Doctoral Thesis

Genetic diversity of Streptococcus thermophilus phages and development of phage-resistant starters for the dairy industry

Author(s):
Lucchini, Sacha

Publication Date:
1999

Permanent Link:
https://doi.org/10.3929/ethz-a-003839329

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Genetic Diversity of *Streptococcus thermophilus* Phages and Development of Phage-Resistant Starters for the Dairy Industry

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETHZ)

for the degree of

Doctor of Natural Sciences

Presented by

SACHA LUCCHINI
Dipl. Biol. ETH
Born March 13, 1970
Citizen of Frasco (TI)

Accepted on the recommendation of
Prof. Dr. Michael Teuber, examiner
Dr. Harald Brüssow, co-examiner

Zürich 1999
Acknowledgments

This work was carried out in the Department of Bioscience, at the Nestlé Research Centre in Vers-chez-les-Blanc. I thank Dr. Andrea Pfeiffer and Jean-Richard Neeser for providing me with excellent working facilities. I would like to extend my thanks to Prof. Michael Teuber for supporting the project and having accepted me as a Ph.D. student. This study was supported by grants from the Swiss National Science Foundation in the framework of the SPP Biotechnology module (grant 5002-044545/1).

My first acknowledgment goes to my supervisor Dr. Harald Brüssow for having guided me throughout my thesis. I would also like to thank Frank Desiere for introducing me to the secret world of Nestlé research; Josette Sidoti, Fanny Scalfo and Déborah Moine for their technical support; Sophie Foley, Marie-Camille Zwahlen and Michelle Delley for useful discussions; and Anne Bruttin for technical advice and not allowing me to transform the lab into something else.

Last but not least, I am grateful to my friends in Lausanne for the good time I had with them.
Contents

1 Introduction 1

1.1 Bacteriophages 2

1.1.1 Classification of bacteriophages 2
1.1.2 Evolution of viruses 3
1.1.3 The life cycle of viruses 4

1.1.3.1 Bacteriophage T4 5
1.1.3.2 Bacteriophage lambda 6

1.1.4 Bacteriophages of lactic acid bacteria 8

1.2 Phage resistance mechanisms in LAB 10

1.2.1 Natural resistance mechanisms 10

1.2.1.1 Phage adsorption inhibition 10
1.2.1.2 Block of phage DNA injection 11
1.2.1.3 Abortive infection mechanisms 11
1.2.1.4 Restriction and modification 12

1.2.2 Novel phage resistance mechanisms 12

1.3 Outline of the thesis 13

1.3.1 Phage diversity 13
1.3.2 Phage resistant starter strains 14

2 Results 21

2.1 The genetic relationship between virulent and temperate Streptococcus thermophilus bacteriophages: whole genome comparison of cos-site phages Sfi19 and Sfi21 21

2.2 The structural gene module in Streptococcus thermophilus bacteriophage φSfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts 35

2.3 Comparative genomics of Streptococcus thermophilus phage species supports a modular evolution theory 47

2.4 Similarly organized lysogeny modules in temperate Siphoviridae from low GC content Gram-positive bacteria 71
2.5 A short noncoding viral DNA element showing characteristics of a replication origin confers bacteriophage resistance to *Streptococcus thermophilus*

2.6 Construction of food-grade *Streptococcus thermophilus* starters with potent, broad range phage resistance phenotypes

3 Summary and conclusions
Abstract

To address the question of the diversity of *Streptococcus thermophilus* phages we sequenced three lytic phages. The phage genomes underwent pairwise alignments by including two temperate and one further virulent *S. thermophilus* phages to localize areas of difference. Pairwise and multiple alignments of these sequences supports the theory of modular evolution of bacteriophages: genes are clustered into functional units (modules), all *S. thermophilus* phages have a similar genomic organization (DNA-packaging followed by head morphogenesis, tail morphogenesis, host lysis, lysogeny module, DNA replication and finally a region possibly involved in phage development control). Individual *S. thermophilus* phages appear to result from recombination of exchangeable genetic elements drawn from a common gene pool. Recombination is apparently the basis of differences in life style (lytic, temperate), host range, DNA packaging, and structural protein pattern. The units of genetic exchange are either entire functional modules (several adjacent structural genes), part of a module (individual genes in the lysogeny module) or even parts of a gene corresponding to likely protein domains (host range determinant of the likely tail fibers, DNA binding domain of regulatory proteins). A detailed comparison of the DNA sequences from phages Sfi19 and Sfi18 indicates that point mutations may also play an important role in phage adaptation. Very few base pair changes seem to be under the escape of phage Sfi19 from the inhibition mediated by the Sfi21 encoded origin of replication.

We examined the genetic relationship between *S. thermophilus* phages and other phages focusing on modules that are under different evolutionary constraints, the structural gene cluster and the lysogeny module. The structural gene cluster from lytic group II *S. thermophilus* bacteriophage Sfi11 was compared to the corresponding region from other *Siphoviridae*. The analysis revealed a hierarchy of relatedness and a common genomic organization. We hypothesize that the morphogenesis module of B1 *Siphoviridae* shares a common ancestry. Conservation of the genetic organization was also found when comparing the lysogeny modules of temperate *Siphoviridae* from the low GC content Gram positive bacteria. The gene organization of the lysogeny module of this group of phages distinguishes it from the *Siphoviridae* infecting Gram-negative bacteria (e.g.: coliphages λ, 933W and P22) or high GC-content Gram-positive bacteria (mycobacteriophage L5). In contrast, similarly organized lysogeny modules were found in the temperate P2-like *Myoviridae*.

To obtain *S. thermophilus* strains with increased phage resistance, we tested a range of molecular approaches. First, it was shown that the phage Sfi21 replication origin, when cloned in a high copy number plasmid provides protection against a high number of phages. Few point mutations in this region (< 15-nt differences over a stretch of 302 nt) seem sufficient to allow phages to escape the inhibition. However, substitution of the Sfi21 ori with that of insensitive phages was shown to extend the range of inhibitory activity of the origin. The second approach relied on the protection provided by the immunity functions of the Sfi21 prophage. A targeted deletion in an essential tail gene of the prophage resulted in a lysogen, which proved to be unable to release infectious phage particles, but maintained intact superinfection control. Finally, we demonstrated that the random inactivation of host genes followed by a selection for phage-resistant phages could lead to extremely phage-resistant strains.
Résumé

Pour examiner la diversité des phages de *Streptococcus thermophilus*, trois phages lytiques ont été séquencés. Afin de localiser des régions de différenciation, les séquences ainsi obtenues ont été comparées en y encluant deux phages tempérés et un autre phage lytique. L'alignement de ces séquences supporte la théorie d'une évolution modulaire des bactériophages: les gènes sont regroupés en unités fonctionnelles (modules), tous les phages de *S. thermophilus* ont une organisation génomique similaire (l'ordre des modules est le suivant: gènes pour l'insertion du DNA phagique dans la tête, morphogenèse de la tête, morphogenèse de la queue, lyse de la cellule hôte, module de lysogénie, replication de l'ADN et une région probablement impliquée dans le contrôle du développement du phage). Pour certains modules plusieurs allèles existent. Les phages de *S. thermophilus* semblent donc être le résultat de la recombinaison d'éléments génétiques interchangeables puisés dans un pool génétique commun. Les recombinations sont apparentement à la base des différences dans le style de vie (virulent, tempéré), spectre d'hôte, mécanisme d'insertion du DNA dans la tête, protéines de structure. Les unités d'échange génétique peuvent être un module fonctionnel complet (gènes de structure), une partie d'un module (échange de gènes isolés dans le module de lysogénie), ou même une partie d'un gène (déterminants du spectre d'hôte des probables fibres de la queue du phage). Une comparaison détaillée des séquences des phages Su19 et Sfi18 indique que les mutations ponctuelles pourraient aussi avoir un rôle important dans l'adaptation des phages. Un nombre très limité de changements dans la séquence d'ADN de Sfi19 semble être à la base de l'insensibilité de Sfi19 à l'inhibition médiaée par l'origine de réplication du phage Sfi21. 

Nous avons examiné les relations génétiques entre les phages de *S. thermophilus* et les phage infectants d'autres cellules hôtès. Nous nous sommes concentrés sur deux modules fonctionnels qui subissent différentes contraintes évolutionnaires: le module structural et le module de lysogénie. En premier, nous avons comparé le module structural du phage Sfi11 de *S. thermophilus* (groupe lytique II) avec la région correspondante d'autres *Siphoviridae*. L'analyse a révélé un gradient de similitudes et une organisation génomique similaire. Nous avons donc formulé l'hypothèse que le module structural des *Siphoviridae* appartenant au groupe B1 a une origine commune. Une conservation de l'organisation génomique a aussi été détectée pour les modules de lysogénie des *Siphoviridae* tempérés infectant des bactéries Gram⁺ positives à contenus en GC bas. L'organisation des gènes du module de lysogénie de ce groupe de phages les distingue des *Siphoviridae* isolés dans bactéries Gram⁻ (coli phages λ, 933W et P22). Par contre, des modules de lysogénie similaires ont été trouvés chez les *Myoviridae* du groupe P2.

Pour obtenir des souches de *S. thermophilus* possédant une plus grande résistance aux phages, nous avons testé plusieurs approches moléculaires. Premièrement, il a été montré que l'origine de réplication du phage Sfi21 protège contre un grand nombre de phage quand elle est clonée dans un plasmide à grand nombre de copies. Cependant, quelques mutations ponctuelles dans cette région (< 15 mutations dans une région de 302 nucléotides) semblent être suffisantes pour permettre aux phages d'échapper à l'inhibition. La substitution de l'origine de réplication du phage Sfi21 avec la région correspondante de phages non soumis à l'inhibition permet d'étendre le spectre d'action de l'origine de réplication. La seconde approche a été basée sur l'effet protectif du phage Sfi21 intégré dans le chromosome d'une cellule lysogène. La délétion ciblée d'un gène du prophage essentiel pour la morphogénése de la queue a produit une souche lysogène incapable de produire des phages infectieux. Mais la protection médieée par le prophage n'a pas été modifiée. Enfin, nous avons pu démontrer que l'inactivation aléatoire de certains gènes de la cellule hôte, suivie par la sélection de mutants résistants aux phages, a amené à l'isolement de souches extrêmement résistantes aux phages.
1 Introduction

Lactic acid bacteria (LAB) have been used for millennia for the fermentation of various food products (cereals, vegetables, meat and milk). The low pH generated by their metabolism and the action of other fermentation products inhibits the growth of many food spoilage bacteria. In addition to providing an effective preservation, lactic acid bacteria determine the flavor and texture of the products. If the growth of the starter is inhibited, fermentation is delayed or even abolished resulting in the complete loss of the product (64).

Bacteriophage infection is the main cause of starter failure during the industrial production of cheese and yogurt (83). This problem is increased in modern manufacturing where a limited number of defined starter cultures are used to transform large volumes of milk. Industrial fermentation has opened new possibilities for phages to flourish. The most susceptible target starter species are *Lactococcus lactis* and *Streptococcus thermophilus*. The dairy industry has used various practical methods to control the phage problem; regular disinfecting of equipment, use of closed vats, strain rotation and multiple phage-resistant strain starters (3,75).

In *L. lactis* many natural anti-phage mechanisms have been found, mainly on conjugative plasmids (3). These plasmids can be transferred into phage-sensitive commercial strains to increase their resistance to phages (84). Recently, genetic engineering is used to design novel phage resistance mechanisms.

In contrast to *L. lactis*, very few natural phage resistance mechanisms were identified in *S. thermophilus* (55). Plasmid-containing strains are rare (65) and few have phage inhibitory effects (73,78). Spontaneous bacteriophage-insensitive mutants (BIM) have provided a source of replacement strains for *L. lactis*. However, such mutants are usually slow acid producers and can also revert to phage sensitivity (51). Moreover, attempts to obtain *S. thermophilus* BIM were unsuccessful (15). Thus, the design of phage resistant *S. thermophilus* starter cultures has to rely mostly on genetic engineering. As many of the novel phage resistance mechanisms are based on the knowledge of the phage biology, a better understanding of bacteriophage genetics is needed. In addition, bacteriophages show an extremely high diversity. Knowledge of the phage population involved should therefore assist the design of efficient anti-phage systems.
1.1 Bacteriophages

Bacterial viruses are small nucleoprotein complexes, where nucleic acid is enclosed in a protein coat. In some cases a lipid envelope is present. The genetic material can be either DNA or RNA, single- or double-stranded. As they have no metabolic capabilities, phages need to parasitize their cellular host for reproduction. In principle, all cellular organisms are susceptible to virus infection.

Bacteriophages are probably the most abundant biological entities found in the environment. They are present in virtually all biotopes where bacteria are found. Researchers have reported phage counts 5-25 times higher than bacterial counts in marine water (35). More than 4500 different phages have been found in about 130 bacterial genera until now, both from *Eubacteria* and *Archaea* (1).

1.1.1 Classification of bacteriophages

Morphology and nucleic acid type are the main criteria employed to classify observed phages. Based on these characteristics phages have been classified in 13 families (Table 1).

