus* hosts in lactose M17 broth as described previously (6, 10). *Escherichia coli* JM 101 (Stratagene) was grown in Luria-Bertani broth or on Luria-Bertani broth solidified with 1.5% (wt/vol) agar. Ampicillin, IPTG (isopropyl- β -D-thiogalactopyranoside), and X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) (all from Sigma) were used at concentrations of 100 μ g/ml, 1 mM, and 0.002% (wt/vol), respectively. Pertinent features of the phages are given in Table 1.

DNA techniques. Phage purification and DNA extraction were done as described previously (6, 10). Plasmid DNA was isolated with Qiagen midiplasmid isolation columns. Restriction enzymes were obtained from Boehringer Mannheim and used according to the supplier's instructions.

Sequencing. DNA sequencing was started with universal forward and reverse primers on pUC19 or pNZ124 shotgun clones and continued with synthetic oligonucleotide (18-mer) primers (Microsynth, Balgach, Switzerland). Both strands of the cloned DNA were sequenced by the Sanger method of dideoxy-mediated chain termination with the *fml* DNA sequencing system of Promega (Madison, Wis.). The sequencing primers were end labelled with [γ -³³P]ATP according to the manufacturer's protocol. The thermal cycler (Perkin-Elmer) was programmed at 30 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min.

In addition, pUC19 clones of *Sau*3A-digested phage Sfi11 DNA were sequenced with the Amersham Labstation sequencing kit based on Thermo Sequenase-labelled primer cycle sequencing with 7-deaza-dGTP (RPN2437). Sequencing was done on a Licor 6000L automated sequencer with fluorescence-labelled universal reverse and forward pUC19 primers.

PCR. PCR was used to span regions which were not obtained through random cloning. PCR products were generated with the synthetic oligonucleotide pair designed according to the established ϕ Sfi11 DNA sequence, purified phage DNA, and Super Taq polymerase (Stehelin, Basel, Switzerland). PCR products were purified with the QIAquick-spin PCR purification kit.

Sequence analysis. The Genetics Computer Group (University of Wisconsin) sequence analysis package was used to assemble and analyze the sequences. Nucleotide and predicted amino acid sequences were compared to those in the databases (GenBank, release 109; EMBL [abridged], release 56; PIR-Protein, release 57; SWISS-PROT, release 36; and PROSITE, release 15.0) with FastA (29) and BLAST (1) programs. Sequence alignments were performed by the CLUSTALW version 1.74 method (43) and with the Multalign program (15) and the SIM alignment tool (27).

Nucleotide sequence accession numbers. The complete ϕ Sfi11 and ϕ Sfi18 genome sequences were deposited in the GenBank database under accession no. AF158600 and AF158601.

RESULTS

Comparison of the Sfi11 and O1205 phage genomes. Phage Sfi11 (31) was chosen for complete-genome sequencing, since no complete genome sequence was available for lytic *pac*-site *S. thermophilus* phages. Phage Sfi11 has a 39,807-bp genome consisting of 52 open reading frames (ORFs) longer than 50 codons (Fig. 1). The bioinformatic analysis suggested a modular structure of the Sfi11 genome which closely resembled that of four other sequenced *S. thermophilus* phage genomes. Phages Sfi11 and O1205 differ in life style, host range, and geographical origin, while they belong to the same DNA-packaging group and have similar structural protein patterns and similar ecological origins (Table 1). Major differences between both genomes should therefore identify candidate genes involved in life style decision and host range determination. With the exception of the lysogeny module and two differences in the rightmost parts of their genomes, a one-to-one correspondence of the predicted gene map was observed (Fig. 1). This similarity extended to the nucleotide sequence level. The dot plot analysis (Fig. 2A) showed a nearly uninterrupted straight line over the putative DNA-packaging, morphogenesis, and lysis modules. Two gaps were found at the position of ORF 695 (Sfi11 numbering), encoding a likely tail fiber protein possibly involved in host range determination (31).

Over the genome region covering the lysogeny module, the dot plot indicated a combination of DNA rearrangements and deletion processes between the phages. Over the DNA replication modules, the dot plot again showed a straight line. This similarity continued with one interruption as far as the right ends of the two compared phage genomes. The interruption was an indel of an ORF 66 followed by a partial duplication of the origin of replication in Sfi11.

Over the DNA-packaging and morphogenesis module, the phages differed on average by only 10% at the nucleotide level, while over the lysis module the average nucleotide difference was about 20%. The DNA replication modules had nearly identical sequences, while the rightmost DNA segments showed on average a greater-than-20% nucleotide difference.

Comparison of the Sfi11 and Sfi19 phage genomes. Phages Sfi11 and Sfi19 differ in DNA-packaging mechanisms and structural protein patterns, host ranges, serotypes, and ecological origins. In contrast, the two phages have the same life style and geographical origin (Table 1). The phages showed similar genome organizations (Fig. 1). However, with one exception, no nucleotide sequence similarity was seen over the left 18 kb of the aligned genomes in dot plot analysis (Fig. 2B), and the deduced proteins lacked sequence similarity. This region covers the likely DNA-packaging and head and tail morphogenesis modules. Over ORF 695 (Sfi11 numbering), regions of nucleotide sequence similarity alternated with regions of non-

