VIROLOGY 246, 63–73 (1998)
ARTICLE NO. VY989190

The Structural Gene Module in *Streptococcus thermophilus* Bacteriophage φSfi11 Shows a Hierarchy of Relatedness to Siphoviridae from a Wide Range of Bacterial Hosts

Sacha Lucchini, Frank Desiere, and Harald Brüssow[1]

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

Received March 5, 1998; accepted April 15, 1998

The structural gene cluster and the lysis module from lytic group II *Streptococcus thermophilus* bacteriophage φSfi11 was compared to the corresponding region from other Siphoviridae. The analysis revealed a hierarchy of relatedness. φSfi11 differed from the temperate *S. thermophilus* bacteriophage φO1205 by about 10% at the nucleotide level. The majority of the changes were point mutations, mainly at the third base position. Only a single gene (orf 695) differed substantially between the two phages. Over the putative minor tail and lysis genes, φSfi11 and the lytic group I *S. thermophilus* φSfi19 shared regions with variable degrees of similarity. Orf 1291 from φSfi19 was replaced by four genes in φSfi11, two of which (orf 1000 and orf 695) showed a complicated pattern of similarity and nonsimilarity compared with φSfi19. The predicted orf 695 gp resembles the receptor-recognizing protein of T-even coliphages in its organization, but not its sequence. No sequence similarity was detected between φSfi11 and φSfi19 in the region covering the major head and tail genes. Comparison of the structural gene map of φSfi11 with that of Siphoviridae from gram-positive and -negative bacterial hosts revealed a common genomic organization. Sequence similarity was only found between φSfi11 and Siphoviridae from gram-positive hosts and correlated with the evolutionary distance between the bacterial hosts. Our data are compatible with the hypothesis that the structural gene operon from Siphoviridae of the low G+C group of gram-positive bacteria is derived from a common ancestor.

INTRODUCTION

One of the seminal conceptual advances in contemporary biology was the populational definition of the biological species. In this concept, a species is defined as a group of individuals belonging to a closed interbreeding population whose genes can be considered a common pool. The essence of speciation lies in the reproductive isolation of members of a given species. The notion of a gene pool makes sense only if the members of that pool recombine at a significant rate under natural conditions. For the gene pool to be closed, gene flow between members of the species and other sources should ideally be 0 or at least be small compared to intraspecific recombination (Campbell, 1988). The species concept is less well rooted in prokaryotes. There are fundamental problems with the species concept in bacteria since accessory DNA elements like plasmids, transposons, and phages are disseminated among bacteria that are very distantly related taxonomically (Ochman and Lawrence, 1996). In addition, a bacterial species like *Escherichia coli* covers strains that might vary by about 5% at the sequence level (Milkman, 1996; Whittam, 1996). Biologically defined species cover much less sequence diversity, e.g., it is estimated that different individuals of *Homo sapiens* might differ by 0.05% at the DNA level. In fact, the range of diversity covered by a single bacterial species is greater than that between different biological genera, e.g., man and chimpanzee differ by about 2% at the DNA level. Nevertheless, the chromosome of a given bacterium constitutes a coadapted complex that apparently evolved separately despite the existence of pathways for gene transfer with other bacteria. Therefore, in practical terms the evolutionary definition of a bacterial species is commonly accepted. The species concept in virology is a debated issue (Murphy, 1996; Ackermann and DuBow, 1987). The present universal system of virus taxonomy defined arbitrarily hierachical levels of order (-virales), family (-viridae), genus, species, and strain. The taxonomy of animal viruses is relatively developed (Murphy, 1996), as demonstrated by the following example : order Mononegavirales, family Paramyxoviridae, genus *Morbillivirus,* species measles virus, strain Schwarz. Most families of viruses have distinct morphology, genome structure, and strategies of replication. The virus family is being recognized as a taxon uniting viruses with a common, even if distant, phylogeny. Consequently the relatedness, for example, between Paramyxoviridae was studied by phylogenetic tree analyses based on nucleic acid or amino acid sequences (Griffin and Bellini, 1996). In contrast, the relationship between different families within the Mononegavirales (e.g., Rhabdoviridae and Paramyxoviridae) is seen only at the genomic organization level and not any longer at the sequence level (Strauss *et al.,* 1996). The

[1] To whom correspondence and reprint requests should be addressed. E-mail: Harald.Bruessow@CHLSNR.NESTRD.CH.

0042-6822/98 $25.00

molecular taxonomy of bacterial viruses is much less developed. The International Committee on Taxonomy of Viruses classified bacterial viruses on the basis of two criteria : at the first level the genome type (DNA or RNA, double-stranded or single-stranded) and then at the level of phage morphology. In the group of bacteriophages with double-stranded DNA genomes eight morphological types are currently distinguished (Fauquet, 1997). Numerically by far the most prominent are three groups: phages with contractile tails (Myoviridae, prototype:T4 phage), long and noncontractile tails (Siphoviridae, prototype: phage lambda), and short tails (Podoviridae, prototype: T7 phage). Approximately half of the about 3000 known bacteriophages belong to the group Siphoviridae (Ackermann and DuBow, 1987). Much less is known about the phylogenetic relationships of bacterial viruses than those of animal viruses. Comparative data are mainly available for phages belonging to restricted phage groups like lambdoid phages (Botstein, 1980; Campbell and Botstein, 1983) or T-even phages (Monod *et al.,* 1997). In fact, many bacterial virologists see the populational definition of a virus species as clearly inapplicable to bacteriophages. Some virologists imagine that homologous recombination within a phage population is so frequent that the basic units of selection are not the individual phage particles but rather the segments of their genomes that are interchangeable by recombination (Casjens *et al.,* 1992). Even more radical, others see illegitimate recombination as such a dominant force that phage genomes can best be regarded as mosaics of genes from various nonphage sources (Campbell, 1988).

To address these questions, we have started comparative phage genome sequencing projects in our laboratory. Our laboratory is interested in bacteriophages of *Streptococcus thermophilus,* a gram-positive lactic acid bacterium used extensively in industrial milk fermentation (Mercenier, 1990). These studies were motivated by the industrial aim of developing phage-resistant bacterial starter cultures. Therefore, we have to understand the natural variability of *S. thermophilus* phages. Previously we have classified these phages by different taxonomic criteria leading to the definition of different lytic groups (Brüssow *et al.,* 1994a; Brüssow and Bruttin, 1995). Two members of lytic group I were analyzed in sequencing projects (Brüssow *et al.,* 1994b; Bruttin *et al.,* 1997b; Desiere *et al.,* 1997, 1998). Here we describe the genome sequence of a representative lytic group II phage and compare it, first, with more or less related Siphoviridae from the same host species *S. thermophilus* (Le Marrec *et al.,* 1997; Stanley *et al.,* 1997), then with Siphoviridae from a related host, *Lactococcus lactis* (Chandry *et al.,* 1997; van Sinderen *et al.,* 1996; Johnsen *et al.,* 1996; Boyce *et al.,* 1995a), and finally with Siphoviridae from hosts showing decreasing phylogenetic relatedness (*Bacillus subtilis, Mycobacteria, Streptomyces, Escherichia coli*) with *S. thermophilus* (Becker *et al.,* 1997; Anné *et al.,* 1990; Hatful and Sarkis, 1993). This approach revealed a common genome organization of these phages over the morphogenesis module. The relatedness extended to the sequence level for bacteriophages from the low GC group of gram-positive bacteria. The implications of these findings for the understanding of Siphoviridae evolution are discussed in the framework of the species concept.