<table>
<thead>
<tr>
<th>Morphotype</th>
<th>Shape</th>
<th>Nucleic acid</th>
<th>Family</th>
<th>Particulars</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 to A3</td>
<td>Tailed</td>
<td>dsDNA, linear</td>
<td>Myoviridae</td>
<td>Contractile tail</td>
</tr>
<tr>
<td>B1 to B3</td>
<td></td>
<td>Siphoviridae</td>
<td></td>
<td>Long, noncontractile tail</td>
</tr>
<tr>
<td>C1 to C3</td>
<td></td>
<td>Podoviridae</td>
<td></td>
<td>Short tail</td>
</tr>
<tr>
<td>D1</td>
<td>Polyhedral</td>
<td>ssDNA, circular</td>
<td>Microviridae</td>
<td>Conspicuous capsomers</td>
</tr>
<tr>
<td>D3</td>
<td></td>
<td>ssDNA, circular, S</td>
<td>Coticoviridae</td>
<td>Complex capsid, lipids</td>
</tr>
<tr>
<td>D4</td>
<td></td>
<td>dsDNA, linear</td>
<td>Tectiviridae</td>
<td>Double capsid, pseudotail, lipids</td>
</tr>
<tr>
<td>E1</td>
<td></td>
<td>ssRNA, linear</td>
<td>Leviridae</td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td></td>
<td>dsRNA, linear, seg.</td>
<td>Cystoviridae</td>
<td>Envelope, lipids</td>
</tr>
<tr>
<td>F1</td>
<td>Filamentous</td>
<td>ssDNA, circular</td>
<td>Inoviridae</td>
<td>a. Long filaments</td>
</tr>
<tr>
<td>F2</td>
<td></td>
<td></td>
<td></td>
<td>b. Short rods</td>
</tr>
<tr>
<td>F3</td>
<td></td>
<td>dsDNA, linear</td>
<td>Lipoviridae</td>
<td>Envelope, lipids</td>
</tr>
<tr>
<td>G1</td>
<td>Pleomorphic</td>
<td>dsDNA, circular, S</td>
<td>Plasmaviridae</td>
<td>Envelope, lipids, no capsid</td>
</tr>
<tr>
<td>G2</td>
<td></td>
<td>dsDNA, circular, S</td>
<td>Fuselloviridae</td>
<td>Same, lemon-shaped</td>
</tr>
</tbody>
</table>

Table 1

Tailed phages account for about 96% of all observed phages. They are differentiated into 3 families, *Myoviridae*, *Siphoviridae* and *Podoviridae* (Table 1). Each family is further differentiated by head morphology; small isometric, prolate
(elongated) or large prolate (morphotypes 1-3, respectively). Tailed phages have been found in all major prokaryotic groups. This ubiquity led to the hypothesis that tailed phages preceded the Eubacteria and Archaea differentiation. Apart from having a tail, these phages have other common characteristics; they all have linear dsDNA genomes and similarly organized head morphogenesis gene clusters. On that basis, Ackermann (2) proposed to group the families Myoviridae, Siphoviridae and Podoviridae in the order Caudovirales.

1.1.2 Evolution of viruses

Morphology and nucleic acid type are useful criteria for a rapid and easy classification of viruses, but care has to be taken when trying to attribute phylogenetic meaning to this subdivision. In fact, the comparison of bacteriophages by either heteroduplex analysis or sequencing has shown a mosaic structure in the pattern of similarities. When two phage strains from the same host species are compared, regions of high similarity are interspersed with regions of low or no sequence similarity. This is most likely the result of the exchange of genes or part of genes between phages. The comparison of the lambdoid phages led to the formulation of a theory of modular phage evolution (10): the product of evolution is not a given virus but a family (group) of interchangeable genetic elements (modules) each of which carries out a particular biological function. Evolution is perceived to act at the level of modules by selecting for a good execution of function and for optimal recombination events of functional genetic units. If the gene order is the same in all members of an interbreeding population, homologous or non-homologous recombination events could lead to increased diversity without giving rise to individuals that lack any essential function. Actually, phages seem to be very recombinogenic. Isolates from nature that are indistinguishable from the well-studied type isolates are rarely found (21). Viruses sharing the same gene pool can then differ widely in many characteristics, including morphology and host range, since these are aspects of individual modules. Good examples are bacteriophage λ and P22. These phages are morphologically distinct, λ is a siphovirus P22 a podovirus. Nevertheless, they have very similar genetic maps and can form many different viable hybrids (11;39).

Among bacteriophages four types of genetic exchange seem to occur (21): (1) homologous recombination between regions of very high sequence similarity, (2) recombination between regions of micro-homology such as promoters or terminators;
(3) illegitimate recombination; (4) site specific recombination such as the tail fiber gene inversion system of coliphages Mu and P1.

1.1.3 The life cycle of viruses

The first step in a viral infection is the attachment of the virus particle to the host. Each strain of bacteria has characteristic proteins, carbohydrates and lipopolysaccharides on its surface. Each of these molecules can act as a receptor for specific phages. Because of the specificity of the interactions, the diversity of the cell surface is one important factor in the limitation of the host range of bacteriophages. After the attachment, the phage genome enters the cell. At this stage there are different strategies used by viruses to exploit the host for their own proliferation. (1) In lytic infection the cell is reprogrammed by the phage to produce the maximal amount of progeny virions, which are then released by the lysis of the cell. (2) In chronic infection progeny virions are released by extrusion without killing the cell. (3) In temperate infection a bacteriophage maintains a stable genetic relationship with the host. The nucleic acid of the phage becomes part of the genome of the cell, either as a plasmid or integrated in the host chromosome (prophage), and can be reproduced along with the host for many generations. The prophage can be induced by events such as DNA damage or other physiological stress. This activation triggers then a lytic replication cycle.

Two well-characterized phages will be used as examples to illustrate the most representative phage groups: T4 as the prototype lytic phage and λ as the best known temperate phage (7).
1.1.3.1 Bacteriophage T4

The best-characterized example of a lytic phage is the *E. coli* bacteriophage T4. Morphologically it belongs to the *Myoviridae* family. It has an elongated head (78x111 nm), which contains its genome (166 kb, dsDNA) and a long contractile tail (113x20 nm) with six tail fibers attached to the baseplate. The study of the T4 life cycle has revealed the strategies used to exploit the bacterial host.

The tail makes the first contact with the host, first reversibly then irreversibly. The adsorption specificity of T4 is largely determined by gene 37, which encodes the 1026 aa residue distal tail fiber. The ~140 C-terminal residues of this protein contains determinants that can recognize either *E. coli* B type lipopolysaccharides or OmpC protein.

After the tail fibers and the tail tip have bound to the cell surface, the tail sheath contracts, the tail tube crosses the plasma membrane and the DNA is injected into the cytoplasm. Soon after DNA injection, the *E. coli* RNA polymerase start the synthesis of phage mRNA. This mRNA and all other early mRNAs (mRNA transcribed before phage DNA is made) encode the enzymes required to acquire the control over the cell. Synthesis of host DNA, RNA and proteins is stopped, and the cell is forced to produce viral constituents. Some early viral enzymes degrade host DNA to nucleotides. This halts gene expression and provides raw material for virus DNA synthesis. Other early proteins and a few phage head proteins injected along with phage DNA modify the bacterial RNA polymerase. The polymerase then recognizes promoters on viral DNA rather than bacterial promoters. Within 5 minutes, virus DNA synthesis starts. The DNA of T4 has a unique feature; all of its cytosine residues are replaced by 5-hydroxymethylcytosine (HMC). Two viral proteins are employed in its synthesis. This base does not normally occur in *E. coli*. HMC has the same hydrogen-bonding characteristics as normal cytosine, but it offers a reactive site where to attach one molecule of glucose. This difference has important biological functions, firstly the glucose moieties prevent restriction of phage DNA by host cell endonucleases, secondly it blocks the synthesis of cellular DNA because HMC can be used only by the phage polymerase and not the cellular DNA polymerase.

After DNA replication, late mRNA is produced. It directs the synthesis of two kinds of proteins, (1) phage structural proteins and proteins helping phage assembly, (2) proteins involved in cell lysis and phage release. All the proteins requested for phage assembly are synthesized simultaneously and used in independent assembly lines.
The tail, tail fibers and heads are made in parallel, then assembled. At the end of phage production the cell is lysed. One gene directs the synthesis of a lysozyme that attacks the cell wall peptidoglycan. Another phage protein damages the bacterial plasma membrane allowing the lysozyme to reach the cell wall. At 37°C the lysis takes place 22 minutes after infection. About 300 infectious particles are released in the medium.

1.1.3.2 Bacteriophage Lambda

The *E. coli* bacteriophage λ is a member of the *Siphoviridae* family. It has a small icosahedral head (60 nm) and a long non-contractile tail (150x17 nm). Its genome consists in a single linear dsDNA molecule of 48 kb with cohesive ends. Bacteriophage λ can start a normal lytic infection with subsequent host lysis and release of phage particles, or it can silence the expression of its lytic genes and integrate the genome of its host as a prophage. Once integrated, the prophage will be propagated along with the cells. The first steps of the lytic and temperate pathways are identical. First, the virus attaches reversibly to the host through the tip of the tail fiber, and then the end of the tail binds irreversibly a component of the maltose permeation system. The phage DNA can now enter the cell, probably by simple diffusion. RNA transcription begins promptly (immediate early transcription). Starting from the promoter P_L, the protein N is expressed, Cro is expressed from the P_R promoter (Fig. 1).

N is an antiterminator and allows the RNA polymerase to bypass the t_L and t_R1 terminators (delayed early transcription). Additional proteins are thus expressed; cII is one of the most important. This protein is a key factor for the establishment of lysogeny. It has two important functions. First, it activates the P_I promoter leading to the expression of the integrase. The integrase is responsible for the integration of the λ genome into the host chromosome by site specific recombination. Second, cII also activates the P_{RE} (repression establishment) promoter allowing the expression of the CI repressor. CI will bind to two operators, O_L and O_R, blocking both P_L and P_R promoters and activating P_{RM}. This will result in the silencing of all genes except for itself. But, CI is in competition with Cro (control of repressor and other things) for the same operators. Both repressors have different affinities for the binding sites of the
operators; CI binds preferentially to site 1, cro to site 3. Thus, while CI blocks P_L, P_R and activates P_RM, cro will block P_RM.

At least three bacterial and one phage factor do determine the stability and expression level of CII. According to the concentration of these factors at the moment of the infection, there will be more or less CI. If not enough CI protein is produced, cro concentration prevails, Cro will block P_M and CI concentration will further decrease. Now, N will be fully expressed allowing the transcription of all genes downstream of P_R, including Q. Q is a strong transcriptional activator, which will permit the expression of both structural and lysis genes. At this stage, lytic development will proceed unrestrained.

In the case there is enough CI to block the transcription at P_L and P_R, λ will lysogenize the cell. Activation of P_RM will provide enough CI to maintain the lysogeny and prevent the superinfection of the lysogen by the same phage. The prophage can be induced by environmental factors such as UV light or chemical mutagens that damage host DNA. These damages cause the RecA protein to act as a protease and cleave the CI repressor, inducing the transcription of the phage genes.

![Figure 1: Regulation of early λ transcription. Transcription of early λ genes is initiated at P_L and P_R, and is subject to repression by CI acting at Q_L and Q_R. Dashed lines indicate transcripts. The expanded diagrams show the relationship between PR, PL and PRM with the three repressor binding sites. CI binds preferentially to the sites 1, thus blocking P_L and P_R. Cro binds preferentially to the sites 3, blocking P_RM. Adapted from E. Birge, Bacterial and Bacteriophage genetics (7).](image-url)
1.1.4 Bacteriophages of lactic acid bacteria

Phage infection is a permanent threat in the dairy industry; therefore many dairy research institutes focused on the characterization of the phage population involved. A detailed classification of the various isolates has been developed. All viruses infecting LAB are tailed phages and contain linear dsDNA. Genome sizes generally range between 18 and 55 kb. According to the classification scheme of Bradley (19) phages of LABs are differentiated into three main classes; morphotype A (contractile tails), B (long, noncontractile tails) and C (short noncontractile tails) (48).

*L. lactis* phages have been divided into 10 species according to morphology and DNA/DNA hybridization. Some of the best characterized phages are listed in Table 2. Species 936, c2 and recently P335 are the predominant species (48). P335 is the only species to comprise both temperate and lytic phages (20;82).