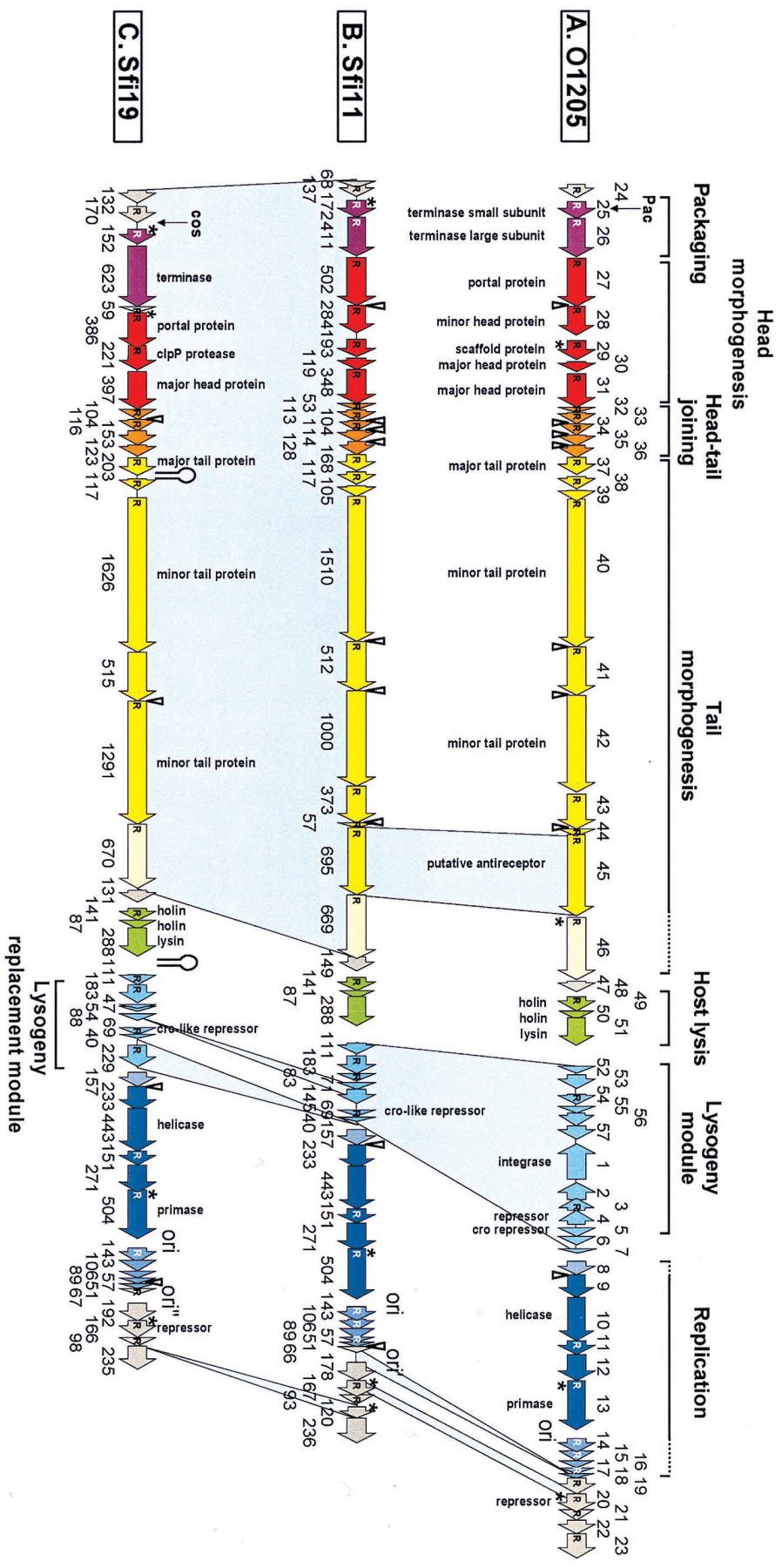


FIG. 1. Prediction of ORFs in the complete genomes of the temperate *pac*-site *S. thermophilus* phage O1205 (top), the virulent *pac*-site *S. thermophilus* phage Sfi11 (center), and the virulent *cos*-site *S. thermophilus* phage Sfi19 (bottom). To provide a better reference to previous publications, the ORFs from O1205 are numbered as in Stanley et al. (40), while the Sfi11 and Sfi19 ORFs are marked with their lengths in codon numbers. Probable gene functions identified by bioinformatic analysis or biological experiments are noted. The phage genomes were divided into functional units according to previous bioinformatic and comparative evolutionary analysis (32). Genes belonging to the same unit have the same color. Grey shading indicates lack of information about the function of the ORF. ORFs preceded by a potential ribosome binding site (23) are marked with an R inside the arrow. ORFs starting with an unconventional initiation codon are indicated with an asterisk, and possible rho-independent terminators are indicated with a hairpin. The overlap of start and stop codons is indicated with a triangle. Areas of blue shading connect regions of major sequence difference between the compared phage genomes. The genome maps were oriented to maximize alignments; the *cos* sites and *pac* sites are thus not at the very ends of the maps.

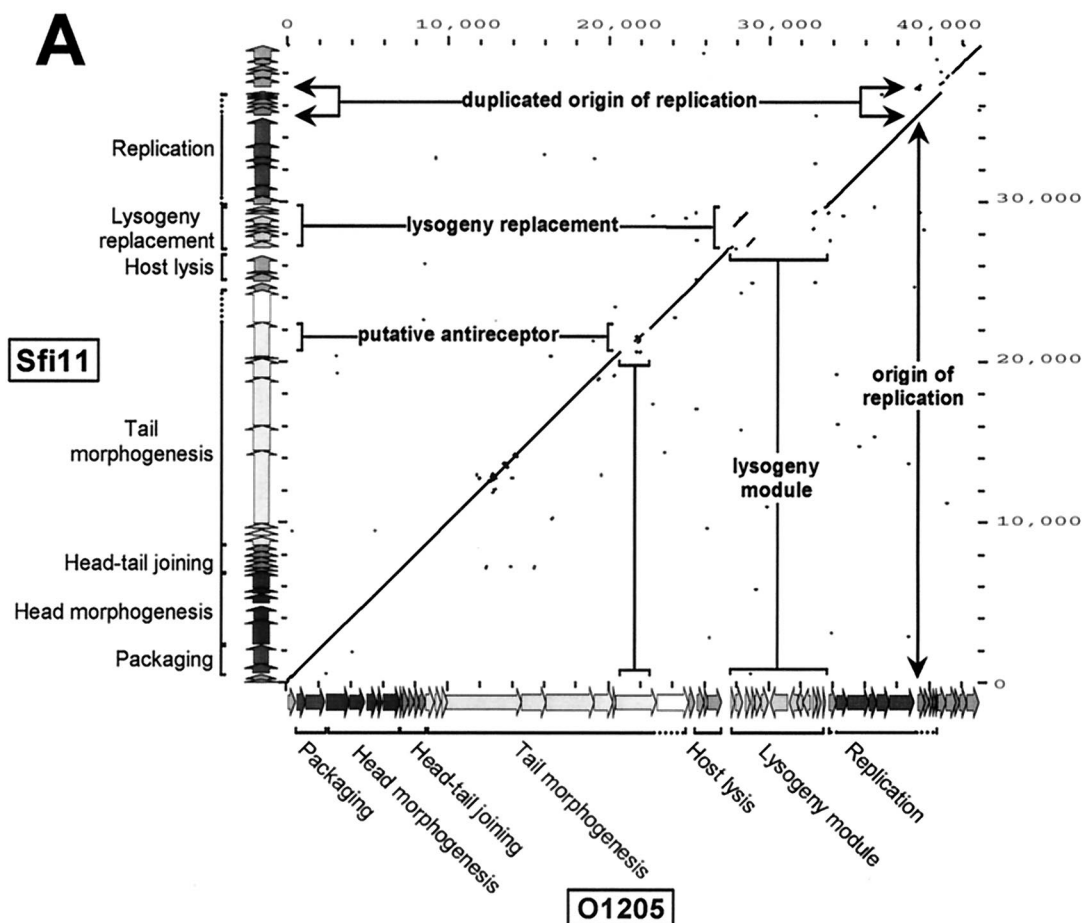


FIG. 2. Dot plot analysis. (A) Dot plot calculated for the DNA genome sequences of the virulent *pac*-site phage Sfi11 (y axis) and the temperate *pac*-site phage O1205 (x axis). The comparison window was 50 bp, and the stringency was 30 bp. To provide an easier orientation, the color-coded gene maps of Sfi11 and O1205 are shown at the y axis and the x axis, respectively, together with the map position in base pairs as defined in Fig. 1. Regions with sequence differences are marked and annotated. (B) Dot plot calculated for the DNA genome sequences of the virulent *pac*-site phage Sfi11 (y axis) and the virulent *cos*-site phage Sfi19 (x axis). (C) Dot plot calculated for the DNA genome sequences of the virulent *cos*-site phage Sfi19 (y axis) and the virulent *cos*-site phage DT1 (x axis).

similarity, followed by a nearly uninterrupted dot plot line. Only three insertion-deletion events were observed over the right halves of the two genomes. Two indels were localized in the lysogeny replacement module, and one was localized near the right end of the genome.