RESULTS

ϕSfi11 and ϕSfi19 are clearly distinct phage types

We tested the host range of our prototype lytic group I phage, ϕSfi19, and our prototype lytic group II phage, ϕSfi11 (Brüssow *et al.,* 1994a), on a total of 226 distinct *S. thermophilus* strains. Not a single strain was lysed by both phages. Furthermore, 130 *S. thermophilus* phages from another phage collection (H. Neve, Kiel/Germany) were tested on our lytic group I and II indicator cells. Not a single phage infected both indicator cells. Apparently, lytic group I and II phages differ in a fundamental way, resulting in two nonoverlapping infection patterns.

Further results confirmed the difference between ϕSfi11 and ϕSfi19. First, the two phages differed in the neutralization of their infectivity by hyperimmune sera defining two clearly separated serotypes (Brüssow *et al.,* 1994a). Second, the phages differed in their polypeptide composition when CsCl gradient-purified phage particles were compared by SDS–polyacrylamide gel electrophoresis (data not shown). Third, the two phages differed in tail morphology: tails from ϕSfi11 were thinner than tails from ϕSfi19 and the striation of the tail was less evident in ϕSfi11. In addition, ample numbers of tail fibers were detected in ϕSfi11 (Fig. 1B), while we never observed tail fibers in preparations of ϕSfi19 (Fig. 1A). The latter observation should be interpreted with caution since *S. thermophilus* phages are unstable in CsCl gradients, as exemplified by the isolation of ϕSfi11 heads lacking tails (Fig. 1C) or loss of phage particles during purification (Fayard *et al.,* 1993). Interestingly, the tailless ϕSfi11 particles lacked, in addition to some minor proteins, the 27-kDa major protein (data not shown), suggesting that this protein is likely to be a tail protein.

DNA sequence of ϕSfi11

A 24-kb DNA segment from ϕSfi11 was sequenced, corresponding to about 60 % of the genome. When only ATG start codons were accepted, 33 open reading frames (orf) longer than 60 aa were detected. Twenty-three orfs remained when those located within or opposite to orfs that showed similarity to entries from the database were subtracted (Fig. 2). With one exception, all orfs were located on the same strand. The overall genetic structure of this region was very dense. The start and stop codons of three groups of genes overlapped, indicating potential translational coupling (Fig. 2). Translational coupling (Draper, 1996) is a posttranscription mechanism that helps ensure balanced production of polypeptides that function as part of a multicomponent complex.

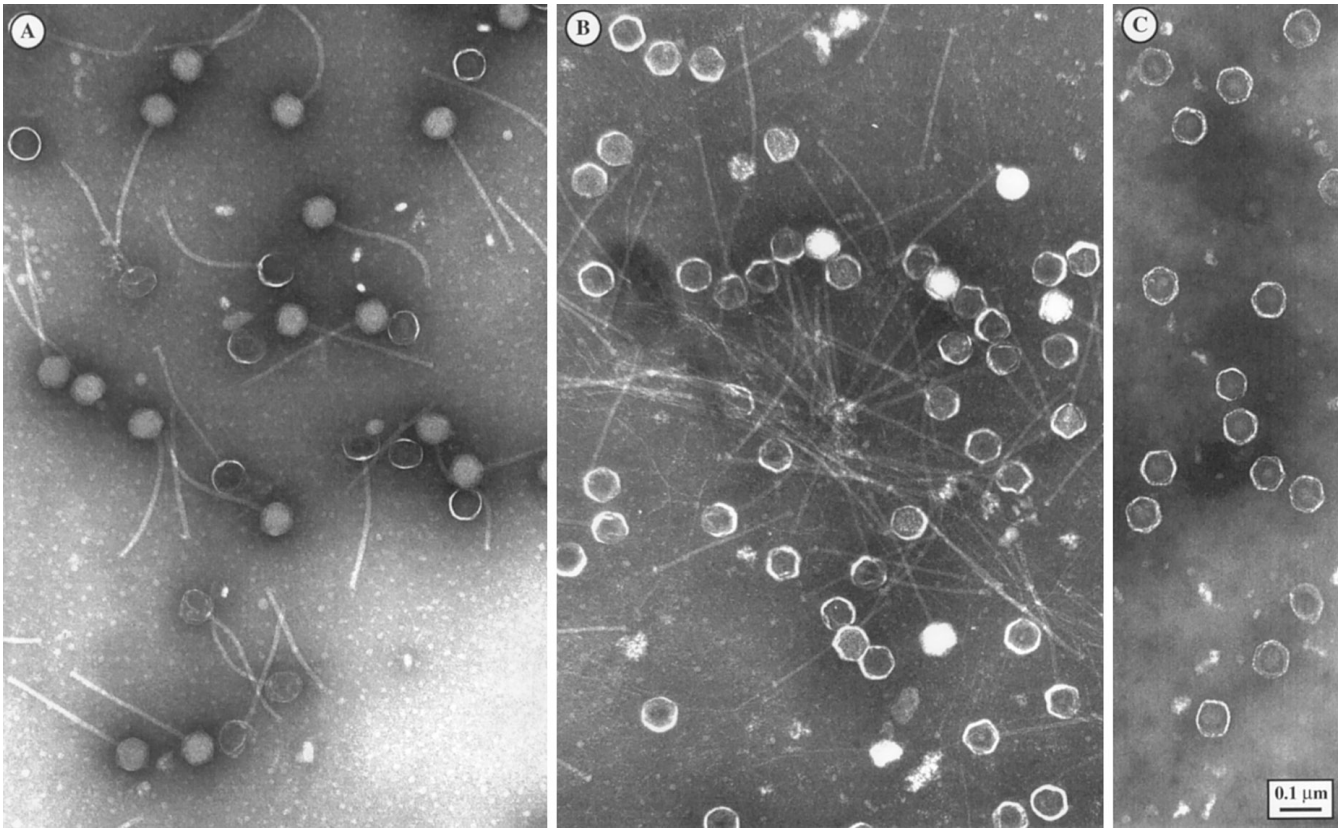

FIG. 1 Electron micrographs of *S. thermophilus* bacteriophages φSfi19 (A), φSfi11 (B), and tailless φSfi11 (C). Bar, 0.1 μm.

If entries from *S. thermophilus* phages were excluded, 20 of the 23 orfs showed similarity to predicted proteins from the database. Without exception the similarities were to proteins from bacteriophages. Most similarities were to proteins from phages infecting taxonomically related bacteria such as *Streptococcus pneumoniae, L. lactis,* and *B. subtilis.* However, two proteins showed similarities to bacteriophages from distantly related gram-positive bacteria, *Streptomyces venezuelae* and *Mycobacterium tuberculosis* (Fig. 2).

Several adjacent φSfi11 genes showed similarity to a gene cluster from the *L. lactis* phage TP 901-1 (Table 1). Interestingly, the similarity in the size and the topological organization of these two phage gene clusters extended to genes showing no sequence similarity, e.g., *mhp* from φTP 901-1 coding for a 348-aa-long protein and orf 348 from φSfi11 (see below, Conservation of genome organization).

Upstream of the lysis cassette, bioinformatic analysis revealed genes coding for likely phage structural proteins. These genes are commonly clustered in phage genomes. The molecular weights of the six minor structural proteins estimated from PAGE (Fig. 3) showed a reasonable match to proteins predicted for all but one of the larger orfs from the φSfi11 DNA segment analyzed (Fig. 2).