<table>
<thead>
<tr>
<th>Host</th>
<th>Phage strain (species)</th>
<th>Morphology</th>
<th>Life style</th>
<th>Relevant information available</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. lactis</em></td>
<td>r1t (P335)</td>
<td>B1</td>
<td>temperate</td>
<td>Totally sequenced, genetic switch.</td>
<td>(74;92)</td>
</tr>
<tr>
<td></td>
<td>LC3 (P335)</td>
<td>B1</td>
<td>temperate</td>
<td>Lysis cassette, integrase.</td>
<td>(8;57)</td>
</tr>
<tr>
<td></td>
<td>sk1 (936)</td>
<td>B1</td>
<td>lytic</td>
<td>Totally sequenced, transcription map, lysis cassette and origin of replication.</td>
<td>(22;23)</td>
</tr>
<tr>
<td></td>
<td>Tuc 2009 (P335)</td>
<td>B1</td>
<td>temperate</td>
<td>Totally sequenced (unpublished), lysis, holin, major structural proteins, putative repressor.</td>
<td>(5;90;91)</td>
</tr>
<tr>
<td></td>
<td>BK5-T (BK5-T)</td>
<td>B1</td>
<td>temperate</td>
<td>Totally sequenced (unpublished), integrase, putative repressor.</td>
<td>(12-14)</td>
</tr>
<tr>
<td></td>
<td>TP901-1 (P335)</td>
<td>B1</td>
<td>temperate</td>
<td>Totally sequenced (unpublished), major structural proteins, integrase.</td>
<td>(24;25;47;62)</td>
</tr>
<tr>
<td></td>
<td>c2 (c2)</td>
<td>B2</td>
<td>lytic</td>
<td>Totally sequenced, structural proteins, replication origin, transcription map.</td>
<td>(59;60;96)</td>
</tr>
<tr>
<td><em>Lb. delbrueckii</em></td>
<td>LL-H</td>
<td>B1</td>
<td>lytic</td>
<td>Totally sequenced. Deletion derivative of a temperate phage.</td>
<td>(66-69)</td>
</tr>
<tr>
<td></td>
<td>mv4</td>
<td>B1</td>
<td>temperate</td>
<td><em>attP</em>, integrase and endolysin.</td>
<td>(9;32)</td>
</tr>
<tr>
<td></td>
<td>JCL 1032</td>
<td>B3</td>
<td>temperate</td>
<td>Contains homologous DNA segments with LL-H and mv4.</td>
<td>(34)</td>
</tr>
<tr>
<td><em>Lb. gasseri</em></td>
<td>adh</td>
<td>B1</td>
<td>temperate</td>
<td>Totally sequenced (unpublished). Integrase, lysis cassette.</td>
<td>(41;81)</td>
</tr>
<tr>
<td><em>Lb. casei</em></td>
<td>A2</td>
<td>B1</td>
<td>temperate</td>
<td>Integrase, DNA packaging proteins and genetic switch.</td>
<td>(4;37;38;53)</td>
</tr>
<tr>
<td><em>Lb. plantarum</em></td>
<td>g1e</td>
<td>B1</td>
<td>temperate</td>
<td>Totally sequenced, major structural proteins and lysis cassette.</td>
<td>(49;52;79)</td>
</tr>
</tbody>
</table>

Table 2: Examples of characterized phages of lactic acid bacteria
Lactobacillus phages generally belong to the Siphoviridae family and have small, isometric heads. A few phages with a prolate head have been also described. Myoviridae have been isolated in Lb. helveticus and Lb. plantarum. Phages of Lb. delbrueckii, the best investigated group, have been classified into four DNA homology groups, a, b, c and d (85). Group a is the most prevalent and contains both temperate and lytic phages. In general, the DNA homology between lytic and temperate phages is much higher for phages of Lactobacillus that for lactococcal phages.

A molecular characterization of several phages of lactic acid bacteria has been initiated. Efforts are mainly focusing on the analysis of the genes involved in phage integration, life cycle control and host lysis, since this can result in molecular tools such as integration vectors, inducible gene expression systems or controlled cell lysis. The complete sequences of 6 L. lactis and 3 Lactobacillus sp. have been determined (Table 2).

Recently, the increasing economical importance of S. thermophilus led an increased interest into their phages. Many phage isolates have been obtained from fermentation failures in the dairy industry resulting in several collections of S. thermophilus phages. S. thermophilus phages have been classified by host range, morphology, serotype, DNA homology, structural protein pattern and DNA packaging mechanism (6;18;54;56;76;80). All phages examined until now have the same basic features (Table 3); they consist of a small isometric head and a long, noncontractile tail. In some cases, a long tail fiber attached to the base of the tail can be seen. They are all classified as group B1 Siphoviridae. The size of their dsDNA genome has been calculated to range between 31 and 45 kb from restriction digests. DNA hybridization studies revealed that all phages belong to a single DNA homology group. All S. thermophilus bacteriophages shared at least one segment of their genome. Host range, on the other hand is very narrow. The majority of the phages are strain specific.

Analysis of the host range and serotype from a subset of phages isolated in French factories during the last 30 years allowed the separation of the phages in two
groups, lytic group I and lytic group II phages (18). Both groups had clearly distinct host ranges. Antibodies raised against a lytic group I phage could neutralize most lytic group I phages, but not lytic group II phages. Conversely, antibodies raised against a lytic group II phage inhibited other lytic group II phages, but not lytic group I phages. This was interpreted as a clear indication for differences in some structural genes. A subsequent analysis based on the protein patterns and DNA analysis could split all *S. thermophilus* phages in two groups (56); one group had two major proteins (32 and 26 kDa), the other group had three major proteins (41, 25 and 13 kDa). The difference in the protein pattern correlated very well with the packaging mechanism used, all phages with two major proteins were *cos*-site phages, and all phages with three major proteins were *pac*-site phages. Until now all lytic group I phages were *cos*-site phages, whereas lytic group II phages were members of the *pac*-site phages.

Both temperate and virulent phages have been isolated, but temperate phages represent only about 1% of all isolates. Dotplot tests showed extensive crosshybridisation between temperate and virulent phages (17).

### 1.2 Phage resistance mechanisms in LAB

#### 1.2.1 Natural resistance mechanisms

Molecular analysis provided insight into natural phage defense systems in *Lactococcus lactis*. These bacteria developed a variety of mechanisms to cope with phage infections, which act at various point of the phage life cycle. Major categories include: (1) adsorption inhibition, (2) block of phage DNA injection, (3) restriction/modification and (4) abortive infection mechanisms.

#### 1.2.1.1 Phage adsorption inhibition

The expression adsorption inhibition is used when the phage is unable to attach to the cell surface. Most of the spontaneous phage resistant mutants selected by phage challenge (*Bacteriophage Inensitive Mutants or BIM*) are modified in their adsorptive characteristics (50,58,63;93). The phenotype can result from many mechanisms, e.g. changes in the cell surface carbohydrate composition (40), masking of cell surface structures (86), or from alteration of specific phage protein receptors (72).

Interference with phage adsorption has also been linked with a number of lactococcal plasmids. Lactococcal plasmids pSK112 (28;86) and pCI528 (61) encode
proteins for the production of Galactosyl-containing lipoteichoic acid, and galactose and rhamnose-containing polymers, respectively. These substances possibly mask the phage receptor. In both cases, mild alkali treatment could remove this material restoring phage sensitivity.

The major problem with adsorption inhibition is that phages are not removed from the media and can attack a sensitive cell present in the same environment.

1.2.1.2 Block of phage DNA injection

Phage resistant strains have been isolated, which allow normal phage adsorption, but which apparently do not permit phage DNA injection (89;95). pNP40 of L. lactis MG1614 permits phage c2 to adsorb to the cell surface (42), but no phage DNA could be detected into the cell after adsorption. The resistance mechanisms were bypassed by electroporating phage DNA into the cell (43).

1.2.1.3 Abortive infection mechanisms

Abortive infection (Abi) is a phage resistance mechanism that acts after phage DNA injection, but is not restriction and modification. Abi results in cell death before phage progeny is released, thus limiting the propagation of new phage particles in the culture. Cells harboring Abi-resistance mechanisms are generally killed either by the Abi protein itself or by the modifications caused by the infection. In lactococci Abi mechanisms are usually encoded on plasmids (42). The phenotype of an abortive infection covers reduction in the efficiency of plaquing (EOP) ranging between 0.5 and < 10^{-6}; smaller plaques due to smaller burst size and a decrease in the efficiency of formation of centers of infection (ECOI) (3).

A few mechanisms of action have been investigated. One is the AbiA system encoded by pTR2030 (46) and pCI829 (26;27). It delays, as other Abi systems, phage DNA replication. It affects 3 lactococcal species (936, P335 and c2). A leucine zipper was identified in the AbiA protein. Site-directed mutagenesis experiments demonstrated its essential role in conferring phage resistance (3).

Phages can be isolated that overcome Abi mechanisms. Point mutations are the most common changes leading to Abi insensitivity (29). Transfers of bacterial genes into the phage genome leading to insensitivity have also been observed (70).
1.2.1.4 Restriction and modification

Many R/M systems have been identified in lactic acid bacteria, reviewed in (3). They operate at varying levels of efficiency (EOP of $10^{-1} - 10^6$). The problem with this system is that it allows some phages to escape restriction resulting in modified progeny virions. The modified phages can then infect a second host bearing an identical R/M system without being restricted (EOP=1.0). However, R/M systems can be extremely useful when combined with Abi defenses. The restriction/modification system provides the first level of protection and promotes cell survival. If the phage escapes restriction, Abi systems will kill the cell and prevent the development of the modified phage.

Hill et al. described a lactococcal phage being able to overcome the phage resistance mechanisms encoded by plasmid pTR2030 (45). This plasmid encodes for two different phage insensitivity mechanisms, a R/M system and the abortive infection system AbiA. To escape the host restriction activity the bacteriophage acquired the functional domain from the type II A methylase encoded by the plasmid. Insensitivity to the second phage-resistance mechanism seems to be due to its phage origin of replication, which apparently differs from those targeted by the AbiA system.

Moineau et al. (71) succeeded in protecting several industrial strains of *S. thermophilus* from phage infection by the introduction of a lactococcal R/M system.

1.2.2 Novel phage resistance mechanisms

In addition to natural resistance mechanisms, resistance mechanisms have been constructed in the laboratory that are not encountered in LAB (Per and antisense RNA). Per is an acronym for Phage Encoded Resistance and relies on the phage inhibitory effect of phage DNA when cloned on a plasmid. For both lactococcal and streptococcal phages, the presence of phage origin of replication (ori) on a high copy number plasmid, can interfere with phage replication in an Abi-like manner (33,44,77). Phage replication proteins are bound to the ori on the high copy instead to the physiological ori on the phage DNA (33).

On theoretical grounds antisense RNA appears a promising approach, but recent experiments had little success (94). Apparently, the choice of the target gene is critical. It should be a crucial phage gene, conserved in a wide range of phages and expressed at low enough levels to permit effective inhibition.
Recently, a further protection system has been constructed for *L. lactis*. A bacteriophage inducible promoter was used to trigger the expression of a bacterial suicide system (31). The lethal genes consisted of the restriction cassette LlaIR deprived of its corresponding methylase. The EOP was lowered to $10^{-5}$. However, mutant phages developed that were less sensitive (EOP of 0.4). Changes were found in their transcription activator, which decreased expression of the suicide system (30). Finally, the loss of the phage receptor in *L. lactis* C2 can completely block the infection by the c2 phage (36).

1.3 Outline of the thesis

1.3.1 Phage diversity

The first aim of my thesis was the definition of phage diversity by comparative genomics of *S. thermophilus* phages by sequencing the genomes of 3 lytic phages. The phage genomes underwent pairwise alignments by including two temperate and one further virulent *S. thermophilus* phages to localize areas of difference (Table 4). The phages were chosen to allow a pairwise alignment of phages differing in specific phenotypes: (A) Life style (*Sfi19* vs. *Sfi21; Sfi11* vs. *O1205*). (B) DNA packaging and structural proteins, (*cos-phage Sfi19* vs. *pac-phage Sfi11*). (C) Host range (*Sfi19* vs. *DT1; Sfi11* vs. *O1205*). (D) Sensitivity vs. resistance to *Sfi21* prophage mediated superinfection exclusion mechanisms (*Sfi19* vs. *Sfi18*). The aim of these comparisons was to correlate phenotypic differences with genotypic differences. Results are reported in chapters 2 to 4.

<table>
<thead>
<tr>
<th>Phage</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sfi11</em></td>
<td>Lytic</td>
<td>Pac-phage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to phage-ori mediated resistance mechanisms</td>
</tr>
<tr>
<td><em>Sfi18</em></td>
<td>Lytic</td>
<td>Cos-phage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to phage-ori mediated resistance mechanisms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Submitted to Sfi21-prophage control</td>
</tr>
<tr>
<td><em>Sfi19</em></td>
<td>Lytic</td>
<td>Cos-phage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not sensitive to phage-ori mediated resistance mechanisms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not submitted to Sfi21-prophage control</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sensitive to phage-ori mediated resistance mechanisms</td>
</tr>
<tr>
<td><em>DT1</em></td>
<td>Lytic</td>
<td>Cos-phage</td>
</tr>
<tr>
<td><em>O1205</em></td>
<td>Temperate</td>
<td>Pac-phage</td>
</tr>
</tbody>
</table>

Table 4: Sequenced *S. thermophilus* phages
The evolutionary relationship of *S. thermophilus* phages to *Siphoviridae* from low GC content Gram-positive bacteria is explored in chapters 3 and 5.

### 1.3.2 Phage resistant starter strains

The second aim of my thesis was the establishment of phage-resistant *S. thermophilus* starter strains. Since plasmids and natural phage resistance mechanisms are rare in *S. thermophilus* we turned our attention to phage DNA segments that inhibited phage infection of strains that contained these elements on high copy number plasmids. We identified the phage replication origin as an inhibitory DNA element. The results of these experiments are described in chapter 6.

A second approach to establish phage resistant starters relied on the anti-phage properties of lysogenic starters. The Sfi21 prophage encodes potent superinfection immunity genes. The drawback is that lysogens continuously release phage particles in the media. These phages could then infect sensitive strains. Therefore, we created deletions in the prophage by targeted insertion mutagenesis to obtain lysogens, which retain superinfection immunity, but do not produce infectious particle. Finally, we hypothesized that *S. thermophilus* phages need host genes for their development and that it should be possible to obtain phage resistant strains of *S. thermophilus* by random insertion mutagenesis of the host genome. If these genes are nonessential for bacterial growth in milk, such insertion mutants can yield phage resistant starters. The results of the insertion mutagenesis experiments are reported in chapter 7.