Comparison of the Sfi19 and DT1 phage genomes. Phages Sfi19 and DT1 differ in host range and geographical and ecological origin. The two phages share life style, DNA-packaging mechanism, and polypeptide pattern (Table 1). Major differences between the genomes should therefore identify candidate genes for host range determinants and polymorphisms that became established in geographically and ecologically separated phage lineages. A dot plot analysis revealed a straight line over the structural gene cluster that was interrupted by substitutions and by insertions and deletions of gene segments in the three largest ORFs from Sfi19 (ORFs 1626, 1291, and 670 [Fig. 2C]). All three genes could therefore play a role in host range determination. The lysis gene from DT1 was interrupted by an intron (20). The lysogeny replacement module from DT1 is the result of a distinct but related DNA rearrangement and deletion process. Multiple small DNA substitutions and indels punctuated the dot plot over the right ends of the genomes of both phages, possibly indicating genome polymorphisms.

Comparison of the Sfi19 and Sfi18 phage genomes. Phages Sfi18 and Sfi19 are related with respect to DNA packaging and structural gene pattern, geographical origin, host range, and life style (Table 1). They differ, however, in their behavior toward prophage Sfi21. Phage Sfi18 was unable to multiply on a cell containing the chromosomally integrated Sfi21 prophage, the cloned ORF 203, or the origin of replication from Sfi21 on a high-copy-number plasmid, while Sfi19 multiplication was unabated (12, 19). Major differences between the phages could thus identify adaptations of Sfi19 to escape from prophage control. PCR with Sfi19 primers revealed identical products for Sfi19 and Sfi18 over the structural gene cluster, while one difference was seen over the functional gene cluster (data not shown). We limited the sequencing to the right half of the genome from Sfi18, since we judged it unlikely that modifications in the structural gene cluster of Sfi19 are responsible for the escape from prophage control. Excluding the indel in ORF 88 from the lysogeny replacement region (see Fig. 4) and a 40-bp DNA substitution within the second origin, the functional gene clusters from Sfi19 and Sfi18 differed by only 145 base pair changes resulting in 34 amino acid changes. The amino acid changes were unevenly distributed: ORF 111 gene product alone showed 13 amino acid changes, and the ORF 166 gene product demonstrated a cluster of 5 amino acid

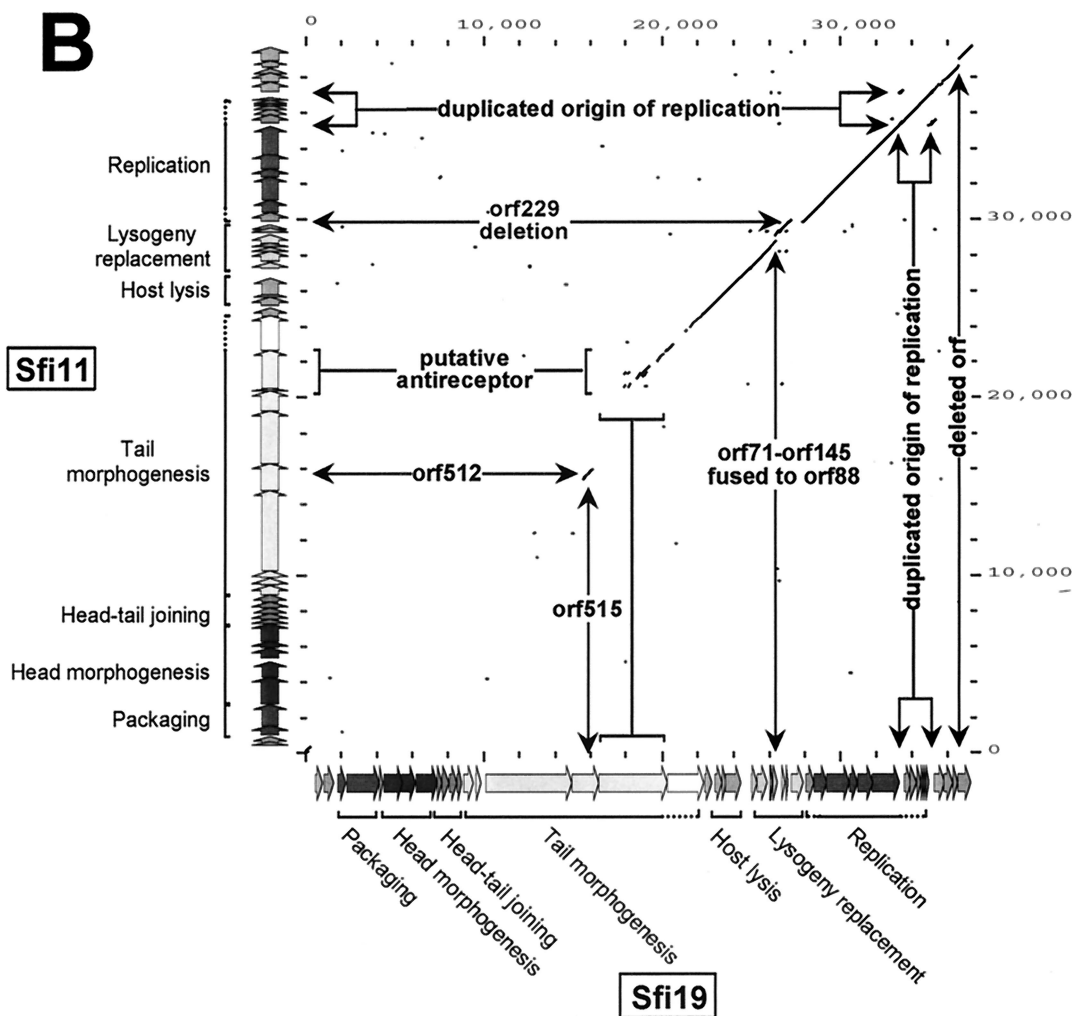


FIG. 2—Continued.

changes. A third of all base pair changes were in two noncoding regions resembling origins of replication. When a group of 20 phages with known susceptibility to prophage control were investigated by PCR, neither the indel in ORF 88 nor the substitution within the second putative origin correlated with escape from prophage Sfi21 control (data not shown), while the sequence of the origin correlated with this phenotype (19).