Comparison with φO1205

S. thermophilus φO1205 (Stanley *et al.,* 1997) and φSfi11 differ in lifestyle (φSfi11 is a virulent phage, while φO1205 is temperate) and in host range (φO1205 was unable to multiply on any of our *S. thermophilus* strains including a number of lytic group II strains). However, except for one gene (orf 695, see below, Comparison with φSfi19) φSfi11 showed very similar genetic organization to φO1205. The genetic similarity between the two phages extended to the nucleotide level (Table 2). Overall, an average base pair change rate of about 10% was calculated. The majority of the changes were point mutations, mainly at the third base position, but an 8.1% average aa change rate was still observed for the predicted gene products. Over the putative lysis cassette, φSfi11 was more closely related to lytic group I phage φSfi19 (Desiere *et al.,* 1998) than to φO1205 (Table 2, Fig. 4). An interesting case is the unattributed gene preceding the holin genes, where the proteins predicted for the two temperate phages differed from those predicted for the two virulent phages by a 14-aa internal deletion (see Fig. 5 in Desiere *et al.,* 1998).

Comparison with φSfi19

Figure 4 shows an alignment of the partial φSfi11 gene map with the corresponding region of lytic group I *S. thermophilus* phage φSfi19 (Desiere *et al.,* 1998). The two phages showed comparable genetic maps if one postulates a fragmentation of orf 1291 from φSfi19 into separate orfs in φSfi11 (orf 1000, 373, 57, and 695), which is supported by the sequence similarity data

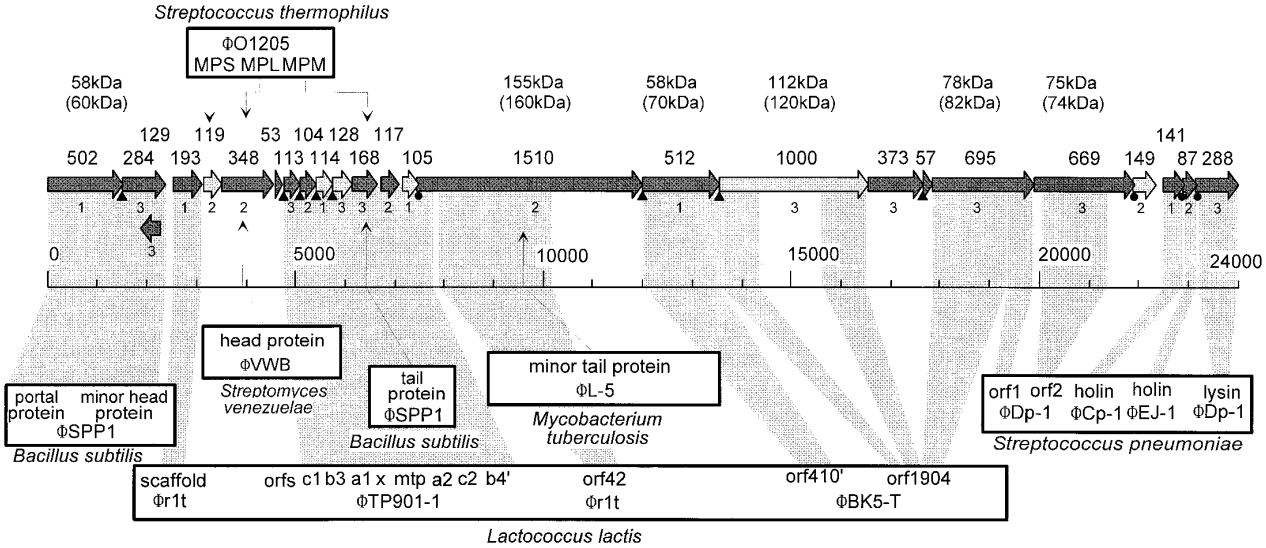

FIG. 2. Prediction of open reading frames in the 24-kb fragment of ϕSfi11. The orfs were marked above the arrows with their length in aa; the reading frame is indicated below the arrow. Filled arrows indicate orfs preceded by a standard ribosomal binding site. Overlaps of orfs are marked by a filled circle, and overlaps of start and stop codons are marked by filled triangles. The ruler provides the nucleotide scale. Similarities to proteins from the database are indicated by shadings. Where the shading does not cover the whole ϕSfi11 gene, only part of the gene showed similarity to the database entry. The boxes at the bottom of the figure identify the similarities; refer to Table 1 for the similarity with ϕTP901-1. The orfs showing similarity to the major structural proteins identified by N-terminal sequencing in ϕO1205 (Stanley *et al.,* 1997) are given above the figure. Genes that could code for the high molecular weight minor proteins are suggested in the top lines of the figure. The gene products are indicated by their calculated mass in kilodaltons; the observed masses of the larger minor proteins are given in parentheses. References: *Streptococcus pneumoniae* ϕCp-1 (Martin *et al.,* 1996), ϕEJ-1 (Lopez *et al.,* 1992) and ϕDp-1 (Sheehan *et al.,* 1996); *Lactococcus lactis* ϕBK5-T (Boyce *et al.,* 1995), ϕTP901-1 (Johnsen *et al.,* 1996), ϕr1t (van Sinderen *et al.,* 1996) and ϕbIL 67 (Schouler *et al.,* 1994); *Bacillus subtilis* ϕSPP1 (Becker *et al.,* 1997); *Streptomyces venezuelae* phage VWB (Anné *et al.,* 1990); and *Mycobacterium tuberculosis* phage L5 (Hatful and Sarkis, 1993).

(Fig. 4). In the left half of the aligned maps the two *S. thermophilus* phages showed no sequence similarity, while each of these phages showed sequence similarity to a phage from a different bacterial genus. Over the right part of the maps a variable degree of sequence similarity was observed between the two phages. Identity exceeding 97 % at the aa level (98.1% bp identity) was found between the genes of the putative lysis cassette (orf 131 to 289). The transition

TABLE 1

Similarity of the Indicated Gene Products from ϕSfi11 to Proteins Coded by a Gene Cluster in *Lactococcus lactis* ϕTP901-1

ϕSfi11 orf (length in aa)	ϕTP901-1 gene	ϕTP901-1 orf (length in aa)	Identical/ aligned aa	*P* value
113	c1	110	25/68	0.0023
104	b3	103	35/102	0.011
114	a1	112	32/109	10^{-13}
128	x	129	37/126	10^{-18}
168	mtp	169	64/167	10^{-25}
117	a2	111	28/85	0.0021
105	c2	120	29/92	10^{-8}
1510	b4	>203	n.a.	0.0039

Note. The first column gives the length of the ϕSfi11 orf in number of encoded amino acids. The orfs were listed according to their order on the ϕSfi11 genome. The second column gives the name of the corresponding ϕTP901-1 gene in the terminology of Johnson *et al.* (1996). mtp means major tail protein. The lengths of the orf in number of encoded amino acids are given in the third column. The fourth column gives the number of identical aa for the two proteins compared over the indicated length of computer-aligned aa (n.a., not applicable since the alignment was over several noncontiguous segments). The fifth column gives the probability derived from BLASTP score for obtaining a match by chance.

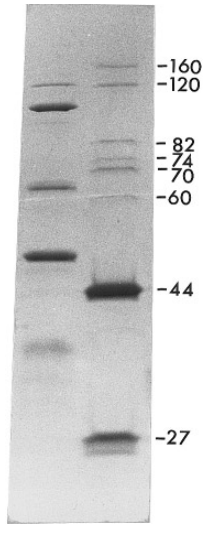

FIG. 3. Structural polypeptides in CsCl density gradient-purified ϕSfi19 (right lane). The gel was stained with Coomassie brilliant blue. Molecular markers are from rotavirus 993/83 (left lane, Brüssow *et al.,* 1992). Molecular masses are given in kilodaltons.