Reference List


72. Morelli, L. Personal Communication


82. Sijtsma, L., J.T.M. Wouters, and K.J. Hellingwerf. 1990. Isolation and characterization of lipoteichoic acid, a cell envelope component involved in


2. RESULTS

2.1 The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison of cos-site phages Sfi19 and Sfi21

(Virology 260, 232-243 (1999))
The Genetic Relationship between Virulent and Temperate *Streptococcus thermophilus* Bacteriophages: Whole Genome Comparison of cos-Site Phages Sfi19 and Sfi21

Sacha Lucchini, Frank Desiere, and Harald Brüssow

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

Received December 29, 1998; returned to author for revision May 1, 1999; accepted May 17, 1999

The virulent cos-site *Streptococcus thermophilus* bacteriophage Sfi19 has a 37,392-bp-long genome consisting of 44 open reading frames all encoded on the same DNA strand. The genome of the temperate cos-site *S. thermophilus* phage Sfi21 is 3.3 kb longer (40,740 bp, 53 orfs). Both genomes are very similarly organized and differed mainly by gene deletion and DNA rearrangement events in the lysogeny module; gene replacement, duplication, and deletion events in the DNA replication module, and numerous point mutations. The level of point mutations varied from <1% (lysis and DNA replication modules) to >15% (DNA packaging and head morphogenesis modules). A dotplot analysis showed nearly a straight line over the left 25 kb of their genomes. Over the right genome half, a more variable dotplot pattern was observed. The entire lysogeny module from Sfi21 comprising 12 genes was replaced by 7 orfs in Sfi19, six showed similarity with genes from temperate pac-site *S. thermophilus* phages. None of the genes implicated in the establishment of the lysogenic state (integrase, superinfection immunity repressor) or remnants of it were conserved in Sfi19, while a Cro-like repressor was detected. Downstream of the highly conserved DNA replication module 11 and 13 orfs were found in Sfi19 and Sfi21, respectively. Two orfs from Sfi21 were replaced by a different gene and a duplication of the phage origin of replication in Sfi19; a further orf was only found in Sfi21. All other orfs from this region, which included a second putative phage repressor, were closely related between both phages. Two noncoding regions of Sfi19 showed sequence similarity to pST1, a small cryptic plasmid of *S. thermophilus*. © 1999 Academic Press

INTRODUCTION

Bacteriophages can be classified according to their lifestyle into virulent and temperate phages. Somewhat generalized temperate phages have the choice to multiply on their host cells leading to the lysis of the cell or to integrate the phage genome into the bacterial chromosome to become a prophage. Prophages are propagated passively by the replication machinery of the bacterial cell. Virulent phages are unable to lysogenize their host cells and have to rely on productive infection cycles for their propagation. Because both lifestyles have different requirements with respect to genome organization, virulent and temperate phages from many bacterial hosts, e.g., *Escherichia coli*, share few, if any genetic relationships. This separation into two fundamentally different classes of bacteriophages seems to be less clear in lactic acid bacteria (lactic streptococci, lactococci, lactobacilli) have recently become a focus of research in food microbiology (Josephsen and Neve, 1998). All *S. thermophilus* phages isolated until now belong to a single class of viruses, isometric headed phages with a long noncontractile tail (*Siphoviridae*) containing ~40 kb of linear double-stranded DNA that showed at least some cross-hybridization (Brüssow et al., 1994; Brüssow and Bruttin, 1995). Ecologically *S. thermophilus* phages are reasonably well characterized. More than 200 distinct *S. thermophilus* phage isolates have been described by several laboratories representing >30 years of systematic surveys including longitudinal ecological surveys (Neve et al., 1989; Prevots et al., 1989; Brüssow et al., 1994; Bruttin et al., 1997a; Le Marrec et al., 1997). Temperate and virulent phages have been detected in this phage group, and they cross-hybridized to a variable extent (Brüssow and Bruttin, 1995). Possibly as an adaptive response to the competition by virulent phages, the temperate *S. thermophilus* phage Sfi21 has developed a powerful double immunity system. The Sfi21 immunity system consists of a phage repressor protecting the lysogen against superinfection with the homologous temperate phage and its derivatives (Foley et al., unpublished data) and a second protein protecting the lysogen from superinfection by the majority of the virulent phages (Bruttin et al., 1997b). However, some virulent phages...
were still able to multiply on the lysogen (Bruttin et al., 1997b). Within this group of phages two types could be differentiated. One type is represented by phage S17; according to DNA–DNA hybridization experiments, S17 DNA is only weakly related to Sf121 DNA (Brüssow and Bruttin, 1995). The second type is represented by Sf19. According to DNA–DNA hybridization experiments Sf19 DNA is closely related to Sf121 DNA over the tail morphology, host lysis, and DNA replication modules (Brüssow and Bruttin, 1995); this was confirmed by sequence analysis (Desiere et al., 1998; Foley et al., 1998). However, Sf19, like the majority of the lytic group I (cos-site phages) S. thermophilus phages, failed to cross-hybridize with the DNA packaging, head morphogenesis and lysogeny modules from Sf121 under stringent hybridization conditions (Brüssow and Bruttin, 1995). To further our understanding of the genetic relationship between temperate and virulent S. thermophilus phages, we completed the sequencing of Sf19 and Sf121 and provide here their whole genome comparison. After the recent comparison of the virulent D29 and temperate L5 mycobacteriophage genomes (Ford et al, 1998), this is to our knowledge only the second whole genome comparison of a genetically closely related virulent/temperate phage pair.

RESULTS

Overall comparison of the S. thermophilus Sf19 and Sf121 phage genomes

The temperate cos-site S. thermophilus phage Sf121 has a 40,740-bp-long genome consisting of 53 open reading frames (ORFs) longer than 50 codons. Two ORFs were of unusual length for prokaryotic genomes in that they exceed 1200 codons. All but four ORFs started with an ATG start codon (Fig. 1A). Twenty-five ORFs were preceded by a standard ribosomal binding site for S. thermophilus (GAG) (Guedon et al., 1995) in appropriate spacing to the initiation codon; three isolated cases of potential translational coupling of two ORFs were observed (Fig. 1A). Factor-independent terminator structures were identified downstream of the major tail protein gene and the lysin gene. The phage DNA was densely packed with ORFs (only six noncoding regions longer than 200 bp were detected, the longest was 450 bp in length (Fig. 1A).

The virulent cos-site S. thermophilus phage Sf19 has a 37,392-bp-long genome, 3.3 kb shorter than Sf121, consisting of 44 open reading frames longer than 50 codons (Fig. 1B). Remarkably, all ORFs are encoded on the same DNA strand in Sf19, while in Sf121 four lysogeny-related genes were encoded on the opposite strand. Both genomes are very similarly organized with respect to modular structure, number, and length of the ORFs, location of intergenic regions, and use of alternative start codons and terminators (Fig. 1). Major differences in genome organization were detected in the lysogeny and DNA replication module (Fig. 1, blue shading).

The similarity between both phages extended to the nucleotide sequence level. A dotplot analysis showed nearly a straight line over the left 25 kb of their genomes (Fig. 2). Over the right genome half, a more variable dotplot pattern was observed. The entire lysogeny module from Sf121 consisting of 12 ORFs (Bruttin et al., 1997b) was replaced by a sequence-unrelated cluster of 7 ORFs in Sf19. In the rightmost part of the compared genome, the dotplot revealed a DNA duplication in Sf19 and two nonalignments (Fig. 2).

The dotplot analysis does not provide a quantitative analysis of the level of nucleotide sequence similarity between corresponding genome regions of the two phages. To depict the differences quantitatively, a plot-similarity analysis was done (Fig. 3). It demonstrated that the level of sequence similarity varied over different regions of the compared genomes. Over the DNA packaging and head morphogenesis modules a relatively high degree of diversification was observed. Both phages differed by ~15% at the basepair level. Over the tail morphogenesis module sequence variation ranged from <1% to >25%; substantial variation was seen even between different regions of the same gene. Except for one insertion the lysis genes were highly conserved, this was also the case for the DNA replication module. The right end of the phage genomes showed an alternating pattern of nonconservation and conservation of nucleotide sequence. Three regions of the Sf19 genome that have not been described previously will be analyzed in detail below.

The DNA packaging and head morphogenesis modules of Sf19

The left end of the two phage genomes showed an identical genetic organization. However, at the nucleotide sequence level, important differences were observed. The noncoding region around the cohesive ends differed substantially between the two phages [64 changes in 189 nt positions (34%), 3 of 15 cohesive positions from the cohesive ends of Sf121 were changed in Sf19] (Fig. 4). Most of the direct and inverted repeats from Sf121 (Desiere et al., 1999) were not conserved in Sf19. Notably, the noncoding region from Sf19 showed highly significant similarity to the S. thermophilus plasmid pST1 (94% bp identity over two 50-bp-long DNA segments, \( P = 10^{-14} \), Janzen et al., 1992), whereas this was not the case for the noncoding region from \( \phi \)Sf121 (\( P = 0.015 \)) (Fig. 4). Previous bioinformatic analysis has suggested that ORF 152 and ORF 623 encode the small and large subunit terminase, respectively (Desiere et al., 1999). The putative large subunit terminase was well conserved between both phages (93% aa identity with evenly distributed aa changes), while the ORF 152 gp...
FIG. 1. Prediction of open reading frames in the complete genomes of the temperate S. thermophilus phage Sfi21 (A) and the virulent S. thermophilus phage Sfi19 (B). The ORFs are marked with their length in codon numbers. ORFs of identical length were differentiated by letters. Probable gene functions identified by bioinformatic analysis were noted. Gene functions identified by biological experiments were underlined. The phage genomes were divided into functional units according to previous bioinformatic and comparative evolutionary analysis. Genes belonging to the same unit have the same color. Gray filling indicates lack of information about the function of the ORF. ORFs preceded by a potential RBS 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. Overlap of start and stop codon is indicated with a triangle. Areas of blue shading connect regions of major sequence difference between the two phage genomes. The location of the cos-site is indicated.
showed only 84% aa identity and the changes were concentrated in the C terminus of both proteins. The Sfi19 genes encoding the putative portal protein, protease and major head protein differed by 13–15% at the bp level from the corresponding genes of Sfi21 (Table 1). The bp changes were evenly distributed and mainly found in the third base position of the codons.

The lysogeny replacement module in Sfi19

The sequence similarity between the left genome halves of both phages ended after orf 111/110 (84% bp identity), located 1 orf downstream of the lysin gene, and resumed over orf 157 (98% bp identity) (Fig. 3), flanking the highly conserved DNA replication module (Desiere et al., 1997). None of the 12 genes from the lysogeny module of Sfi21 (orf 74 to orf 87) showed bp (or deduced aa) sequence similarity with the seven predicted orfs in Sfi19, located between orf 111 and orf 157 (Fig. 5). We called this DNA segment a lysogeny-replacement module. Interestingly, the left half of the lysogeny replacement module from Sfi19 was identical to a DNA segment found ~1 kb downstream of the integrase gene in the temperate pac-site S. thermophilus phages TP-I34 (Neve et al., 1998). Only 2 bp differences were observed over this region: 1 bp change led to a different stop codon (orf 183 vs. orf 103) and 1 bp indel (insertion/deletion) led to the prediction of two orfs in Sfi19 instead of one orf 83 in TP-I34 (Fig. 5).

Over the central part of the lysogeny replacement module, we detected nucleotide sequence similarity with DNA segments of the temperate pac-site S. thermophilus phage 01205 (Stanley et al., 1997). Several points are noteworthy. First, this Sfi19 DNA segment showed sequence similarity with four noncontiguous DNA seg-
FIG. 3. Plot-similarity analysis of the sequence differences between the genomes of Sfi21 and Sfi19. The lower x axis gives the map position in basepairs as defined in Fig. 1. The upper x axis gives the gene map of Sfi21 for orientation. The y axis gives the similarity score (1.0, 0.5 and 0.0 correspond to 100, 50, and 0% sequence identity). The window was 100 bp.

ments of O1205 located, respectively, 1.6 kb downstream and 1.5 kb upstream of the phage integrase gene (Fig. 5). Second, some sequences were highly similar (e.g., the 3' half from orf 88 showed 99% bp identity to orf 53 from O1205), whereas others were at the borderline of nt sequence similarity (orf 69 showed 45% bp identity with orf 5 from O1205). Third, orf 69 gp showed similarity to Cro-like repressors both from the temperate S. thermophilus phage O1205 (42% aa identity) and the temperate Lactococcus lactis phage TP901-1 (Madsen and Hammer, 1998) (48% aa identity; Fig. 6). Fourth, orf 88 from Sfi19 is apparently the product of a gene fusion event between two orfs separated in O1205.

The right part of the lysogeny replacement module is occupied by orf 229, which had no counterpart in temperate S. thermophilus phages. The predicted orf contained a DNA sequence of low complexity: An 18-bp-long sequence was repeated 18 times. The consensus sequence was: CT(A1/G1) TGGA1TA1CGCA1XTX1GGT (the subscripts give the number of positions identical to the specified nt, nt in bold were identical in all 18 repeats). The same sequence could also be read in a different reading frame (orf 129) or on the opposite strand (orf 114), resulting in predicted proteins with GNAxVx, GSxVTx, or PxRYHR peptide repeats. The latter showed similarity to the sra gene product of Yersinia pseudotuberculosis (score 105, P = 10^{-7}) located in a possible pathogenicity island. The noncoding regions preceding and following orf 229 as well as the region preceding orf 69 showed sequence similarity to noncoding regions in O1205 (notably including part of the hypothetical genetic switch region of O1205).