Recombination hot spots. After the pairwise alignments of two phage genomes, we aligned all available *S. thermophilus* phage genomes. This multiple alignment revealed a subdivision into conserved and variable regions. Conserved regions were represented by the DNA packaging, the head and tail gene region (two alleles, excluding the tail fiber genes), the lysis cassette (1 allele; DT1 showed a gene substitution for the first holin gene), and the DNA replication region (1 allele). Three variable regions were identified: the likely tail fiber genes, the lysogeny-replacement region, and the right end of the genome. All three regions demonstrate the importance of recombination processes for the diversification of *S. thermophilus* phage genomes.

The likely tail fiber genes. Multiple alignment of the largest predicted protein from the three *cos*-site phages Sfi19, Sfi21, and DT1 revealed a very conserved 500-amino-acid (aa) N-terminal segment followed by a variable region and then a

highly conserved central region of 75 aa that was flanked by a perfect 12-aa repeat (YKHNKKFKKFVD). After a second variable region, a third conserved region was observed. Weak similarity was detected with the protein encoded by the topologically corresponding ORF 1510 in Sfi11 (23% over 500 aa; $P = 10^{-10}$).

An even more complicated pattern was seen for the second-largest proteins. Over the N-terminal 900 aa, the corresponding proteins from the two *cos*-site phages Sfi19 and DT1 were nearly identical, except for an indel in DT1 which started and ended in a collagen-like repeat. The borders corresponded relatively precisely to a spontaneous deletion in Sfi21 (Fig. 3). Sfi21 differed from Sfi19 by a few amino acid replacements and an indel of a small collagen-like repeat. DT1 phage showed a second indel flanked at both sides by collagen-like repeats (Fig. 3), suggesting the collagen-like repeats were recombination hot spots. This conserved region was followed by a highly variable region between Sfi19 and DT1 and ended in a further highly conserved region. Notably, at a topologically corresponding genome position, *pac*-site phages encoded a protein that shared the highly conserved collagen-like repeat region with *cos*-site phages. This repeat region was followed by a region which differed not only between *cos*- and *pac*-site phages but also between the two *pac*-site phages. The variable

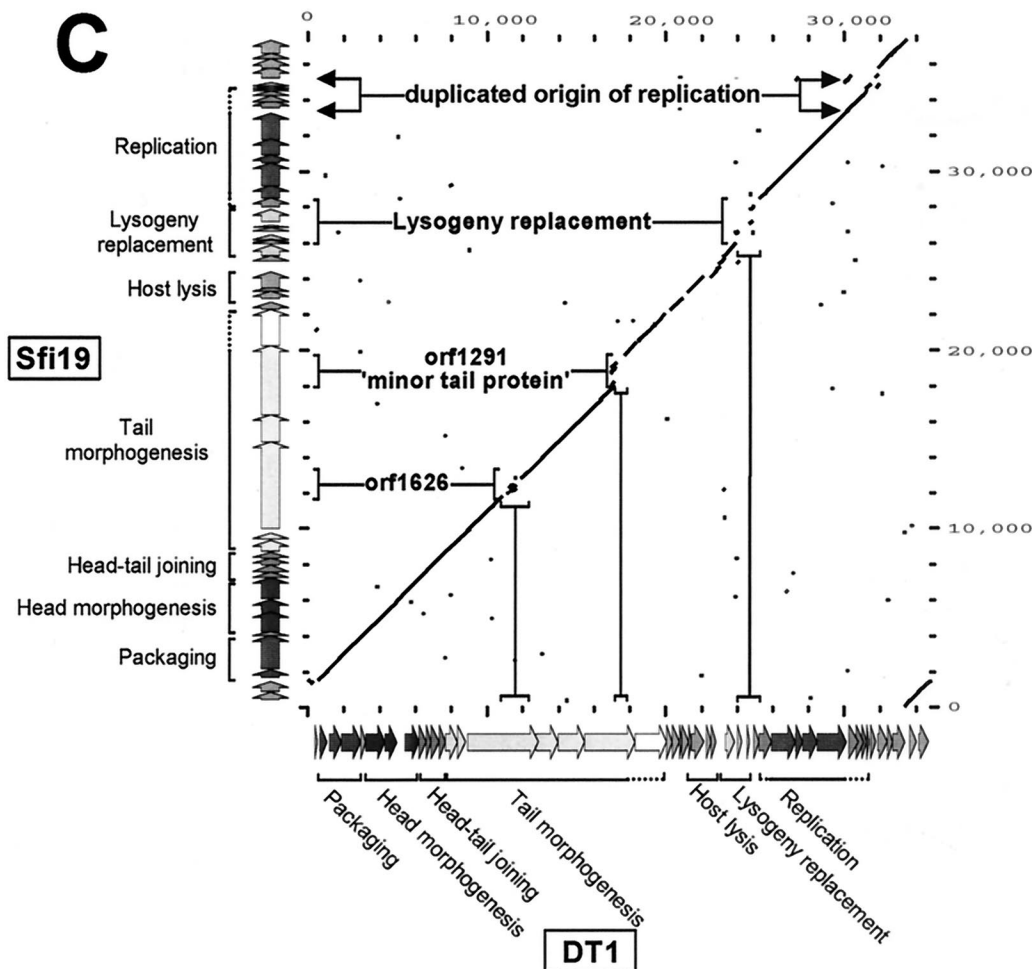


FIG. 2—Continued.

region was followed by a C-terminal domain that was highly conserved between the corresponding proteins from both *pac*-site phages and moderately conserved between the proteins from *pac*- and *cos*-site phages (Fig. 3).

The lysogeny replacement region. Temperate and lytic *S. thermophilus* phages showed the same overall genome organization. The lysogeny module was flanked on one side by the lysis module and on the other side by the DNA replication module. All lytic phages showed a replacement module at the position of the lysogeny module. A multiple alignment of the gene maps demonstrated that the replacement modules were derived from lysogeny modules by a combination of insertion-deletion and DNA rearrangement processes (Fig. 4). Interestingly, the three lytic phage isolates from France showed a replacement module that differed by only three insertion-deletion events. One event was an in-frame fusion of ORF 71 and ORF 145 from Sfi11 and Sfi18 to yield ORF 88 in phage Sfi19 (Fig. 4). This event is likely the consequence of DNA polymerase slippage, since the putative deletion site was precisely flanked by a heptanucleotide repeat, GATGATT, and only one repeat from phage Sfi19 was retained in ORF 88 (Fig. 4). A second event comprised the entire ORF 229 in Sfi18 and Sfi19 (Fig. 4). This ORF is flanked by a perfect 68-bp repeat situated 16 bp upstream and 2 bp downstream, respectively, of the start and stop codons of ORF 229. Sfi11 showed only one 68-bp sequence at the corresponding position, again suggesting DNA

polymerase slippage or homologous recombination as the cause of this polymorphism. Over ORF 111 to ORF 71, the temperate phage TPJ-34 differed from the lytic phages Sfi11, Sfi18, and Sfi19 by a 1-bp change resulting in a premature stop codon in TPJ-34 (Fig. 4) and a 1-bp insertion, resulting in different ORF predictions for Sfi18 and Sfi19 phages compared to phage Sfi11 (Fig. 4).