TABLE 2

Comparison of the Indicated ϕSfi11 Gene and Gene Product with the Corresponding Element from ϕO1205 and Analysis of the Differences

ORF ϕSfi11	ORF ϕ1205	bp differences (%)	aa differences (%)	Distribution of bp differences	Gaps	Deletion/insertion
502	502	10.4	7.2	H	2	—
284	297	10.2	8.5	H	2	D:18,21
193	196	5.9	5.2	H	1	—
119	119	16.2	11.0	C:3′	2	—
348	348	7.0	4.0	H	0	—
113	113	17.4	16.0	H	1	—
104	104	5.1	8.0	C:3′	0	—
114	114	0.0	0	—	—	—
128	128	3.9	5.0	C:3′	0	—
168	168	5.2	2.0	H	0	—
117	117	1.1	2.0	—	—	—
105	105	9.8	3.0	H	1	—
1510	1517	7.7	7.5	C:middle	14	D:21
512	512	8.1	5.5	H	2	—
1000	1006	7.5	5.5	H	1	D:18
373	373	6.9	5.1	H	1	—
695	843	High	High	C	Many	Mosaic
669	669	9.9	8.0	C:3′	8	—
149	117	13.1	13.7	C:3′	0	I:36
141	141	5.2	5.0	C:5′	0	—
80	80	21.6	22.5	H	1	—
281	281	19.5	17.8	H	9	—

Note. ORFϕSfi11: the orfs were listed according to their order on the ϕSfi11 genome (Fig. 3). bp differences: The percentage of basepair differences between the corresponding orfs of the two phages. aa differences: the percentage of amino acid differences between the two proteins. Distribution: The distribution of the basepair differences in the compared orfs was classified as homogeneous (H) or clustered (C); the location of the clustered base pair differences at the 3′ or 5′ end or the middle of the orf is given. Deletion: deletions (D) and insertions (I) observed in the alignment of the corresponding orfs. The numbers give the length of the D/I in basepairs deduced from the SIM alignment. If more than one number is given, multiple D/I were observed. The distinction of D/I is arbitrarily based on the ϕSfi11 sequence.

zone from very high to moderately high sequence similarity coincided with the start codon of orf 131 in ϕSfi19. A complex pattern of similarity was observed between orf 1291 gp from ϕSfi19 and its complements in ϕSfi11. Two adjacent N-terminal segments of the orf 1291 gp showed similarity to the N- and C-terminal parts, respectively, of the orf 1000 gp (Fig. 4). These two segments were shared with *L. lactis* phage BK5-T (Desiere *et al.,* 1998). In addition, part of the intervening segment in orf 1000 gp showed similarity with a minor tail protein from *S. pneumoniae* phage Cp-1 (36% aa identity, $P = 10^{-11}$). Finally, the C-terminus from orf 1000 gp showed significant similarity with the C-terminus from the orf 373 gp (34% aa identity over 90 aa, $P = 10^{-6}$).

An even more complicated pattern emerged from the comparison of ϕSfi11 orf 695 gp with the corresponding proteins from ϕSfi19 and ϕO1205. The triple alignment allowed the demarcation of 10 segments. Over the central 150-aa segment 7, all three proteins showed 90% aa identity. This region contained collagen-like GXY repeats which were also found in the lactococcal phage BK5-T (Fig. 4). Over the 170-aa C-terminal segment 10, the ϕSfi11 protein was 98% identical to the ϕO1205 protein, while the ϕSfi19 protein was only 50% identical. This part of the ϕSfi11 protein showed strong sequence similarity to orf 1 gp from *S. pneumoniae* phage Dp-1 (Fig. 4). In contrast, the short (20 to 70 aa), interspersed segments 3, 5, and 9 from the ϕSfi11 protein demonstrated >80% identity with the ϕSfi19 protein, while the ϕO1205 protein was only distantly related. Segments 3 and 5 demonstrated strong similarities to *L. lactis* phage c2 and BK5-T proteins. Finally, the 130- to 150-aa-long segments 4 and 8 showed less than 15% aa identity between the three *S. thermophilus* phages, while the ϕSfi19 protein showed moderately high sequence identity (38%) to the tail tip protein from the lactococcal phage bIL67 (orf 35 gp). A number of gaps were introduced by the alignment program (e.g., segment 2 which is lacking in ϕSfi11). Many of the gaps in the aa alignment reflected gaps in the same positions of the nucleotide alignment, possibly indicating deletion/insertion processes. Interestingly, a spontaneous deletion covering segments 2, 3, 4, and 5 was observed during serial passage of phage ϕSfi21 in our laboratory (Fig. 4; Desiere *et al.,* 1998).

Conservation of genome organization

Next we did an alignment of the genetic maps from phages which showed significant ($P \leq 10^{-5}$) sequence similarity with ϕSfi11 genes over the structural gene cluster (Fig. 5). We took ϕSfi11 orf 1510 as the starting

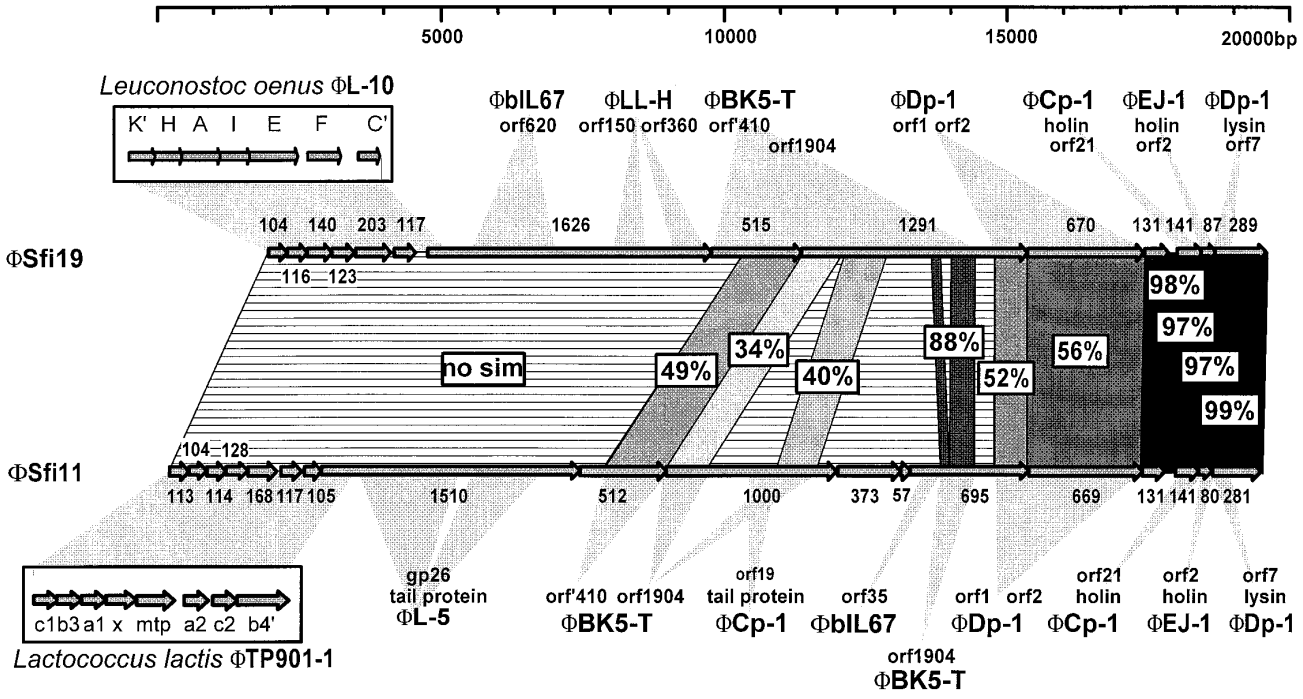

FIG. 4. Alignment of the partial gene maps from *S. thermophilus* bacteriophages ϕSfi19 (top line) and ϕSfi11 (bottom line). The predicted open reading frames are indicated with their orientations and are marked with their length in amino acids. The ruler at the very top gives the scale in base pairs starting with orf 113 from ϕSfi11. Similarities to proteins from the database are indicated by shading and are identified by the phage name, the orf numbering used in the original description, and proposed function (for references see Fig. 3 and Desiere *et al.,* 1998). When the shading does not cover the whole gene, only part of the gene showed similarity to the database entry. The boxes at the left bottom and top of the figure identify similarly organized gene clusters in other phages. Regions of similarity between ϕSfi19 (top line) and ϕSfi11 (bottom line) are indicated by shadings in the central part of the figure. The degrees of aa similarity are graded from black to light grey. The percentages express the percentage of aa identity between the depicted regions. The staggered percentages over the lysis cassette give the aa identity for the gene products predicted for orf 131 to 289, respectively. Zones connected by horizontal crossing are regions showing no significant similarity at the aa sequence level ($P > 0.01$).