The right genome ends of Sfi19 and Sfi21

The highly conserved part of the DNA replication modules from S. thermophilus phages extended from orf 157 to orf 143 when a large panel of phage isolates was investigated (Desiere et al., 1997). Eleven orfs were found between orf 143 and the right cohesive end of the Sfi19 genome, whereas 13 orfs were detected over this region in the temperate cos-site phage Sfi21. Interestingly, a similarly organized genome region was found in the pac-site temperate S. thermophilus phage O1205, consisting of 10 orfs between orf 143 and the pac-site (Fig. 7). The similarity extended to the nucleotide sequence level: orf 106 to orf 51 were identical in Sfi19 and Sfi21 and the corresponding orfs in O1205 differed by 11–14% bp changes. None of the predicted four proteins showed a database match. Downstream of this region the three phages differed. The temperate cos-site phage Sfi21 showed two orfs, orf 61 and orf 130. Orf 130 gp shared significant sequence similarity with gp 64 from mycobacteriophage L5 (40% aa identity, P = 10^{-25}). Gene
COMPARISON OF cos-SITE PHAGES SfI19 AND SfI21

SfI21, SfI19, 1326 TTTACCGAGATACCGGAGCATCAGCAGTCTTCCGGTTTATCGAGTGGTATGCTTTGAGG
R1 1301 AARCCAGAGATACCGGAGCATTACGAGTCTTCCGGTTTATCGAGTGGTATGCTTTGAGG
R2 stop 125 R1 R2
SfI21, SfI19, 1386 AAAAAAGAGAGACGTCGTTAAACACCTCTAAAAAGACCATAGCAAGCCTTTAGAACGTTTG
R1 1361 R1
SfI21, SfI19, 358 CGAAGGAGAGACGTCGTTAAACACCTCTAAAAAGACCATAGCAAGCCTTTAGAACGTTTG
R1 1301 R1
SfI21, SfI19, stop 170 R1 R2
SfI21, 1446 TAGTATTTTTCCCCGTGCCAAATTTTAAAACTACCCCCGCCCCGTAATCTAAAAAGGG
R1 1421 R1
SfI21, 1506 GAGCCGCCACAPGGTGTS
R1 1481 R1
SfI21, 1566 C-TATACCAAACGGAGAGGAGTAATGATGApTGfGTTAAGAATCCATACTTCAAGCAGAAT
R1 1541 R1
SfI21, 1625 tcggggcgcttaccaaecgaccctcc
R1 1600 R1

FIG. 4. Comparison of the cos-region from SfI21 and SfI19 phages. The stop codon from orf 175 and the alternative start codons from orf 152 are boxed. The 15 nucleotide 3' overhang from the cohesive ends is marked by a large box; the region of twofold hyphenated rotational symmetry in SfI21 is marked by a smaller box. Direct repeats (R) are marked by lines, inverted repeats (IR) by arrows. Differences between SfI21 and SfI19 are shaded. The SfI19 regions showing 94% bp identity with S. thermophilus plasmid pST1 (Janzen et al., 1992) are bracketed. 64 is found downstream of a putative DNA replication module from phage L5 (Hatfull and Sarkis, 1993). The virulent cos-site phage SfI19 showed at this position a gene replacement by orf 67. Orf 67 gp resembled the N terminus of orf 22 gp from the L. lactis phage r11 (P = 10^-9). Orf 22 was located downstream of the putative r11 DNA replication module (van Sinderen et al., 1995). Intriguingly, orf 67 was followed by a nearly perfect repeat of the SfI19 origin of replication located downstream of the putative primase gene (FoChe et al., 1998) (Fig. 8). After this region of diversification, the three phages were closely related over the next three orfs. The orf 19 gp from O1205 showed similarity to orf 192 gp from SfI19 over the N- and C-terminal third, while the central part of orf 19 gp showed similarity to orf 19 gp from L. lactis phage r11 (P = 10^-9) (Fig. 7). The corresponding SfI21 gene differed by only 6-bp changes. A similar mosaic-like pattern of relatedness was observed for the next orf (orf 170, 166, 20, Fig. 7). Over the C-terminal halves, the three predicted phage proteins were nearly identical and

TABLE 1

Comparison of the DNA Packaging and Head Morphogenesis Modules in the cos-Site S. thermophilus Phages SfI21 and SfI19

<table>
<thead>
<tr>
<th>Orf SfI21</th>
<th>Orf SfI19</th>
<th>bp</th>
<th>aa</th>
<th>Ks</th>
<th>Ka</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>132</td>
<td>132</td>
<td>14</td>
<td>14</td>
<td>0.37</td>
<td>0.08</td>
<td>D: 3</td>
</tr>
<tr>
<td>176</td>
<td>170</td>
<td>20</td>
<td>18</td>
<td>0.71</td>
<td>0.11</td>
<td>terminase</td>
</tr>
<tr>
<td>152</td>
<td>150</td>
<td>23</td>
<td>16</td>
<td>1.08</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>623</td>
<td>621</td>
<td>14</td>
<td>7</td>
<td>0.25</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>59</td>
<td>15</td>
<td>12</td>
<td>0.72</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>386</td>
<td>13</td>
<td>10</td>
<td>0.63</td>
<td>0.05</td>
<td>D: 6/portal protein</td>
</tr>
<tr>
<td>221</td>
<td>221</td>
<td>15</td>
<td>15</td>
<td>0.60</td>
<td>0.08</td>
<td>ClpP protease</td>
</tr>
<tr>
<td>397</td>
<td>397</td>
<td>14</td>
<td>9</td>
<td>0.52</td>
<td>0.05</td>
<td>major head protein</td>
</tr>
</tbody>
</table>

Note. Orf, the corresponding orfs in the two cos-site phages are indicated by their codon lengths; the orfs are enumerated according to their position on the genome map (see Fig. 1). bp, percentage of basepair differences between the aligned DNA sequences from SfI21 and SfI19. aa, percentage of amino acid differences between the aligned protein sequences from SfI19 and SfI11. Ks, number of substitutions per synonymous site. Ka, number of substitutions per nonsynonymous site. Comments, D deletions with the number of nucleotides affected, arbitrarily defined by reference to the genome of SfI21 phage/attribute of genes.
FIG. 5. Comparison of the lysogeny replacement module in the virulent cos-site S. thermophilus phage Sfi19 with the lysogeny modules from the temperate pac-site S. thermophilus phages O1205 and TPJ-34. The predicted open reading frames are annotated with their length in codon numbers and in the case of O1205 the numbering system described in Stanley et al. (1997). Genes whose functions were identified by bioinformatic analysis are given with their genetic symbols. Regions of greater than 70% bp identity are connected by shading.

showed a significant similarity to the arpR gene product of Enterococcus hirae (P = 10^{-12}, 38% aa identity). Over the N-terminal halves, the Sfi19 and Sfi21 proteins, but not the O1205 protein, demonstrated significant similarity with possible transcriptional regulators from Bacillus subtilis and Methanococcus janaschii as well as repressor proteins from a number of phages infecting a large range of bacterial hosts (Streptococcus, Lactococcus, Bacillus, Lactobacillus, Escherichia). The similarity covered the helix-turn-helix DNA binding region of the well-characterized coliphage 434 repressor (Fig. 9A). The next orf (orfs 98, 114, 21, Fig. 7) differed between the three phages only by 24–43 point mutations.

The following genome position was again variable. Sfi21 showed an orf 152 without database link, this orf was replaced by an unrelated orf 22 in O1205 that also lacked database matches. No gene was found at this position in Sfi19. Very similar proteins without database match were predicted for the next gene in all three phages (orf 235, 236, Fig. 7). The Sfi21 protein was more closely related to the O1205 protein (9% aa differences) than to the Sfi19 protein (19% aa differences). Downstream of orf 235 two related genes were found in the two cos-site phages (Table 1). Orf 175 showed similarity to cos-site associated genes from a Lactobacillus and Staphylococcus phage (Desiere et al., 1999). The pac-site phage O1205 showed at the corresponding genome position orf 24 whose gene product showed significant similarity to cos-site associated genes from a Lactobacillus and Staphylococcus phage (Desiere et al., 1999). The pac-site phage O1205 showed at the corresponding genome position orf 24 whose gene product showed significant similarity to the ArpU protein from Enterococcus hirae (Stanley et al., 1997). ArpU showed significant similarity with orf 31 gp from Lactobacillus phage g1e. Orf 31 is located downstream of the g1e DNA replication module (Kodaira et al., 1997). arpR and arpU belong to the same bacterial operon implicated in muramidase export (Del Mar Lleo et al., 1995). Database searches identified in this operon a further gene which encoded a protein with matches to several bacteriophages from Lactococcus lactis. Due to its many links to phage genes, the arp operon might be part of a prophage from E. hirae.

**DISCUSSION**

The possible temperate origin of new virulent phages isolated in the dairy industry has been a subject of speculation for years. This question is not of purely academic interest because successful phage control programs in industrial milk fermentation depend on knowledge of the source of new phages. In earlier reports, when only lactococcal phages were reasonably well investigated, it has been suggested that temperate dairy phages do not contribute significantly to the emergence of new virulent phages, because virulent and temperate Lactococcus phages are not significantly homologous (Jarvis, 1984a,b; Relano et al., 1987). The present and previous works from our laboratory showed that this conclusion cannot be generalized to other lactic acid bacteria. In lactic streptococci, virulent and temperate phages showed close genetic relationships (Desiere et al., 1998; Lucchini et al., 1998). In addition, we described how the temperate phage Sfi21 gave rise to a virulent phage by a single (Bruttin and Brüssow, 1996), integrase-mediated (Bruttin et al., 1997c) deletion event in the lysogeny module. Similar deletion events occurred under selective laboratory conditions in coliphage λ.

**FIG. 6.** Alignment of orf 69 gp from Sfi19 with the Cro-like repressor proteins from the pac-site temperate Lactococcus lactis phage TP901-1 and S. thermophilus O1205. Amino acid positions shared by two phage proteins are shaded.
COMPARISON OF cos SITE PHAGES Sf19 AND Sf21

Mycobacteriophage L5

gene 64

DNA-binding domain of several phage repressors

FIG. 7 Alignment of the gene maps from the right genome ends of S. thermophilus phages Sf19, Sf21, and O1205. The predicted open reading frames are annotated with their length in codons. The end of the phage genomes is at the right and is given by the cos- or pac-site. Regions of >85% bp sequence identity are connected by shading. The shaded wedges marked regions of significant amino acid sequence similarity to the indicated entries from the database. For O1205, the numbering system described in Stanley et al. (1997) is given. The DNA sequences that were specific to the indicated phage were marked by a different type of striping in the respective orfs.

(Davis and Parkinson, 1971). In S. thermophilus, this deletion process is not limited to specific laboratory selection because it was also observed in the virulent field isolate S3 (Bruttin and Brüssow, 1996). Despite all similarity in genome organization, Sf19 is not a simple deletion derivative of Sf21 as S3. A detailed comparison of both genomes showed that they were punctuated by insertions, deletions, duplications, and substitutions of DNA as well as a large number of point mutations mainly over the structural gene cluster. Because we do not possess a molecular clock for phage genomes, we cannot approximate what time interval separates the structural gene clusters from Sf19 and Sf21. Notably, ecological surveys of S. thermophilus phages covering an ob-

FIG. 8 Alignment of the putative origin of phage replication from Sf19 found downstream of orf 504 (orf1) and downstream of orf 67 (orf2). DNA repeat identified in orf1 are indicated by different types of lines and arrows. The bended arrows located the regions of significant sequence similarity (90 and 87% bp identity) with the S. thermophilus plasmid pST1 (Janzen et al., 1982).
servation period of several years in a highly dynamic industrial environment did not identify an accumulation of point mutations (Bruttin et al., 1997a). Despite a distinctively different genome organization between phages from low and high GC content gram-positive bacteria, the processes that differentiate the virulent/temperate mycobacteriophage pair (Ford et al., 1998) were fundamentally similar to those that differentiate the streptococcal phage pair. However, a marked difference was seen with respect to the lysogeny-related genes. In the virulent mycobacteriophage D29, a 3.6-kb deletion removed the repressor gene that accounted for the inability of D29 to form lysogens. An apparently functional integrase gene was still found on the genome of the virulent phage (Ford et al., 1998). In fact, when the L5 repressor was provided from an extrachromosomally plasmid, the virulent phage D29 could lysogenize its host. In the two lactic streptococcal phages, the virulent phage could conceivably be derived from the closest temperate phage by a precise deletion of the entire Sfi21-like lysogeny module combined with a complicated recombination process with at least two distinct pac-site temperate S. thermophilus phages. Two observations are noteworthy. First, alignments of the lysogeny modules from different temperate S. thermophilus phages suggested that this region has, in contrast to other genome regions from this phage group, been shaped by numerous recombination processes (Neve et al., 1998; Lucchini et al., 1999). Second, the lysogeny module can be divided into two parts: a group of essential lysogeny genes encoded on the opposite strand (integrase, superinfection immunity, repressor) and a group of genes downstream and upstream of the core lysogeny genes that are on the same strand as the rest of the phage genome. Only the latter genes from the two pac-site phages contributed to the genome of the virulent S. thermophilus phage.

Close genetic relationships between temperate and virulent phages were also described in Lactococcus delbrueckii. An alignment of the temperate phage mV4 with the virulent phage LL-H (Vasala et al., 1993; Mikkonen et al., 1996) suggested DNA deletion events over the lysogeny module followed by complex DNA rearrangements as the origin of the virulent phage. In contrast to streptococcal phages, the virulent phage LL-H still possessed a truncated integrase gene and two sites homologous to the mV4 attP site (Mikkonen et al., 1996). In lactic Streptococcus phages, the separation of virulent and temperate phages lines seems therefore to be more evolved because no remnants of essential lysogeny genes were detected. The changes in the replication module could possibly indicate a further adaptation to the virulent lifestyle. However, we need sequence data from further virulent phages and biochemical experiments concerning the role of the second origin to substantiate these suggestions.