Notably, the Canadian virulent phage isolate DT1 showed a distinct DNA rearrangement event over this region (Fig. 4), demonstrating that the majority of the ORFs from the region are nonessential genes, as already indicated by the high number of polymorphisms in the French isolates. Only one gene is conserved among all four virulent phages (ORF 183), and all possess a *cro*-like gene. However, the DT1 phage shared a highly related *cro*-like gene with TPJ-34, while the French lytic phages shared a *cro*-like gene that was distantly related to O1205. Previously, we had attributed ORF 157 to the highly conserved DNA replication module from *S. thermophilus* phages (16). DT1 showed a distinct ORF 104 at this position.

The right end of the genome. Beyond the lysogeny or the replacement module, all six investigated *S. thermophilus* phages showed a common DNA segment which covered the putative DNA replication module and extended until ORF 236, three ORFs upstream of the respective *cos* or *pac* site. The left part of this common genome segment was highly conserved among the investigated phages, while the right part was punc-

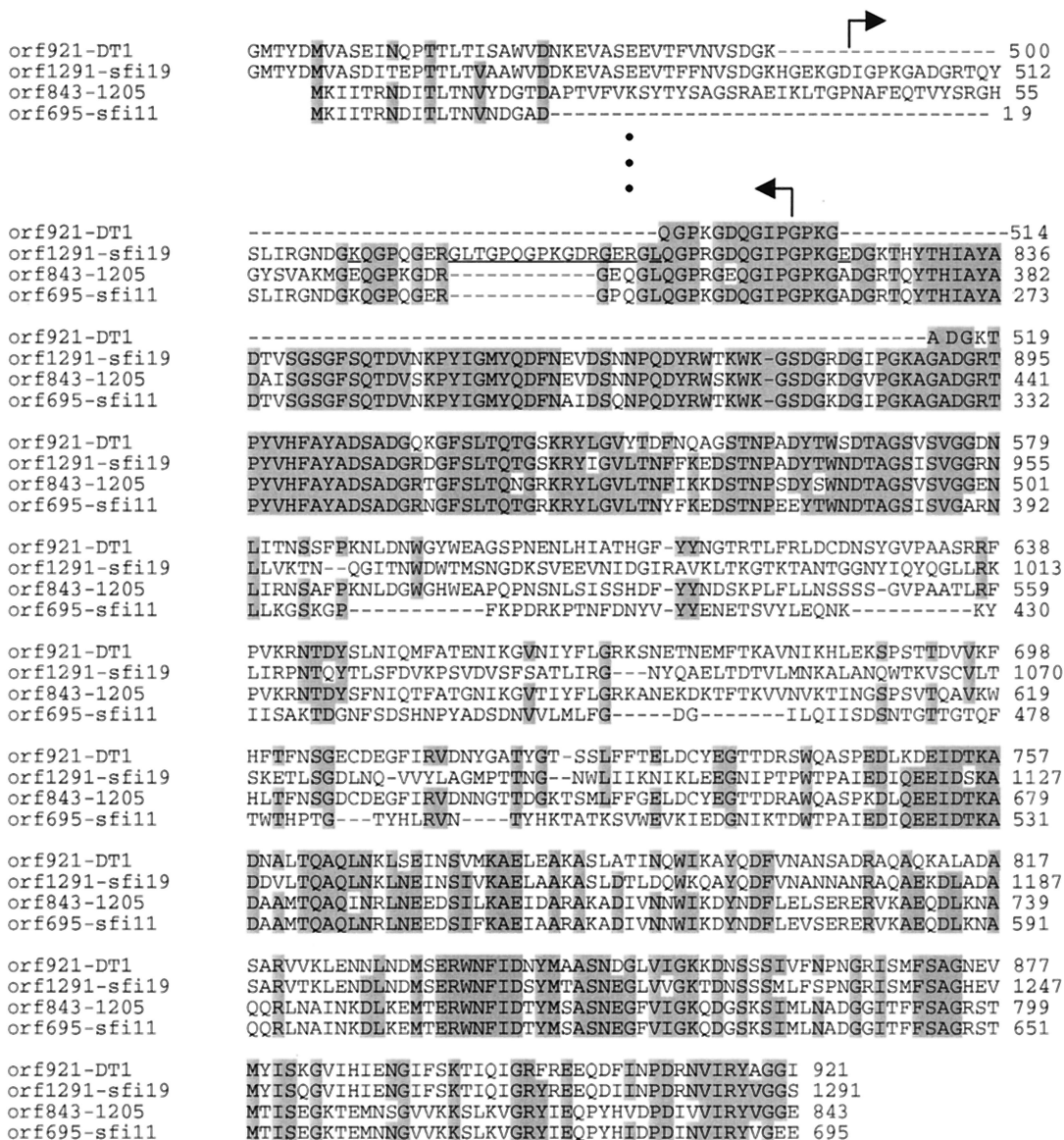


FIG. 3. Multiple alignment of a putative tail fiber protein from *S. thermophilus* phages showing distinct host ranges. The proteins are identified with their corresponding codon lengths. Gaps (-) were introduced for maximal alignment. Amino acid positions that differ in the corresponding proteins from phages Sfi19 and Sfi21 are underlined, and the location of a spontaneous deletion in the Sfi21 protein is marked by bent arrows. Amino acid positions which are identical in at least three proteins are shaded.

tuated by numerous insertions and deletions of entire genes (e.g., ORF 57 and ORF 146 in DT1) and substitutions of genes (e.g., ORF 152 in Sfi21 and ORF 146 in O1205) or a segment of a gene (e.g., ORF 156 in O1205) (Fig. 5). The region downstream of ORF 51 is a recombination hot spot. The three lytic phages isolated in France showed a replacement of ORF 61 and ORF 130 (Sfi21) at this position by a distinct ORF 67 followed by a duplicated origin of phage replication. The lytic Canadian isolate DT1 showed a distinct duplication at this position, while O1205 lacked supplementary DNA between ORF 50 and ORF 156 (Fig. 5).