point for the orientation of the maps since it showed significant sequence similarity with gps from three *Lactococcus* phages, one *Bacillus* phage, and one *Mycobacterium* phage. Three genes upstream from this reference gene we found a gene coding for a major structural (putative or proven tail) protein (exception: ϕ r1t where the distance is five genes). Six genes upstream of the putative major tail gene a further major structural gene was localized in all of the phages coding for the putative or proven head protein (exception: ϕsk1 where the distance was five genes). ϕSfi11 differs from the other phages by having two adjacent major structural genes at this position. Directly upstream of this structural gene, ϕSfi11 had a gene that showed significant similarity to a ϕr1t gene that preceded the major ϕr1t structural gene. Finally, at the left end of the partial ϕSfi11 map, we identified two genes that showed strong and weak similarity, respectively, with two adjacent genes from *Bacillus* phage SPP1 encoding a portal and a prohead protein.

Over the structural gene cluster, no sequence similarity was detected between coliphage lambda, a Siphoviridae from a gram-negative bacterium, and Siphoviridae from gram-positive bacteria. However, if phage lambda gene H is aligned with orf 1510 from ϕSfi11, a strikingly similar gene order was observed in phage lambda. As in the lactococcal, streptococcal, and bacillar Siphoviridae, three genes upstream from this reference gene phage lambda possesses a gene coding for the major tail protein. Five genes upstream from the major tail gene, phage lambda shows the major head gene. This constellation is similar to ϕsk1, but one gene shorter than in the other Siphoviridae from the gram-positive bacteria. The difference might be due to a possible "displacement" of gene W from its morphogenetic context: lambda gps U, Z, FII, and W interact to join the head to the tail (Georgopoulos *et al.,* 1983; Katsura, 1983). Three of these proteins are clustered between the major head and major tail encoding genes, while gene W is with the prehead assembly genes. The interspersed character of gene W was previously noted from comparisons of the gene map of two lambdoid phages (λ and P22, Eppler *et al.,* 1991, but see also Smith and Feiss, 1993).

DISCUSSION

Over the morphogenesis module, our analysis revealed striking parallels in the genome structure of Siphoviridae. Lambda virologists have already observed that the order of action of the gene products during phage lambda assembly is similar to the arrangement of the genes on the lambda genome. It has been argued that the order of action is based on the structural inter-

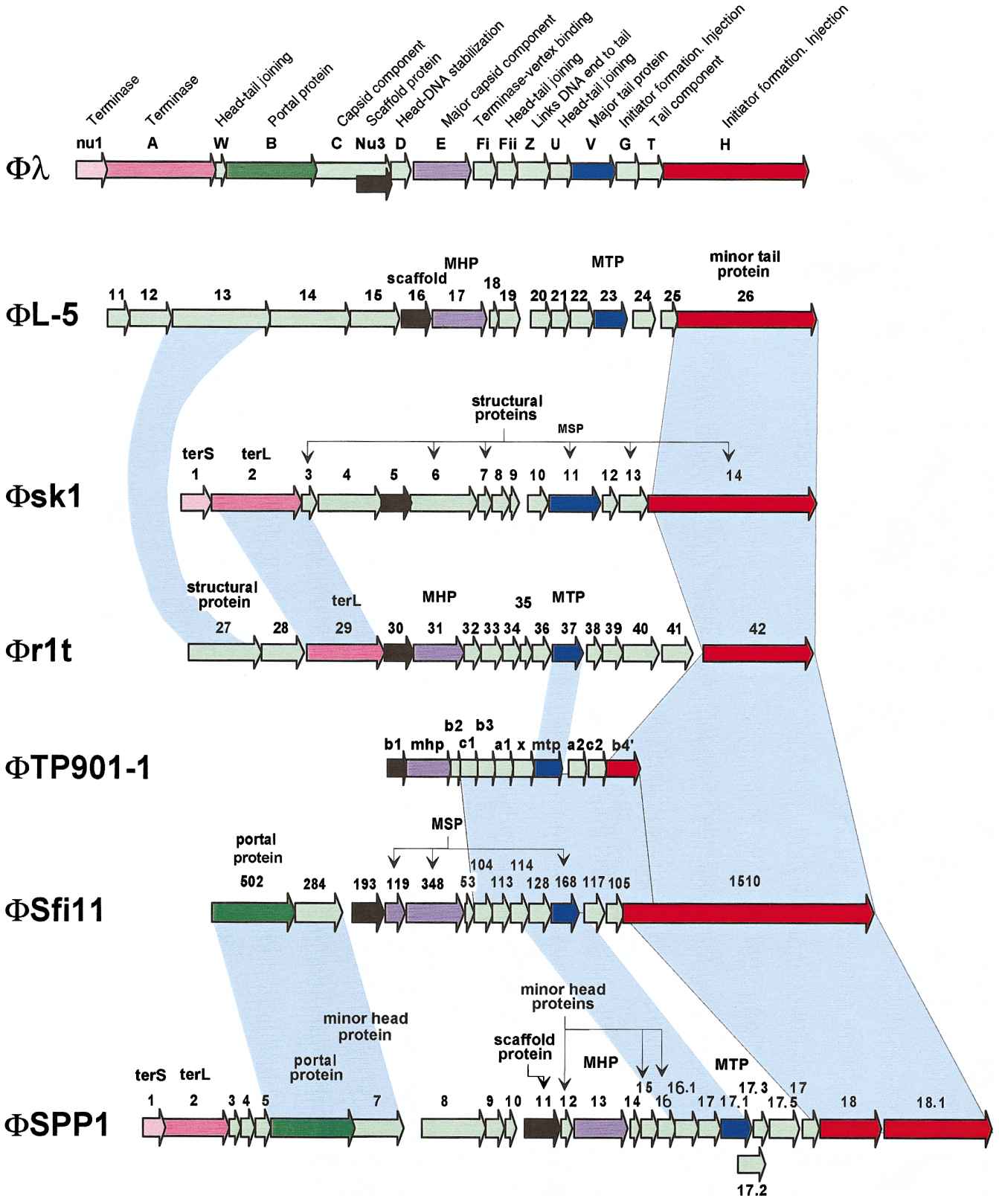

FIG. 5. Comparison of the organization of the head and tail assembly genes in coliphage λ (top line), *Mycobacterium* phage L-5 (second line), *L. lactis* phages sk1 (third line), r1t (fourth line), TP901-1 (fifth line), *S. thermophilus* phage Sfi11 (sixth line), and *B. subtilis* phage SPP1 (bottom line). The genomes are aligned at the right by a region coding for a putative minor tail protein (red) showing sequence similarity over all listed Siphoviridae from gram-positive bacteria. Genes coding for the putative or proven major tail protein (blue), major head protein (purple), scaffolding protein (black), portal protein (dark green), and small subunit terminases (light pink) and large subunit terminase (dark pink) were coded. The length of the arrow is proportional the length of the predicted open reading frame. The individual orfs were identified by the numbering system used in the original publication (references : *L. lactis* phage sk1 (Chandry *et al.,* 1997); for all others see Fig. 3). For phage lambda the functional attribution of the genes is indicated. Zones of light blue link regions showing significant sequence similarity ($P \leq 10^{-5}$).