Classically, dairy microbiologists have perceived the genetic conflict between the phage and the host lactic acid bacterium as the major driving force for the evolution of the phage genomes. There are good arguments for this view in the field of lactococcal phages (Garvey et al., 1995). Lactococci possess numerous plasmids containing phage resistance functions (Allison and Klaenhammer, 1998), e.g., restriction systems (Garvey et al., 1995). Notably, many lactococcal phages showed a conspicuous avoidance of restriction sites in their genomes (Moineau et al., 1993) and phages have been isolated...
that had acquired bacterial genes to undermine cellular resistance functions based on restriction systems (Moineau et al., 1994). Little evidence for this type of phage-host conflict has been documented for S. thermophilus (Larbi et al., 1992). Plasmids are conspicuously few in S. thermophilus, and few have shown an anti-phage function (O'Sullivan et al., 1999; L. Morelli, personal communication). We documented even a possible indication for coadaptation of the bacterial and phage genome: the bacterial attachment site (attB) contained a DNA sequence that complemented the C terminus of the phage integrase that would otherwise have been disrupted by the prophage integration process (Bruttin et al., 1997c). The genome comparison of Sfi21 and Sfi19 suggested that the genetic conflict between temperate and virulent S. thermophilus phages might be a further, if not the major driving force for phage genome evolution in this group of bacteria. Once again, we need additional temperate/virulent S. thermophilus phage genome comparisons and ecological phage surveys outside of the dairy environment to substantiate this hypothesis.

Virologist have been intrigued by the question of where do viruses belong in the biological world? What are their origins and closest relatives? At what point did their evolution start to diverge from that of the genetic elements now found in cells? Of all living things, viruses are least likely to have a monophyletic origin because they always replicate in the presence of large amounts of nonviral nucleic acids, which can become incorporated into the viral genome. The bioinformatic analysis of the genomes from S. thermophilus phages has not identified a single case where a bacterial gene and not another phage gene was the closest relative of a S. thermophilus phage gene. This conclusion also applies to the ClpP-like putative protease gene from Sfi21. ClpP proteases are common constituents of prokaryotic and eukaryotic cells and have not been described in any viral system. However, phylogenetic analysis of the Sfi21 ClpP-like protein did not demonstrate a close relationship with bacterial ClpP proteins (Desiere et al., 1999). In another case, a plasmid and not a phage gene turned out to be the nearest database match: Sfi21orf 504 gp showed strong similarity to a putative DNA primase from the cryptic plasmid pWS68 of Lactobacillus delbrueckii (Desiere et al., 1997). In addition, we detected sequence similarity between the putative minus origin from the cryptic S. thermophilus plasmid pST1 and the phage origin, located directly downstream of orf 504 (Desiere et al., 1997). Intriguingly, the noncoding DNA surrounding the cos-site of Sfi19 but not of Sfi21 showed strong similarity to a closely adjacent (80-bp distance) noncoding region of the same plasmid. pST1 is a small cryptic plasmid that encodes only its replicase (Jansen et al., 1992). The sequence similarity between S. thermophilus phages and plasmids is also interesting in view of classical hypotheses that linked the evolution of bacterial viruses to extrachromosomal elements. Plasmids share important properties with temperate phages and indeed can be generated by some phages. Thus phage P1, which showed strong sequence similarity with the Sfi21 antirepressor (Bruttin and Brüssow, 1996), lyses as a plasmid. Also phage λ can multiply as a plasmid after mutation or large deletions. Integrative plasmids resemble temperate phage genomes in their ability to integrate into the bacterial chromosome. It was proposed that evolution proceeded from simple plasmids, which direct only their own replication, to more complex plasmids (F and R factors), then to those bacteriocins that showed the structure of complex phage tails and finally to complete phages. Do we have in S. thermophilus with plasmid pST1 and phage Sfi19 representatives for both ends of this evolutionary sequence? Does S. thermophilus contain intermediate stages of this hypothetical scenario that could make this sequence of events more likely?

METHODS

Phages, strains, and media

The phages were propagated on their appropriate S. thermophilus hosts in lactose M17 broth as described previously (Brüssow and Bruttin, 1995; Bruttin and Brüssow, 1996). 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 Xgal (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 (Brüssow and Bruttin, 1995; Bruttin and Brüssow, 1996). 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.

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 dyeoxy-mediated chain termination using the fmol DNA Sequencing System of Promega (Madison, WI). The sequencing primers were end-labeled using [γ-32P]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 Sau3A-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 Sfi1 DNA sequence, purified phage DNA and Super Taq Polymerase (Stehlin, 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 with those in the databases (GenBank, Release 109; EMBL (Abridged), Release 56; PIR-Protein, Release 57; SWISS-PROT, Release 36; PROSITE, Release 15.0) using FastA (Lipman and Pearson, 1985), and BLAST (Altschul et al., 1990) programs. Sequence alignments were performed using the CLUSTALW 1.74 method (Thompson et al., 1994), the Multalin program (Corpet, 1988), and the SIM align tool (Huang and Miller, 1991).

The complete Sfi1 and Sfi21 genome sequences were deposited in the GenBank database under the Accession No. AF115102 and AF115103.

**ACKNOWLEDGMENTS**

Special thanks go to D. van Sinderen, who shared unpublished data with us that demonstrated that sequence similarity with the cos-site of *S. thermophilus* phages is a widespread property of *S. thermophilus* plasmids. We thank the Swiss National Science Foundation for the financial support of F. Desiere and S. Lucchini in the framework of its Biotechnology Module (Grant 5002-044545/1) and S. Foley for critical reading of the manuscript.

**REFERENCES**


resistance mechanisms in *Streptococcus salivarius* subsp. thermophilus. *J. Dairy Res.* 59, 349–357.


Lucchini, S., Desiere, F., and Brüssow, H. (Submitted). The lysogeny modules from temperate Siphoviridae of low GC gram-positive bacteria share a common genetic organization. *Virology*.


2.2 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 (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

Sacha Lucchini, Frank Desiere, and Harald Brüssow

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 \(\phi Sfi11\) was compared to the corresponding region from other Siphoviridae. The analysis revealed a hierarchy of relatedness. \(\phi Sfi11\) differed from the temperate \(\phi S\) thermophilus bacteriophage \(\phi O1205\) 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, \(\phi Sfi11\) and the lytic group I \(\phi S\) thermophilus \(\phi Sfi19\) shared regions with variable degrees of similarity. Orf 1291 from \(\phi Sfi19\) was replaced by four genes in \(\phi Sfi11\), two of which (orf 1000 and orf 695) showed a complicated pattern of similarity and non-similarity compared with \(\phi Sfi19\). The predicted orf 695 gp resembles the receptor-recognizing protein of \(\phi\) even coliphages in its organization, but not its sequence. No sequence similarity was detected between \(\phi Sfi11\) and \(\phi Sfi19\) in the region covering the major head and tail genes. Comparison of the structural gene map of \(\phi Sfi11\) with that of Siphoviridae from gram-positive and -negative bacterial hosts revealed a common genomic organization. Sequence similarity was only found between \(\phi Sfi11\) 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 may be 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 co-adapted 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 hierarchical levels of order (-virales), family (-viri-), 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 unifying 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).
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 imagine 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 some 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 (Brussow et al., 1994a; Brüssow and Bruttin, 1996). 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 (Racillus subtilis, Mycobacteria, Streptomyces, Escherichia coli) with S. thermophilus (Becker et al., 1997; Arné et al., 1990; Haful 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, φSfi11, and our prototype lytic group II phage, φSfi19 (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 (Brussow 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 multi-component complex.
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 orf11 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 pTP 901-1 coding for a 348-aa-long protein and orf 348 from orf11 (see below, Conservation of genome organization).

Upstream of the lysis cassette, bioinformatic analysis revealed genes coding for likely phage structural proteins. 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 orf11 DNA segment analyzed (Fig. 2).

**Comparison with orf1205**

*S. thermophilus* orf1205 (Stanley *et al.*, 1997) and orf11 differ in lifestyle (orf11 is a virulent phage, while orf1205 is temperate) and in host range (orf1205 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 orf19) orf11 showed very similar genetic organization to orf1205. 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, orf11 was more closely related to lytic group I phage orf19 (Desiere *et al.*, 1998) than to orf1205 (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 orf19**

Figure 4 shows an alignment of the partial orf11 gene map with the corresponding region of lytic group I *S. thermophilus* phage orf19 (Desiere *et al.*, 1998). The two phages showed comparable genetic maps if one postulates a fragmentation of orf 1291 from orf19 into separate orfs in orf11 (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 \( \phi 3S11 \). 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 \( \phi 3S11 \) 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 \( \phi TP901-1 \). The orfs showing similarity to the major structural proteins identified by N-terminal sequencing in \( \phi O1205 \) (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 \( \phi CP-1 \) (Martin et al., 1996), \( \phi E-1 \) (Lopez et al., 1992) and \( \phi Op-1 \) (Sheehan et al., 1996); Lactococcus lactis \( \phi KS-T \) (Boyce et al., 1995); \( \phi TP901-1 \) (Johnsen et al., 1996), \( \phi 11 \) (van Sinderen et al., 1996) and \( \phi E-1 \) (Schouler et al., 1994); Bacillus subtilis \( \phi SP1 \) (Becker et al., 1997); Streptomyces venezuelae phage VW3 (Anné et al., 1993); and Mycobacterium tuberculosis phage LB (Hatful and Säkis, 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

<table>
<thead>
<tr>
<th>( \phi 3S11 ) orf (length in aa)</th>
<th>( \phi TP901-1 ) gene</th>
<th>( \phi TP901-1 ) orf (length in aa)</th>
<th>Identical/aligned aa</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>113 a1 110</td>
<td>26/68</td>
<td>0.0023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>104 b3</td>
<td>103</td>
<td>35/102</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>114 a1</td>
<td>112</td>
<td>32/108</td>
<td>10^{-9}</td>
<td></td>
</tr>
<tr>
<td>128 x</td>
<td>129</td>
<td>37/126</td>
<td>10^{-11}</td>
<td></td>
</tr>
<tr>
<td>168 mtp</td>
<td>169</td>
<td>64/167</td>
<td>10^{-23}</td>
<td></td>
</tr>
<tr>
<td>117 a2</td>
<td>111</td>
<td>28/85</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td>105 c2</td>
<td>120</td>
<td>20/92</td>
<td>10^{-9}</td>
<td></td>
</tr>
<tr>
<td>1510 b4</td>
<td>&gt;203</td>
<td>n.a.</td>
<td>0.0039</td>
<td></td>
</tr>
</tbody>
</table>

Note. The first column gives the length of the \( \phi 3S11 \) orf in number of encoded amino acids. The orfs were listed according to their order on the \( \phi 3S11 \) genome. The second column gives the name of the corresponding \( \phi TP901-1 \) 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 \( \phi 3S11 \) (right lane). The gel was stained with Coomassie brilliant blue. Molecular markers are from rotavirus 993/83 (left lane, Brussow et al., 1992). Molecular masses are given in kilodaltons.
zone from very high to moderately high sequence similarity coincided with the start codon of orf 131 in φSfi11. 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 significant similarity with a minor tail protein 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 ≤ 10^-5) sequence similarity with φSfi11 genes over the structural gene cluster (Fig. 5). We took φSfi11 orf 1510 as the starting