DISCUSSION

Comparative sequence analysis revealed that genetic diversity is created in *S. thermophilus* phages by two distinct pro-

cesses. One process of obvious importance is genetic recombination. Recombination is apparently the basis of differences in life cycle, host range, DNA packaging, and structural protein pattern. The second process creating diversity is the accumulation of point mutations. The degree of sequence diversity differed substantially over the various phage genome segments. Practically no sequence variability was found over the replication module, while up to 30% base pair differences were detected over the morphogenesis genes. Interestingly, morphogenesis genes from *Lactococcus lactis* phages showed up to 50% base pair identity with phage Sfi21 (18), raising the possibility that the accumulation of point mutations is an important driving force of phage evolution. In comparison with phage Sfi18, relatively few base pair changes could free the lytic *S. thermophilus* phage Sfi19 from superinfection exclusion

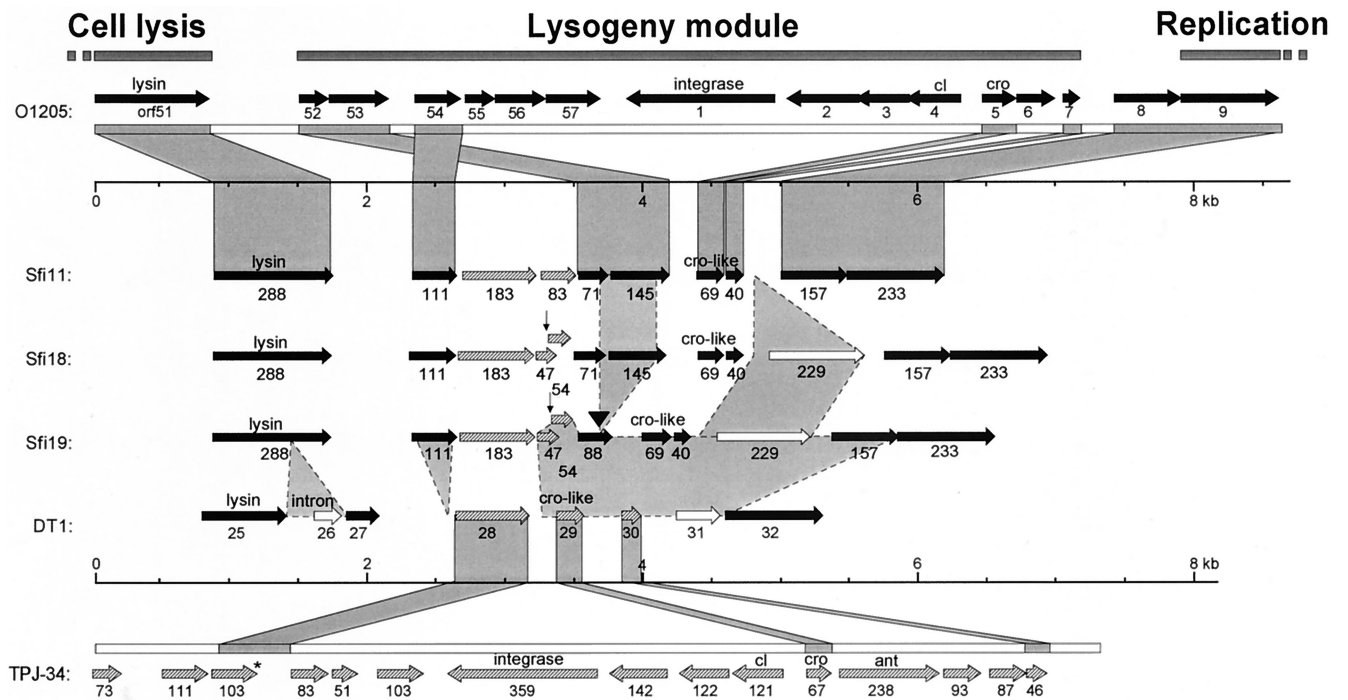


FIG. 4. Comparison of the lysogeny replacement module from the virulent *pac*-site phage Sfi11 with that of the virulent *cos*-site phages Sfi19, Sfi18, and DT1. The predicted ORFs are annotated with their codon lengths. The 1-bp indels are indicated by vertical arrows, and the position of a second indel is marked by a solid arrowhead. The third indel corresponds to ORF 229 flanked by 68-bp direct repeats (open triangle). For comparison, the genome maps of the lysogeny modules from the *pac*-site *S. thermophilus* phages O1205 (top) and TPJ-34 (bottom) are shown. Regions of sequence similarity between the phages O1205 and Sfi11 on one side and DT1 and TPJ-34 on the other side are connected by dark shading. Regions of sequence diversity between the four virulent phages are connected by light shading. The solid triangle indicates one repeat in ORF 88 (Sfi18).

by the Sfi21 prophage. The analysis of competitive interactions among RNA phages within the framework of game theory demonstrated that defection (selfishness) evolved, despite the greater fitness payoff that would result if all phage players were

to cooperate (46). Phage Sfi19 could thus be a selfish phage player, which might explain the relative rarity of lysogeny in *S. thermophilus*.

Genetic recombination apparently plays a role in *S. ther-*

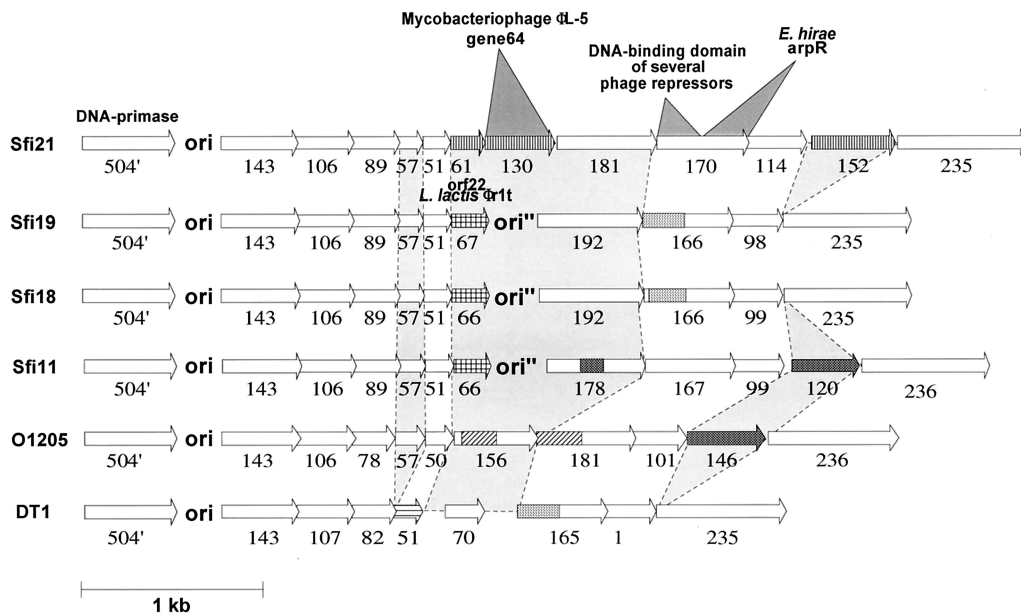


FIG. 5. Alignment of the gene maps from the right ends of the genomes of the indicated *S. thermophilus* phages. The predicted ORFs are annotated with their codon lengths. Insertions and deletions are connected by light shading. Sequence diversity greater than 20% at the amino acid level is indicated by different shadings of the arrows representing the ORFs. *E. hirae*, *Enterococcus hirae*.