action between proteins involved in adjacent steps of the assembly pathway (Katsura, 1983). In addition, it was argued that the conservation of the gene order in lambdoid phages is of clear evolutionary advantage since it minimizes the number of unproductive interactions that arise by recombination between phages having partially homologous chromosomes (Casjens and Hendrix, 1974). However, this hypothesis can explain the conservation of the gene order only within closely related phage groups such as the *E. coli* lambdoid phages where such hybrids can and do form. It cannot be a present day force responsible for maintaining the gene order between λ and the *Salmonella* phage P22 since there is not enough similarity in sequence over the morphogenesis operon (Eppler *et al.,* 1991). This argument is even more evident when Siphoviridae from gram-positive and gram-negative bacterial hosts are compared. Alternatively, the conserved gene order could be a relic from an ancient common ancestor or an undefined selection force that favors this particular order.

The similarity of the lambda structural gene map with that of *L. lactis* phage sk1 and *B. subtilis* phage SPP1 was observed recently by Chandry *et al.* (1997) and Becker *et al.* (1997), respectively. On the basis of our multiple comparisons of gene maps we propose that the gene map from phage lambda can be used to predict tentative gene functions in uncharacterized Siphoviridae. For an industrial laboratory dealing with dairy phages, the genetic basis of the phage host range is of obvious practical importance. Therefore, we tested the predictive power of this method by searching for the streptococcal phage complement to the tail tip protein of phage lambda responsible for mediating phage adsorption. In phage lambda, this protein is encoded by gene J, which is the last gene of the morphogenesis operon (Katsura, 1983). Our bioinformatic analysis localized the end of the structural gene cluster in φSfi11 at orf 695 or 669. In fact, orf 695 gp fulfils several requirements for a phage adsorption protein. First, it differed among phages that show distinct host ranges (φSfi11/O1205; Sfi11/Sfi19), while it was nearly identical among phages showing an extensively overlapping host range (φSfi19/Sfi21). Second, orf 695 gp resembles the anti-receptor of phage T4 (protein 38) with its alternating stretches of high, low, and no homology in multiple alignments with related phages (Montag *et al.,* 1987). In addition, when a number of lambdoid phages were compared by heteroduplex analysis, gene J showed also conserved and variable segments (Highton *et al.,* 1990). Third, the DNA region covered by orf 695 is a recombinational hotspot in streptococcal phages. A large spontaneous deletion was observed several times in the orf 695 orthologue from φSfi21 (Bruttin and Brüssow, 1996). This deletion was flanked by a nearly perfect 53-bp repeat (Desiere *et al.,* 1998) containing collagen-like GXY repeats. In addition, collagen-like repeats were also found in orf 1904 gp from lactococcal phage BK5-T. Interestingly, this protein possesses four perfect tandem repeats containing the collagen-like motifs (Boyce *et al.,* 1995a) which also suffered spontaneous deletions (Boyce *et al.,* 1995b). Further oligoglycine repeats were found in the conserved segments of orf 695 gp. The corresponding DNA repeats could lead to deletions by slippage of the DNA polymerase or to DNA exchanges or DNA expansion by unequal crossover events. Interestingly, a structure prediction for the adsorption protein from phage T4 suggested a number of hypervariable loops mediating the receptor recognition held together by conserved oligoglycine stretches (Henning and Hashemolhosseini, 1994). The *S. thermophilus* phages might thus resemble coliphages in which the evolution of tail fiber genes apparently occurs by recombinational reshuffling (Haggard-Ljungquist *et al.,* 1992).

We analyzed the similarity of *S. thermophilus* phages to other phages at the sequence level and in an evolutionary context. The sequence comparisons revealed a hierarchy of relatedness. At the first level are relationships between *S. thermophilus* phages belonging to the same lytic group: they showed a nearly identical gene map and differed at the DNA sequence level by about 10% (Desiere *et al.,* 1998). At the second level of relatedness are relationships between *S. thermophilus* phages belonging to different lytic groups. They showed over the whole genome a comparable gene order, but at the sequence level they are patchy, demonstrating regions of no and very high (> 99%) DNA sequence identity (Brüssow *et al.,* 1994b). At the third level of relatedness are *S. thermophilus* φSfi21 and the *L. lactis* φBK5-T (Boyce *et al.,* 1995a). The gene order is very similar and many genes showed high sequence similarity (up to about 63% at the aa level). It should be noted that this is the level of relatedness between two lambdoid coliphages (Smith and Feiss, 1993). The similarity was not restricted to the morphogenesis operon, but was also found over the lysogeny module (Bruttin *et al.,* 1997a). At a similar level are the relationships between *S. thermophilus* φSfi11 and several *Lactococcus* phages (φTP901-1, r1t, 7-9) or *S. thermophilus* φSfi21 with *Leuconostoc* φL10 (Desiere *et al.,* 1998). High sequence similarity (up to 56% at the aa level) was detected over several adjacent genes. At the fourth level are relationships between *B. subtilis* φSPP1 and φSfi11. Adjacent genes from two different modules (structural genes, DNA packaging, Becker *et al.,* 1997) showed aa similarity (≤ 30% identity). At the same level are the relationships with *S. pneumoniae* phages over the lysis module. At the fifth level are relationships between *S. thermophilus* phages and Siphoviridae from taxonomically more distant bacterial hosts like *Lactobacillus* (Desiere *et al.,* 1998), *Streptomyces,* and *Mycobacterium.* Sequence similarity was found for individual genes (≤ 30% aa identity), but no longer over adjacent genes. Finally, we have the distant relationship between *S. thermophilus* phages and phage lambda from gram-negative bacteria. Over the structural gene operon both

phages showed a relatively conserved gene order but no sequence similarity.

The observation of this series of graded relatedness is notable since it is the hallmark of any biological system undergoing evolutionary changes. The fact that the degrees of relatedness are correlated approximately with the evolutionary distance between the bacterial hosts (*S. thermophilus*→*Lactococcus*→*Leuconostoc*→*Bacillus*→*Lactobacillus* (all low GC group of gram-positive bacteria)→*Mycobacterium*→*Streptomyces* (both high GC group of gram-positive bacteria)→*Escherichia* (gram-negative bacterium)) is intriguing. It will be important to confirm this correlation from the perspective of other Siphoviridae isolated from an evolutionarily distant host. This is currently not possible since only relatively few phage sequences of Siphoviridae are in the database and their bacterial hosts do not represent the evolutionary diversity of eubacteria. Our data are compatible with the hypothesis that the morphogenesis operon from Siphoviridae of the low G + C group of gram-positive bacteria is derived from a common ancestor. In contrast, the sequence similarity between Siphoviridae of the high and low G + C group of gram-positive bacteria (which was limited to individual genes) is more likely to represent horizontal gene transfer in a relatively distant past than common ancestry. The similarity between the morphogenesis operons from phage lambda and Siphoviridae from gram-positive bacteria allows two interpretations: convergent evolution (then a very astonishing one) or splitting of the two lines in a relatively distant past which obscured all aa similarities. Current data do not constrain the time scale of the latter process. It is possible that all extant Siphoviridae diverged very rapidly from a common ancestor. The high rate of bp changes observed between coliphages is indirect evidence for a rapid pace of coliphage evolution. Therefore, the impression of a coevolution of the phages with the host bacteria does not necessarily implicate descent from a phage ancestor which already existed when the bacterial genera split apart. We suspect that the ancestor of Siphoviridae is much younger than the separation of gram-positive and gram-negative bacteria. A putative ancestor phage could have invaded different bacterial genera in an evolutionary not too distant past and the Siphoviridae split into distinct lines due to separations of the gene pools affected by more or less tight host range barriers between bacterial species or genera.