---

**TABLE 2**

<table>
<thead>
<tr>
<th>ORF φSfi11</th>
<th>ORF φO1205</th>
<th>bp differences (%)</th>
<th>aa differences (%)</th>
<th>Distribution of bp differences</th>
<th>Gaps</th>
<th>Deletion/insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>502</td>
<td>502</td>
<td>10.4</td>
<td>7.2</td>
<td>H</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>284</td>
<td>297</td>
<td>10.2</td>
<td>8.5</td>
<td>H</td>
<td>2</td>
<td>D18,21</td>
</tr>
<tr>
<td>193</td>
<td>196</td>
<td>5.9</td>
<td>5.2</td>
<td>C3'</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>119</td>
<td>119</td>
<td>16.2</td>
<td>11.0</td>
<td>H</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>348</td>
<td>348</td>
<td>7.0</td>
<td>4.0</td>
<td>H</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>113</td>
<td>113</td>
<td>17.4</td>
<td>16.0</td>
<td>C3'</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>104</td>
<td>104</td>
<td>5.1</td>
<td>8.0</td>
<td>C3'</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>114</td>
<td>114</td>
<td>0.0</td>
<td>0</td>
<td>—</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>128</td>
<td>128</td>
<td>3.9</td>
<td>5.0</td>
<td>C3'</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>168</td>
<td>168</td>
<td>5.2</td>
<td>2.0</td>
<td>—</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>117</td>
<td>117</td>
<td>1.1</td>
<td>2.0</td>
<td>—</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>105</td>
<td>105</td>
<td>0.6</td>
<td>3.0</td>
<td>—</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>1510</td>
<td>1517</td>
<td>7.7</td>
<td>7.6</td>
<td>Cmiddle</td>
<td>14</td>
<td>D21</td>
</tr>
<tr>
<td>512</td>
<td>512</td>
<td>8.1</td>
<td>5.5</td>
<td>H</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>1000</td>
<td>1000</td>
<td>7.5</td>
<td>5.5</td>
<td>C</td>
<td>1</td>
<td>D18</td>
</tr>
<tr>
<td>373</td>
<td>373</td>
<td>6.0</td>
<td>5.1</td>
<td>H</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>695</td>
<td>643</td>
<td>High</td>
<td>High</td>
<td>C</td>
<td>Many</td>
<td>Mosaic</td>
</tr>
<tr>
<td>699</td>
<td>669</td>
<td>9.9</td>
<td>8.0</td>
<td>C3'</td>
<td>8</td>
<td>—</td>
</tr>
<tr>
<td>149</td>
<td>117</td>
<td>13.1</td>
<td>13.7</td>
<td>C3'</td>
<td>0</td>
<td>I36</td>
</tr>
<tr>
<td>141</td>
<td>141</td>
<td>5.2</td>
<td>5.0</td>
<td>C8'</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>21.6</td>
<td>22.5</td>
<td>H</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>281</td>
<td>281</td>
<td>19.5</td>
<td>17.8</td>
<td>H</td>
<td>9</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. ORF φSfi11: 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 DI in basepairs deduced from the SIM alignment. If more than one number is given, multiple DI were observed. The distinction of DI is arbitrarily based on the φSfi11 sequence.
FIG. 4. Alignment of the partial gene maps from *S. thermophilus* bacteriophages *φ*Sf19 (top line) and *φ*Sf11 (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 *φ*Sf11. 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 *φ*Sf19 (top line) and *φ*Sf11 (bottom line) are indicated by shadings in the central part of the figure. The degrees of aa similarity are graded from black to light gray. 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 coding for a major structural (putative or proven tail) protein (exception: *φ*r11 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 *φ*s1k1 where the distance was five genes). Six genes upstream of the putative major tail gene another major structural gene was localized in all of the phages coding for the putative or proven head protein (exception *φ*s1k1 where the distance was five genes). Six genes upstream of the putative major tail gene another major structural gene was localized in all of the phages coding for the putative or proven head protein (exception *φ*s1k1 where the distance was five genes). *φ*Sf11 differs from the other phages by having two adjacent major structural genes at this position. Directly upstream of this structural gene, *φ*Sf11 had a gene that showed significant similarity to a *φ*s11 gene that preceded the major *φ*r11 structural gene. Finally, at the left end of the partial *φ*Sf11 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 *φ*Sf11, 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 *φ*s1k1, 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 phage sk1 (third line), rlt (fourth line), TP901-1 (fifth line), S. thermophilus phage Sfi11 (sixth line), and E. coli 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 to the length of the predicted open reading frame. The individual ORFs were identified by the numbering system used in the original publication (references for L. lactis phage sk1 [Chandy 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 ≤ 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 Chandy 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 ΔSfi11 at orf 695 or 669. In fact, orf 695 gp fulfills several requirements for a phage adsorption protein. First, it differed among phages that show distinct host ranges (ΔSfi11/O1205; Sfi1/Sfi19), while it was nearly identical among phages showing an extensively overlapping host range (ΔSfi19/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 ΔSfi21 (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 (Boyece et al., 1995a) which also suffered spontaneous deletions (Boyece 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-Ljunquist 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 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 ΔSfi21 and the L. lactis ΔBK5-T (Boyece 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 ΔSfi11 and several Lactococcus phages (ΔTP901-1, 11, 7-9) or S. thermophilus ΔSfi21 with Leuconostoc 4L10 (Desiere et al., 1998). High sequence similarity (up to 66% at the aa level) was detected over several adjacent genes. At the fourth level are relationships between B. subtilis ΔSPP1 and ΔSfi11. Adjacent genes from two different modules (structural genes, DNA packaging, Becker et al., 1997) showed aa identity (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 as 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, φ8K5-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 φ8K5-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, 1996). E. coli strain JM 101 (Strategene) 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 mid-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 6-step preformed CsCl gradient (n0 = 1.4, 1.372, 1.3698, 1.3687, 1.367). The phage bands were recovered with a Pasteur pipette, diluted in phage buffer.
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-32P]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 Sau3A-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 dSfi21 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, release 98; Swiss-PROT, release 34; PROSITE, release 13.0) using the BLAST (Altschul et al., 1990) programs. Sequence alignments were performed using the CLUSTALW 1.6 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 dSfi11 sequence was deposited in the GenBank database under Accession No. AF057033.


2.3 Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory

(Journal of Virology 73, 8647-8656 (1999))
Comparative Genomics of Streptococcus thermophilus Phage Species Supports a Modular Evolution Theory

(Journal of Virology 73: 8647-8656 (1999))

Sacha Lucchini, Frank Desiere and Harald Brüssow

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

Abstract

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 covered lytic and temperate phages of both pac-site and cos-site phages from three distinct geographical areas. The units of genetic exchange were either large and comprised the entire morphogenesis gene cluster excluding the putative tail fiber genes or small thereby 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 range 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 hotspots. Downstream of a highly conserved DNA replication region, all lytic phages showed a short duplication; in three isolates the origin of replication was repeated. Conceptionally, the lytic phages could 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.

Introduction

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 genome 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 (43), coronaviruses (33), alphaviruses (24, 47), arterivirus (28) and in 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 (polyomaviruses, adenoviruses, reviewed in 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 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 more recent 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 (lytic/temperate life style, polypeptide composition, immunity groups, host range, DNA packaging mechanism, geographical and ecological origin). 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-site containing Streptococcus thermophilus phages Sfi19 and Sfi21, respectively, demonstrated that their genomes were similarly organized and differed in addition to numerous point mutations by gene deletion/insertion, duplication and DNA rearrangement events (32). However, comparisons between more than two related phage genomes are necessary to decipher the processes that shaped the genomes of a given phage species. Streptococcus thermophilus phages (5,7) could be a suitable object 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 (7 and references therein). A longitudinal factory survey has documented the ecological dynamics of phage infection (11). All phages characterized until now 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 phages, the French yogurt isolate Sfi19 (17,32) and the Canadian cheese isolate DT1 (44). 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 now 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 / temperate), DNA packaging mechanisms and structural polypeptide pattern (pac-site, cos-site), host ranges, serotype, susceptibility to Sfi21 prophage control, ecological environment and geographical origin. We first used pairwise comparisons of phages differing in defined phenotypes in order to establish associations between phenotype and genotype 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). 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 Xgal (all from Sigma) were used at concentrations of 100 $\mu$g /ml, 1 mM and 0.002% (w/v), respectively. Pertinent features of the compared phages are given in Table 1.

DNA techniques. Phage purification and DNA extraction were done as described previously (6,10). 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.

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-labelled using $\gamma^{33}$P ATP according to the manufacturer’s protocol. The thermal cycler (Perkin Elmer) was programmed at 30 cycles of 95°C for 30 sec, 50°C for 30 sec, and 72°C for 1 min.
In addition, pUC19 clones of Sau3A-digested phage Sfi11 DNA were sequenced using 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 using the synthetic oligonucleotide pair designed according to the established Sfi11 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 109; EMBL (Abridged), Release 56; PIR-Protein, Release 57; SWISS-PROT, Release 36; PROSITE, Release 15.0) using FastA (29) and BLAST (1) programs. Sequence alignments were performed using the CLUSTALW 1.74 method (45), the Multalign program (15) and the SIM alignment tool (27). The complete Sfi11 and Sfi18 genome sequences were deposited in the GenBank database under the accession number AF158600 and AF158601.

Results

Comparison of the Sfi11 and O1205 phage genomes

Phage Sfi11 (31) was chosen for complete genome sequencing since until now no complete genome sequence was available for lytic pac-site S. thermophilus phages. Phage Sfi11 has a 39,807 bp genome consisting of 52 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 share similar structural protein pattern and a similar ecological origin (Table 1). Major differences between both genomes should therefore identify candidate genes involved in life style decision and host range determination. 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 part of their genomes, a one-by-one correspondence of the predicted gene map was observed (Fig. 1). This similarity extended to the nucleotide sequence level.
Figure 1. Prediction of open reading frames 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). For a better reference to previous publications the orfs from O1205 were numbered as in Stanley et al. (40), while the Sfi11 and Sfi19 orfs were marked with their length in codon numbers. Probable gene functions identified by bioinformatic analysis or biological experiments were noted. The phage genomes were divided into functional units according to previous bioinformatic and comparative evolutionary analysis (32). Genes belonging to a same unit have the same color. Grey filling indicates lack of information about the function of the orf. Orfs preceded by a potential RBS (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. Overlap of start and stop codon 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-site and pac-site, respectively, are thus not at the very ends of the maps.
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).

<table>
<thead>
<tr>
<th>Phage</th>
<th>Life Cycle</th>
<th>DNA packaging</th>
<th>Origin</th>
<th>Ecology</th>
<th>Host range</th>
<th>Sero-type</th>
<th>Immunity</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1205</td>
<td>Temp.</td>
<td>pac</td>
<td>Greece</td>
<td>Yoghurt</td>
<td>Non-LGI,II</td>
<td>n.t.</td>
<td>n.a.</td>
<td>40</td>
</tr>
<tr>
<td>Sfi11</td>
<td>Lytic</td>
<td>pac</td>
<td>France</td>
<td>Yoghurt</td>
<td>LG II</td>
<td>2</td>
<td>n.a.</td>
<td>31</td>
</tr>
<tr>
<td>Sfi21</td>
<td>Temp.</td>
<td>cos</td>
<td>France</td>
<td>Yoghurt</td>
<td>LG I</td>
<td>1</td>
<td>+</td>
<td>12,16-19,32</td>
</tr>
<tr>
<td>Sfi19</td>
<td>Lytic</td>
<td>cos</td>
<td>France</td>
<td>Milk</td>
<td>LG I</td>
<td>1</td>
<td>-</td>
<td>32</td>
</tr>
<tr>
<td>Sfi18</td>
<td>Lytic</td>
<td>cos</td>
<td>France</td>
<td>Yoghurt</td>
<td>LG I</td>
<td>1</td>
<td>+</td>
<td>this report</td>
</tr>
<tr>
<td>DT1</td>
<td>Lytic</td>
<td>cos</td>
<td>Canada</td>
<td>Cheese</td>
<td>n.t.</td>
<td>n.t.</td>
<td>n.t.</td>
<td>44</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of the investigated *Streptococcus thermophilus* phages. Notes: For the definition of host range/lytic group (LG) and serotype see (8,31); n.t. not tested. Immunity: The indicated phage is inhibited in its growth (+) on a lysogenic cell containing the Sfi21 prophage or it is not inhibited on the lysogen (-); n.a. non applicable, the O1205 lysogen shows no immunity to superinfection. The Sfi21/Sfi19 comparison has been reported previously (32).

Over the genome region covering the lysogeny module the dotplot indicated a combination of DNA rearrangements and deletion processes between both phages. Over the DNA replication modules the dotplot showed again a straight line. This similarity continued with one interruption until the right end 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 both phages differed on the average by only 10 % at the nucleotide level, while over the lysis module the average nucleotide difference was about 20 %. The DNA replication module was nearly sequence identical, while the rightmost DNA segment showed on the average a greater than 20 % nucleotide difference.
Comparison of the Sfi11 and Sfi19 phage genomes

Phages Sfi11 and Sfi19 differ in the DNA packaging mechanism and structural protein pattern, host range, serotype and ecological origin. In contrast, the two phages share the same life style and geographical origin (Table 1). Both phages showed a similar genome organization (Fig. 1). However, with one exception no nucleotide sequence similarity was seen over the left 18 kb of the aligned genomes in dotplot analysis (Fig. 2B) and the deduced proteins lacked sequence similarity. This region
covers the likely DNA packaging, head and tail morphogenesis modules. Over orf 695 (Sfi11 numbering) regions of nucleotide sequence similarity alternated with regions of non-similarity followed by a nearly uninterrupted dotplot 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 near the right genome end.

**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 both genomes should therefore identify candidate genes for host range determinants and polymorphisms that became established in
geographically and ecologically separated phage lineages. A dotplot analysis revealed a straight line over the structural gene cluster that was interrupted by substitutions and insertion/deletion of gene segments in the three largest orfs from Sfi19 (orf 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 dotplot over the right genome end 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 towards 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 the Sfi19 multiplication was unabated (12, 19). Major differences between both phages could thus identify adaptations of Sfi19 to escape from prophage control. PCR using 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 genome half 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 bp changes resulting in 34 aa changes. The aa changes were unevenly distributed: orf 111 gp alone showed 13 aa changes and the orf 166 gp demonstrated a cluster of 5 aa changes. A third of all bp changes were in two non-coding 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, 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 genome end. All three regions demonstrate the importance of recombination processes for the diversification of S. thermophilus phage genomes.
Figure 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 length. Gaps (−) were introduced for maximal alignment. Aa positions that differ between the corresponding proteins from phages Sfi19 and Sfi21 are underlined and the location of a spontaneous deletion in the Sfi21 protein is marked by broken arrows. Aa positions which were identical in at least three proteins were shaded.
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 aa N-terminal segment, followed by a variable region, 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 aa 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 as recombination hotspots. 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 did not only differ between cos- and pac-site phages, but also between the two pac-site phages. The variable 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 at the position of the lysogeny module a replacement 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 only by three insertion / deletion events: One event was an in frame fusion
Figure 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 were annotated with their codon lengths. The one bp indel was indicated by an arrow, the position of a second indel is marked by a filled 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) were given. 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 light shading.

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 was retained in orf 88 from phage Sfi19 (filled triangle in 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 codon of orf 229. Sfi11 showed at the corresponding position only one 68 bp sequence suggesting again DNA polymerase slippage or homologous recombination as the cause for this polymorphism. Over orf 111 to orf 71 the temperate phage TPJ-34 differed from the lytic phages Sfi11, Sfi18 and Sfi19 by one bp change resulting in a premature stop codon in TPJ-34 (asterisk in Fig. 4) and one bp insertion resulting in different orf predictions for Sfi18 and Sfi19 phages when compared to phage Sfi11 (arrow in Fig. 4).

Notably, the Canadian virulent phage isolate DT1 showed a distinct DNA
Figure 5. Alignment of the gene maps from the right genome ends of the indicated *S. thermophilus* phages. The predicted open reading frames are annotated with their codon lengths. Insertions / deletions are connected by light gray shading. Sequence diversity greater than 20% at the aa level are indicated by different shadings of the arrows representing the orfs.