thermophilus phages comparable to that of the much-better-investigated (although less intensively sequenced) lambdoid coliphages (14). Lambdoid coliphages showed a similar division into a structural and a functional gene cluster. As in *S. thermophilus* phages, only two different structural gene clusters were identified in lambdoid phages. Apparently due to the multiple interactions of the structural proteins, modular exchanges are not allowed within the morphogenesis gene cluster. The similarity between the phage systems is not surprising, since the structural gene map from *S. thermophilus* phages could be aligned with that from lambdoid coliphages (18). In both phage systems, modular exchange reactions were limited to the putative tail fiber genes and the functional gene cluster. In *S. thermophilus* phages, the units of genetic exchange over the functional gene cluster are small and comprise single genes or even segments of genes. Similar observations were made for mycobacteriophages (21), and there is also increasing evidence that units of genetic exchange in lambdoid phages can be as small as segments of an individual gene (26, 38).

As seen previously from the comparison of temperate *S. thermophilus* phages (39), the lysogeny region is a recombination hot spot. All four virulent phages showed a lysogeny replacement module, which may be derived from temperate phages by two processes: the deletion of genes essential for the establishment of the lysogenic state (integrase and repressor genes) and the rearrangement of DNA regions flanking the lysogeny module in temperate *pac*-site *S. thermophilus* phages. A common denominator was the conservation of a *cro*-like repressor. Such a close similarity between virulent and temperate phages has also been described for phages from other gram-positive bacteria, such as *Lactobacillus* (37) and *Mycobacterium* (21).

The putative tail fiber genes are a further recombination hot spot. Multiple alignment of these proteins revealed a pattern of conservation and nonconservation reminiscent of the mechanism with which T-even phages create host range diversity in their adhesins (42). Further similarities with T-even phage adhesins were glycine-rich segments in the conserved protein regions: Tétart et al. (42) described oligoglycine stretches in T4, while Desiere et al. (17) reported collagen-like glycine-X-Y repeats in *S. thermophilus* phages. We speculate that these elements play a role in the recombinational reshuffling of the putative phage antireceptors. Similar collagen-like repeats were observed in *Lactococcus* phage BK5-T (3). During serial passage in the laboratory, BK5-T and Sfi21 phages showed spontaneous deletions which originated in the collagen-like repeats (4, 10). The conserved C terminus of the putative *S. thermophilus* phage antireceptor was also detected in a corresponding protein from the *S. pneumoniae* phage Dp-1 (17). Since this protein segment gave a strong coiled-coil prediction, it could represent an important structural determinant for the assembly of a fibrous protein.

The prominent role of recombination in the diversification of bacteriophage genomes has prevented the assessment of phylogenetic relationships among the different phage groups. In phage modules, such as the morphogenesis gene cluster, which do not allow exchange reactions, we have observed evidence for the role of point mutations in the evolution of temperate *Siphoviridae* from low-G+C-content gram-positive bacteria. With an increasing database of phage sequences, it might therefore become possible to delineate evolutionary relationships, at least for selected phage modules. The prominent role of recombination has also complicated all taxonomical approaches to phages. The International Committee for the Taxonomy of Viruses (ICTV) defined a virus species as "a polythetic class of viruses that constitutes a replicating lineage

and occupies a particular ecological niche" (35). According to this definition, all *S. thermophilus* phages investigated in this report should belong to the same phage species, since they have extensive DNA homology indicative of a replicative lineage and they infect a single bacterial species from a defined dairy environment. In our opinion, this definition has the advantage of biological plausibility. The ICTV definition of a virus genus is vague: "a virus genus is a group of species sharing certain common characters" (35). Maniloff and Ackermann (34) have delineated a taxonomy of bacterial viruses in which they have tentatively defined six viral genera in the family *Siphoviridae*. Two criteria were used to define the tailed-virus genera: properties related to DNA replication and packaging, and specific features, such as the ability to establish a temperate infection or bacterial host. According to this proposition it could be argued that *cos*-site and *pac*-site *S. thermophilus* phages represent distinct and new phage genera of *Siphoviridae* (if the criteria of Maniloff and Ackermann [34] are stretched somewhat, *cos*-site *S. thermophilus* phages could be grouped with the genus λ -like phages"). The attribution of *S. thermophilus* phages to two genera seems counterintuitive to us. According to the concept of Maniloff and Ackermann (34), many more new *Siphoviridae* genera will be created as more *Siphoviridae* from underinvestigated bacterial groups are characterized. We point out that less than 1% of the bacteria from the natural environment can be cultivated and that viruses are the most common biological agents, for example, in the sea (22). The ICTV rules state that it is not obligatory to use all levels of the taxonomical hierarchy (35). We propose to renounce the attribution of *Siphoviridae* genera for the moment until we have at our disposal a reasonable sequence database for bacterial viruses. In view of the current ease of sequence acquisition, a coordinated international effort to obtain whole-genome sequences for available phages from all major classes of bacteria is warranted.

ACKNOWLEDGMENTS

We thank the Swiss National Science Foundation for the financial support of S. Lucchini and F. Desiere in the framework of its Biotechnology Module (grant 5002-044545/1).

We thank S. Foley for critical reading of the manuscript.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
- Botstein, D. 1980. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354**:484-491.
- Boyce, J. D., B. E. Davidson, and A. J. Hillier. 1995. Sequence analysis of the *Lactococcus lactis* temperate bacteriophage BK5-T and demonstration that the phage DNA has cohesive ends. *Appl. Environ. Microbiol.* **61**:4089-4098.
- Boyce, J. D., B. E. Davidson, and A. J. Hillier. 1995. Spontaneous deletion mutants of the *Lactococcus lactis* temperate bacteriophage BK5-T and localization of the BK5-T *attP* site. *Appl. Environ. Microbiol.* **61**:4105-4109.
- Brüssow, H. 1999. Phages of *Streptococcus thermophilus*. In R. Webster and A. Granoff (ed.), *Encyclopedia of virology*, 2nd ed. Academic Press, London, England.
- Brüssow, H., and A. Bruttin. 1995. Characterization of a temperate *Streptococcus thermophilus* bacteriophage and its genetic relationship with lytic phages. *Virology* **212**:632-640.
- Brüssow, H., A. Bruttin, F. Desiere, S. Lucchini, and S. Foley. 1998. Molecular ecology and evolution of *Streptococcus thermophilus* bacteriophages—a review. *Virus Genes* **16**:95-109.
- Brüssow, H., M. Frémont, 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.
- Brüssow, H., A. Probst, M. Frémont, and J. Sidoti. 1994. Distinct *Streptococcus thermophilus* bacteriophages share an extremely conserved DNA fragment. *Virology* **200**:854-857.
- Bruttin, A., and H. Brüssow. 1996. Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. *Virology* **219**:96-104.