What does the comparative sequencing approach tell us about the definition of a phage species? Since lytic group I and II *S. thermophilus* phages have clearly distinct structural genes, we might define phages from a single lytic group as a phage species. However, any biologically meaningful species definition should take exchange of genetic material as an inclusion criterion and lack of this exchange as an exclusion criterion. In view of the possibility of horizontal gene transfer between different phage systems, the species barrier cannot be absolute even between clearly distinct phage systems. Between lytic group I and II *S. thermophilus* phages we observed very high DNA sequence identity in the DNA replication module (>99%, Brüssow *et al.,* 1994b; Desiere *et al.,* 1997) and the lysis cassette. There is no process which can explain this high sequence conservation except recent DNA exchanges. Sequence similarity of >90% was observed over approximately a quarter of the genomes from lytic group I and II phages (Desiere *et al.,* manuscript in preparation). For us it is not meaningful to separate these phages into two species. In contrast, ϕBK5-T and *S. thermophilus* ϕSfi21 differed by at least 35% at the aa level. As long as no lactococcal or streptococcal phages are described that narrow this gap, we anticipate that these phage groups share a common ancestor and are not currently exchanging DNA and thus do not belong to the same phage species. On the other end of taxonomical hierarchy, one might lift the family Siphoviridae to order level (Siphovirales) and reserve the family level, for example, for Siphoviridae from the low GC group of gram-positive bacteria and genus level to phages which are as similar as lactococcal ϕBK5-T and streptococcal ϕSfi21. However, any definition of higher taxonomic groups in Siphoviridae depends on the postulated model of phage evolution.

MATERIALS AND METHODS

Phages, strains, and media

The phages were propagated on their appropriate *S. thermophilus* hosts in lactose M17 broth as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995). *E. coli* strain JM 101 (Stratagene) was grown in LB broth or on LB broth solidified with 1.5% (w/v) agar. Ampicillin, IPTG, and X-gal (all from Sigma) were used at concentrations of 100 μg /ml, 1 mM, and 0.002% (w/v), respectively.

DNA techniques

Phage purification and DNA extraction were done as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995; Bruttin *et al.,* 1997a). Plasmid DNA was isolated using Qiagen midi-plasmid isolation columns. Restriction enzymes were obtained from Boehringer Mannheim and used according to the supplier's instructions.

Protein techniques

Phage particles were concentrated by PEG precipitation and purified by two rounds of CsCl density gradient centrifugation (3 h at 40,000 rpm using a Beckman SW55.5 rotor) on a 5-step preformed CsCl gradient (n_D = 1.4, 1.372, 1.3698, 1.3682, 1.367). The phage bands were recovered with a Pasteur pipette, diluted in phage buffer

(Brüssow and Bruttin, 1995) and concentrated by high speed centrifugation (1 h, 40,000 rpm, SW55.5 rotor, Beckman). The purified phage particles were then denatured for 2 min at 100°C using SDS gel-loading buffer with β-mercaptoethanol. SDS–PAGE was done on 8 % acrylamid slab gels which were subsequently stained with Coomassie brilliant blue (Bio-Rad).

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, Switzerland). Both strands of the cloned DNA were sequenced by the Sanger method of dideoxy-mediated chain termination using the *fmol* DNA Sequencing System of Promega (Madison, WI). The sequencing primers were end-labeled using [γ-^{33}P]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 Sfi21 DNA were sequenced using the Amersham Labstation sequencing kit based on Thermo Sequenase-labeled primer cycle sequencing with 7-deaza-dGTP(RPN2437). Sequencing was done on a Licor 6000L automated sequencer with fluorescence-labeled universal reverse and forward pUC19 primers.

PCR

PCR was used to span regions, which were not obtained through random cloning. PCR products were generated using the synthetic oligonucleotide pair designed according to the established ϕSfi21 DNA sequence, purified phage DNA, and Super *Taq* polymerase (Stehelin, Basel, Switzerland). PCR products were purified using the QIAquick-spin PCR purification kit.

Sequence analysis

The Genetics Computer Group sequence analysis package (University of Wisconsin) was used to assemble and analyze the sequences. Nucleotide (nt) and predicted amino acid (aa) sequences were compared to those in the databases (GenBank, release 102; EMBL (abridged), release 51; PIR-protein, release 53; SWISS-PROT, release 34; PROSITE, release 13.0) using the FastA (Lipman and Pearson, 1985) and BLAST (Altschul *et al.,* 1990) programs. Sequence alignments were performed using the CLUSTALW 16 method (Thompson *et al.,* 1994), the Multalign program (Corpet, 1988), http://www.toulouse.inra.fr/multalin.html, and the SIM alignment tool (Huang and Miller, 1991, http://expasy.hcuge.ch/sprot/sim-nucl.html).

The ΦSfi11 sequence was deposited in the GenBank database under Accession No. AF057033.

ACKNOWLEDGMENTS

We thank V. Fryder for electron microscopy, Anne Bruttin for her contributions in the initial phase of the work, Fanny Scalfo for valuable technical assistance, Sophie Foley for critical reading of the manuscript, and the Swiss National Science Foundation for financial support of Sacha Lucchini and Frank Desiere in the framework of its Biotechnology Module (Grant 5002-044545/1).

REFERENCES

Ackermann, H.-W., and DuBow, M. S. (1987). Bacteriophage taxonomy. *In* "Viruses of Prokaryotes" (H.-W. Ackermann and M. S. DuBow, Eds.), Vol. I, Chap. 2, pp. 13–28. CRC Press, Baca Raton, Florida.

Anné, J., van Mellaert, L., Decock, B., van Damme, J., van Aerschoot, A., Herdewijn, P., and Eyssen, H. (1990). Further biological and molecular characterization of actinophage VWB. *J. Gen. Microbiol.* 136, 1365–1372.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Becker, B., de la Fuente, N., Gassel, M., Günther, D., Tavares, P., Lurz, R., Trautner, T. A., and Alonso, J. C. (1997). Head morphogenesis genes of the *Bacillus subtilis* bacteriophage SPP1 *J. Mol. Biol.* 268, 822–839.

Botstein, D. (1980). The theory of modular evolution in bacteriophages. *Ann. N. Y. Acad. Sci.* 354, 484–491.

Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1995b). 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.

Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1995a). Sequence analysis of the lactococcus lactis temperate bacteriophage BK5-T and demonstration thsat the phage DNA has cohesive ends. *Appl. Environ. Microbiol.* 61, 4089–4098.

Brüssow, H., and Bruttin, A. (1995). Characterization of a temperate *Streptococcus thermophilus* bacteriophage and its genetic relationship with lytic phages. *Virology* 212, 632–640.

Brüssow, H., Frémont, M., Bruttin, A., Sidoti, J., Constable, A., and Fryder, V. (1994a). Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. *Appl. Environ. Microbiol.* 60, 4537–4543.

Brüssow, H., Probst, A., Frémont, M., and Sidoti, J. (1994b). Distinct *Streptococcus thermophilus* bacteriophages share an extremely conserved DNA fragment. *Virology* 200, 854–857.

Bruttin, A., and Brüssow, H. (1996). Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. *Virology* 219, 96–104.

Bruttin, A., Desiere, F., d'Amico, N., Guerin, J. P., Sidoti, J., Huni, B., Lucchini, S., and Brüssow, H. (1997a). Molecular ecology of *Streptococcus thermophilus* bacteriophage infections in a cheese factory. *Appl. Environ. Microbiol.* 63, 3144–3150.

Bruttin, A., Desiere, F., Lucchini, S., Foley, S., and Brüssow, H. (1997b). Characterization of the lysogeny DNA module from the temperate *Streptococcus thermophilus* bacteriophage ϕSfi21. *Virology* 233, 136–148.

Campbell, A. (1988). Phage evolution and speciation. *In* "The Bacteriophages" (R. Calendar, Ed.), Chap. 1, pp. 1–14. Plenum, New York.

Campbell, A., and Botstein, D. (1983). Evolution of the lambdoid phages. *In* "Lambda II" (R. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, Eds.), pp.365–380. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Casjens, S., Hatfull, G., and Hendrix, R. (1992). Evolution of dsDNA tailed-bacteriophage genomes. *Semin. Virol.* 3, 383–397.

Casjens, S., and Hendrix, R. (1974). Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. *J. Mol. Biol.* 90, 20–23.

Chandry, P. S., Moore, S. C., Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1997). Analysis of the DNA sequence, gene expression, origin of

replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. *Mol. Microbiol.* 26, 49–64.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic. Acids. Res.* 16, 10881–10890.

Desiere, F., Lucchini, S., Brüssow, H. (1998). Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small insertions. *Virology* 241, 61–72.

Desiere, F., Lucchini, S., Bruttin, A., Zwahlen, M.-C., and Brüssow, H. (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.

Draper, D. E. (1996). Translational Initiation. *In* "*Escherichia coli* and *Salmonella:* Cellular and Molecular Biology" (F. C. Neidhardt, Ed.), Chap. 59, pp. 902–908. ASM Press, Washington, DC.

Eppler, K., Wyckoff, E., Goates, J., Parr, R., and Casjens, S. (1991). Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. *Virology* 183, 519–538.

Fauquet, C. M. (1997). International Committee on Taxonomy of Viruses. Chart available from ILTAB@scripps.edu.

Fayard, B., Haeflinger, M., and Accolas, J.-P. (1993). Interactions of temperate bacteriophages of *Streptococcus salivarius* subsp. *thermophilus* with lysogenic indicators affect phage DNA restriction patterns and host ranges. *J. Dairy Res.* 60, 385–399.

Georgopoulos, C., Tilly, K., and Casjens, S. (1983). Lambdoid phage head assembly. In "Lambda II" (R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, Eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Griffin, D. E., and Bellini, W. J. Measles virus. (1996). *In* "Fields Virology" B. Fields, (Ed.), Chap. 43, pp. 1267–1312. Lippincott–Raven, Philadelphia.

Guédon, G., Bourgoin, F., Pébay, M., Roussel, Y., Colmin, C., Simonet, J. M., and Decaris, B. (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.

Haggard-Ljungquist, E., Halling, C., and Calendar, R. (1992). DNA sequences of the tail fibre genes of bacteriophage P2: Evidence for horizontal transfer of tail fibre genes among unrelated bacteriophages. *J. Bacteriol.* 174, 1462–1477.

Hatful, G. F., and Sarkis, G. J. (1993). DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol. Microbiol.* 7, 395–405.

Henning, U., and Hashemolhosseini, S. (1994). Receptor recognition by T-even-type coliphages. *In* "Molecular Biology of Bacteriophage T4" (J. D. Karam, Ed.). ASM Press, Washington, DC.

Highton, P. J., Chang, Y., and Myers, R. J. (1990). Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. *Mol. Microbiol.* 4, 1329–1340.

Huang, X., and Miller, W. (1991). A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* 12, 337–357.

Johnsen, M. G., Appel, K. F., Madsen, P. L., Vogensen, F. K., Hammer, K., and Arnau, J. (1996). A genomic region of lactococcal temperate bacteriophage TP901-1 encoding major virion proteins. *Virology* 218, 306–315.

Katsura, I. (1983). Tail assembly and injection. *In* "Lambda II" (R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, Eds.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

LeMarrec, C., van Sinderen, D., Walsh, L., Stanley, E., Vlegels, E., Moineau, S., Heinze, P., Fitzgerald, G., and Fayard, B. (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.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441.

Lopez, R., Garcia, J. L., Garcia, E., Ronda, C., and Garcia, P. (1992). Structural analysis and biological significance of the cell wall lytic enzymes of *Streptococcus pneumoniae* and its bacteriophage. *FEMS Microbiol. Lett.* 79, 439–447.

Martin, A. C., Lopez, R., and Garcia, P. (1996). Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1. *J. Virol.* 70, 3678–3687.

Mercenier, A. (1990). Molecular genetics of *Streptococcus thermophilus. FEMS Microbiol. Rev.* 87, 61–78.

Milkman, R. (1996). Recombinational exchange among clonal populations. *In* "*Escherichia coli* and *Salmonella:* Cellular and Molecular Biology" (F. C. Neidhardt, Ed.), Chap. 145, pp. 2663–2684. ASM Press, Washington, DC.

Monod, C., Repoila, F., Kutateladze, M., Tetart, F., and Krisch, H. M. (1997). The genome of the pseudo T-even bacteriophages, a diverse group that resembles T4. *J. Mol. Biol.* 267, 237–249.

Montag, D., Riede, I., Eschbach, M.-L., Degen, M., and Henning, U. (1987). Receptor-recognizing proteins of T-even type bacteriophages. Constant and hypervariable regions and an unusual case of evolution. *J. Mol. Biol.* 196, 165–174.

Murphy, F. A. (1996). Virus taxonomy. *In* "Fields Virology" (B. Fields, Ed.), Chap. 2, pp. 15–58. Lippincott–Raven, Philadelphia.

Ochman, H., and Lawrence, J. G. (1996). Phylogenetics and the amelioration of bacterial genomes. *In Escherichia coli* and *Salmonella:* Cellular and Molecular Biology" (F. C. Neidhardt, Ed.), Chap. 142, pp. 2627–2637. ASM Press, Washington, DC.

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

Sheehan, M. M., Garcia, J. L., Lopez, R., and Garcia, P. (1996). Analysis of the catalytic domain of the lysin of the lactococcal bacteriophage Tuc2009 by chimeric gene assembling. *FEMS Microbiol. Lett.* 140, 23–28.

Smith, M. P., and Feiss, M. (1993). Sequence analysis of the phage 21 genes for prohead assembly and head completion. *Gene* 126, 1–7.

Stanley, E., Fitzgerald, G., Le Marrec, C., Fayard, B., and van Sinderen, D. (1997). Sequence analysis and characterization of ϕO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology* 143, 3417–3429.

Strauss, E. G., Strauss, J. H., and Levine, A. J. (1996). Virus evolution. *In* "Fields Virology" (B. Fields, Ed.), Chap. 6, pp. 153–172. Lippincott–Raven, Philadelphia.

Susskind, M., and Botstein, D. (1978). Molecular genetics of bacteriophage P22. *Microbiol. Rev.* 42, 385–413.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (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.

van Sinderen, D., Karsens, H., Kok, J., Terpstra, P., Ruiters, M. H. J., Venema, G., and Nauta, A. (1996). Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage rlt. *Mol. Microbiol.* 19, 1343–1355.

Whittam, T. S. (1996). Genetic variation and evolutionary processes in natural populations of *Escherichia coli. In* "*Escherichia coli* and *Salmonella:* Cellular and Molecular Biology" (F. C. Neidhardt, Ed.), Chap. 148, pp. 2708–2722. ASM Press, Washington, DC.