11. **Bruttin, A., F. Desiere, N. d'Amico, J. P. Guérin, J. Sidoti, B. Huni, S. Lucchini, and H. Brüßow.** 1997. Molecular ecology of *Streptococcus thermophilus* bacteriophage infections in a cheese factory. *Appl. Environ. Microbiol.* **63**:3144–3150.
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 phi Sfi21. *Virology* **233**:136–148.
13. **Bujarski, J. J., and P. Kaesberg.** 1986. Genetic recombination between RNA components of a multipartite plant virus. *Nature* **321**:528–531.
14. **Casjens, S., G. Hatful, and R. Hendrix.** 1992. Evolution of dsDNA tailed bacteriophage genomes. *Semin. Virol.* **3**:383–397.
15. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic. Acids. Res.* **16**:10881–10890.
16. **Desiere, F., S. Lucchini, A. Bruttin, M.-C. Zwahlen, and H. Brüßow.** 1997. A highly conserved DNA replication module from *Streptococcus thermophilus* phages is similar in sequence and topology to a module from *Lactococcus lactis* phages. *Virology* **234**:372–382.
17. **Desiere, F., S. Lucchini, and H. Brüßow.** 1998. Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. *Virology* **241**:345–356.
18. **Desiere, F., S. Lucchini, and H. Brüßow.** 1999. Evolutionary analysis of the DNA packaging and head morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage phi Sfi21. *Virology* **260**:244–253.
19. **Foley, S., S. Lucchini, M.-C. Zwahlen, and H. Brüßow.** 1998. A short non-coding viral DNA element showing characteristics of a replication origin confers bacteriophage resistance to *Streptococcus thermophilus*. *Virology* **250**:377–387.
20. **Foley, S., A. Bruttin, and H. Brüßow.** Widespread distribution of a group I intron and its three deletion derivatives in the *lys* gene of *Streptococcus thermophilus* bacteriophages. Submitted for publication.
21. **Ford, M. E., G. J. Sarkis, A. E. Belanger, R. Hendrix, and G. F. Hatful.** 1998. Genome structure of mycobacteriophage D29: implications for phage evolution. *J. Mol. Biol.* **279**:143–164.
22. **Fuhrman, J. A.** 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**:541–548.
23. **Guédon, G., F. Bourgoin, M. Pébay, Y. Roussel, C. Colmin, J. M. Simonet, and B. Decaris.** 1995. Characterization and distribution of two insertion sequences, IS1191 and iso-IS981, in *Streptococcus thermophilus*: does intergenic transfer of insertion sequences occur in lactic acid bacteria co-cultures? *Mol. Microbiol.* **16**:69–78.
24. **Hahn, C. S., S. Lustig, E. G. Strauss, and J. H. Strauss.** 1988. Western equine encephalitis virus is a recombinant virus. *Proc. Natl. Acad. Sci. USA* **85**:5997–6001.
25. **Hendrix, R., M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatful.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. USA* **96**:2192–2197.
26. **Highton, P. J., Y. Chang, and R. J. Myers.** 1990. Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. *Mol. Microbiol.* **4**:1329–1340.
27. **Huang, X., and W. Miller.** 1991. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* **12**:337–357.
28. **Li, K., Z. Chen, and P. Plegemann.** 1999. High-frequency homologous genetic recombination of an arterivirus, lactate dehydrogenase-elevating virus, in mice and evolution of neuropathogenic variants. *Virology* **258**:73–83.
29. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435–1441.
30. **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.
31. **Lucchini, S., F. Desiere, and H. Brüßow.** 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts. *Virology* **246**:63–73.
32. **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.
33. **Makino, S., J. G. Keck, S. T. Stohman, and M. M. C. Lai.** 1986. High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**:729–737.
34. **Maniloff, J., and H.-W. Ackermann.** 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. *Arch. Virol.* **143**:2051–2063.
35. **Mayo, M. A.** 1996. Recent revisions of the rules of virus classification and nomenclature. *Arch. Virol.* **141**:2479–2484.
36. **Mercenier, A.** 1990. Molecular genetics of *Streptococcus thermophilus*. *FEMS Microbiol. Rev.* **87**:61–78.
37. **Mikkonen, M., L. Dupont, T. Alattosava, and P. Ritzenthaler.** 1996. Defective site-specific integration elements are present in the genome of virulent bacteriophage LL-H of *Lactobacillus delbrueckii*. *Appl. Environ. Microbiol.* **62**:1847–1851.
38. **Moore, D. D., K. J. Denniston, and F. R. Blattner.** 1981. Sequence organization of the origins of DNA replication in lambdoid coliphages. *Gene* **14**:91–101.
39. **Neve, H., K. I. Zenz, F. Desiere, A. Koch, K. J. Heller, and H. Brüßow.** 1998. Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: implications for the modular theory of phage evolution. *Virology* **241**:61–72.
40. **Stanley, E., G. F. Fitzgerald, M. C. LeMarrec, B. Fayard, and D. van Sinderen.** 1997. Sequence analysis and characterization of phi O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology* **143**:3417–3429.
41. **Strauss, E. G., J. H. Strauss, and A. J. Levine.** 1996. Virus evolution, p. 153–171. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), *Virology*, 3rd ed. Raven Press, New York, N.Y.
42. **Tétart, F., C. Desplats, and H. Krisch.** 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. *J. Mol. Biol.* **282**:543–556.
43. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
44. **Tolskaya, E. A., L. A. Romanov, M. S. Kolesnikova, and V. I. Agol.** 1983. Intertypic recombination in poliovirus: genetic and biochemical studies. *Virology* **124**:121–132.
45. **Tremblay, D., and S. Moineau.** 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. *Virology* **255**:63–76.
46. **Turner, P. E., and L. Chao.** 1999. Prisoner's dilemma in an RNA virus. *Nature* **398**:441–443.
47. **Weaver, S. C., W. Kang, Y. Shirako, T. Rümepf, E. G. Strauss, and J. H. Strauss.** 1997. Recombinational history and molecular evolution of Western equine encephalomyelitis complex alphavirus. *J. Virol.* **71**:613–623.
48. **Webster, R. G., W. G. Laver, G. M. Air, and G. C. Schild.** 1982. Molecular mechanisms of variation in influenza viruses. *Nature* **296**:115–121.