The rearrangement event over this region (Fig. 4) demonstrating that the majority of the orfs from this region are non-essential genes as already indicated by the high number of polymorphisms in the French isolates. Only one gene is conserved between all four virulent phages (orf 183) and all possess a cro-like gene. However, the DT-1 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). DT-1 showed a distinct orf 104 at this position.

### The right genome end

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 between the investigated phages, while the right part was punctuated by numerous insertion/deletions of entire genes (e.g. orf 57, 146 in DT-1), substitution 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 hotspot. The three lytic phages isolated in France showed at this position a replacement of orf 61 and orf 130 (Sfi21) by a distinct orf 67 followed by a duplicated origin of phage replication. The lytic Canadian isolate DT-1 showed at this position a distinct duplication, while O1205 lacked supplementary DNA between orf 50 and orf 156 (Fig. 5).

Discussion

The comparative sequence analysis revealed that genetic diversity is created in *S. thermophilus* phages by two distinct processes. One process of obvious importance is genetic recombination. Recombination is apparently at 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 bp sequence diversity differed substantially over the various phage genome segments. Practically no sequence variability was found over the replication module, while up to 30 % bp differences were detected over the morphogenesis genes. Interestingly, morphogenesis genes from *Lactococcus lactis* phages showed up to 50 % bp 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 bp changes could free the lytic *S. thermophilus* phage Sfi19 from the superinfection exclusion by the Sfi21 prophage. The analysis of competitive interactions among RNA phages within the framework of the game theory demonstrated that defection (selfishness) evolved, despite the greater fitness pay-off 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 comparable role in *S. thermophilus* phages 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 both 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 hotspot. 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, repressor gene) 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 hotspot. Multiple alignment of these proteins revealed a pattern of conservation and non-conservation 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 anti-receptors. 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 anti-receptor 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 until now prevented the assessment of phylogenetic relationships between the different phage groups. In phage modules, such as the morphogenesis gene cluster, which do not allow exchange reaction, we have observed evidence for the role of point mutations in the evolution of temperate Siphoviridae from low GC 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 share 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). Manillof and Ackermann (34) have delineated a taxonomy of bacterial viruses where 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 from Manillof and Ackermann (34) are somewhat stretched, cos-site S. thermophilus phages could be grouped with the genus "λ-like phages"). The attribution of S. thermophilus phages into two genera seems counterintuitive to us. According to the concept of Manillof and Ackermann (34) many more new Siphoviridae genera will be created as more Siphoviridae from underinvestigated bacterial groups will be characterized. We remind that less than 1 percent 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 dispose of 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.

Acknowledgements

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) and S. Foley for critical reading of the manuscript.

References


2.4 Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria

*(Virology 263, 427-435 (1999))*
Similarly Organized Lysogeny Modules in Temperate Siphoviridae from low GC content Gram-Positive Bacteria

(Virology 263: 427-435 (1999))
Sacha Lucchini, Frank Desiere and Harald Brüssow

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

Abstract

Temperate Siphoviridae from an evolutionary related branch of low GC content Gram-positive bacteria share a common genetic organization of lysogeny-related genes and the predicted proteins are linked by many sequence similarities. Their compact lysogeny modules (integrase/1-2 orfs [phage exclusion? and metalloproteinase motif proteins]/cl-like repressor/ cro-like repressor/ anti-repressor [optional]) differ clearly from that of λ-like and L5-like viruses, the two currently established genera of temperate Siphoviridae, while they resemble those of the P2-like genus of Myoviridae. In all known temperate Siphoviridae from low GC content Gram-positive bacteria the lysogeny module is flanked by the lysis module and the DNA replication module. This modular organization is again distinct from that of the known genera of temperate Siphoviridae. On the basis of comparative sequence analysis we propose a new genus of Siphoviridae: «Sfi21-like» phages. With a larger database of phage sequences it might be possible to establish a genomics-based phage taxonomy and to retrace the evolutionary history of selected phage modules or individual phage genes. The anti-repressor of Sfi21-like phages has an unusual wide-spread distribution since proteins with high aa similarity (40 %) were not only found in phages from Gram-negative bacteria, but also in insect viruses.

Introduction

Principally all cellular organisms are susceptible to viral infection, often by more than one type of virus. A given type of virus usually has a restricted host range, often a single species. In theory this implies that viruses are probably the most diverse creatures on Earth. Viruses might also be the most abundant biological objects typically exceeding the bacterial count in uncultured environmental samples (Fuhrman, 1999). About 4500 bacterial viruses have been examined in the electron microscope. Three quarters are from just 12 bacterial genera (Ackermann, 1996). If one considers that only a tiny fraction of bacteria can presently be cultivated, it becomes clear that we have a very incomplete picture of the diversity of bacterial viruses.

The International Committee on Taxonomy of Viruses (ICTV) has developed a taxonomic system for bacterial viruses. Notably, 96 per cent of all described phages are tailed phages (new order Caudovirales). Three families are distinguished in this order: Myoviridae (viruses with contractile tails), Siphoviridae (viruses with long, noncontractile
tails) and Podoviridae (viruses with short noncontractile tails). At present six genera of phages have tentatively been defined in Myoviridae (e.g. the genus of “P2-like viruses”), six in Siphoviridae (e.g. the genus of “λ-like viruses”) and three in Podoviridae (e.g. the genus of “P22-like viruses”). The genera were defined based on properties involving viral DNA replication, packaging and some specific features (e.g. genome size, temperate/lytic life style; Maniloff and Ackermann, 1998).

A contentious issue is whether these taxonomic units delineate natural phage groups (Ackermann, 1999). The Podovirus P22 (a Salmonella phage), for example, showed a strikingly similar genome organization to the Siphovirus λ (an E. coli phage), both phages were considered members of the lambdoid phage group (Casjens et al., 1992) and indeed viable hybrid phages between P22 and λ have been described (Gemske et al., 1972). The attribution of phages into different phage families on the sole basis of a distinct morphology seems therefore biologically implausible.

Comparative sequence analysis should theoretically shed some light on the phylogenetic relationships between phages. However, two facts limit the use of this straightforward method. On the one hand, many phages encode recombination proteins and are thus very recombinogenic. To account for this fact the modular theory of phage evolution (Botstein, 1980) states that the units of evolution are not individual phage genomes, but groups of functionally related genes (modules) which can relatively freely be exchanged between phages differing otherwise in many respects. On the other hand, only few phage genomes have been sequenced and even less comparative phage genomics studies have been conducted (Hendrix, 1999).

Previously, we conducted comparative and evolutionary analyses for Streptococcus thermophilus phages (Desiere et al., 1998, 1999; Lucchini et al., 1998, 1999a, b), a group of phages of economical importance to the dairy industry (Brüssow et al., 1998). Notably, over the morphogenesis module the degree of sequence similarity between S. thermophilus phages and phage sequences from the database correlated with the evolutionary distance of their bacterial hosts. The Lactococcus lactis phage BK5-T (Boyce et al., 1995) showed still nucleotide sequence similarity with the S. thermophilus phage Sfi21 (Desiere et al., 1999). Phages from more distantly related low GC content Gram-positive bacteria like Bacillus and Lactobacillus showed only sequence similarity at the amino acid level with S. thermophilus phages (Lucchini et al., 1998, Desiere et al.,
1999). Finally, over the morphogenesis genes *S. thermophilus* phage Sfi21 showed a very similar gene map as lambdoid coliphages, but amino acid sequence similarity was no longer detected. The observation of such a hierarchy of relatedness is the hallmark of any biological system undergoing evolutionary changes.

Have *Siphoviridae* an evolutionary history that can partially be deduced by comparative sequencing? To address this question we extended our comparative analysis to another phage module from *Siphoviridae* that is under different functional constraints than the morphogenesis genes. The structural proteins show intensive protein-protein interactions during morphogenesis. In contrast, phage proteins implicated in the control of the virulent / lysogenic lifestyle decision (integrate, excisionase, CI and Cro repressor) interact competitively with target DNA segments while showing relatively few protein-protein interactions. Our database analysis of the lysogeny modules revealed again a close relationship between temperate *Siphoviridae* infecting low GC content Gram-positive bacteria with respect to overall genetic organization and sequence similarity. Their lysogeny modules differed clearly from that of other genera of temperate *Siphoviridae*, while similarly organized lysogeny modules were found in P2-like *Myoviridae*. We discuss the implications of these observations for a sequence-based theory of bacteriophage evolution and phage taxonomy.

**RESULTS**

*Comparable lysogeny modules are found in Siphoviridae from low GC content Gram-positive bacteria*

Probably the best characterized lysogeny module from low GC content Gram-positive phages is that of *S. thermophilus* phage Sfi21 (Fig. 1A and Table 1). From the database we retrieved DNA sequences from temperate *Siphoviridae* and localized their putative lysogeny-related genes. Bioinformatic analysis revealed similarly organized lysogeny modules in *Siphoviridae* from five evolutionary related bacterial genera: *Streptococcus* (n=3), *Lactococcus* (n=3), *Lactobacillus* (n=3), *Staphylococcus* (n=1) and *Bacillus* (n=1).
Figure 1. Comparative genetic organization of the lysogeny related genes in selected temperate phages: A) *Siphoviridae* infecting divers genera of low GC Gram positive bacteria: *Lactobacillus* (galE, adh, A2), *Streptococcus* (Sfi21, TP-J34, O1205, prophage 296) *Lactococcus* (TP901-1, BK5-T, rIt), *Bacillus* (phi-105) and *Staphylococcus* (PVL). B) P2-like genus of *Myoviridae* infecting the Gram negative bacteria *Vibrio cholerae* (K139) and *Hemophilus influenzae* (S2, HP1). C) For comparison the organization of lysogeny related genes is given for the L5-like genus and the λ-like genus of *Siphoviridae* infecting *Mycobacterium tuberculosis* (a high GC content gram-positive bacterium) and *Escherichia coli* (a gram-negative bacterium), respectively. The genomes are aligned at the left by their integrase genes. Gene functions attributed by bioinformatic analysis or biological experiment are color-coded as indicated in the figure legend (mpm: metalloproteinase motif (Jongeneel et al., 1989)). The length of the arrow is proportional to the length of the predicted open reading frame. Note the change in scale in Fig. 1C (reduced). D) Alignment of the putative lysogeny modules and the adjacent region in *L. lactis* rIt and BK5-T. The predicted open reading frames are color-coded (see Fig. 1C) and are annotated as in their original publication (see table 2 for the references). The colored bars give the percentage of nt sequence identity between the two compared sequences as defined in the color code at the top of the figure. The figure was created with the SIM Alignment tool (Huang and Miller, 1991) and LALNview (Duret et al., 1996).
The compact organization of lysogeny-related genes in this group of phages differed clearly from the genetic organization of lysogeny-related genes in *Siphoviridae* from high GC content gram-positive bacteria (L5-like genus) and from Gram-negative bacteria (λ-like genus). In the *E. coli* phage λ, fourteen genes separate the integrase/ excisionase genes from the immunity/ genetic switch region and in the mycobacteriophage L5 thirty-five orfs separated the integrase gene from the repressor gene and no genetic switch region was observed (Fig. 1C).

<table>
<thead>
<tr>
<th>Orf</th>
<th>Likely gene function</th>
<th>Evidence</th>
<th>Comments</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>359</td>
<td>integrase</td>
<td>Bioinformatics, cloning and expression</td>
<td>Integration vector</td>
<td>Bruttin et al., 1997a,b</td>
</tr>
<tr>
<td>203</td>
<td>Superinfection exclusion</td>
<td>cloning and expression</td>
<td>Phage resistance</td>
<td>Bruttin et al., 1997b</td>
</tr>
<tr>
<td>122</td>
<td>?</td>
<td>-</td>
<td>Metalloproteinase motif</td>
<td>Neve et al., 1998</td>
</tr>
<tr>
<td>127</td>
<td>Cl-like repressor</td>
<td>Bioinformatics, cloning and expression</td>
<td>Phage resistance, genetic switch binding, H-T-H motif</td>
<td>Bruttin et al., 1997 b, Foley et al., unpublished</td>
</tr>
<tr>
<td>75</td>
<td>Cro-like repressor</td>
<td>Bioinformatics</td>
<td>H-T-H motif</td>
<td>Bruttin et al., 1997 b</td>
</tr>
<tr>
<td>287</td>
<td>Anti-repressor</td>
<td>Bioinformatics</td>
<td>H-T-H motif</td>
<td>Bruttin and Brüssow, 1996; Neve et al., 1998</td>
</tr>
<tr>
<td>93</td>
<td>?</td>
<td>-</td>
<td>weak excisionase motif</td>
<td>Boccard et al., 1989</td>
</tr>
</tbody>
</table>

Table 1: The genes from the lysogeny module of *S. thermophilus* phage Sfi21.

Temperate phages are also found in other groups of tailed phages (P1-like, P2-like, Mu-like, ΦH-like genera of *Myoviridae* and P22-like genus of *Podoviridae*). Interestingly, the organization of the lysogeny genes of P2-like *Myoviridae* which infect diverse Gram-negative bacteria is very similar to that of temperate *Siphoviridae* from low GC content Gram-positive bacteria (Fig. 1B, Table 3 provides a compilation of the phages investigated in the current report). For example, the *Vibrio cholerae* phage K139 showed the gene order: *attP*- *int*- *glo* (superinfection exclusion)-orf 2-*cl-cox*- (a combination of a *cro*-like transcriptional and a *xis*-like recombinational switch; Esposito and Scocca, 1994)-*cll* (Nesper et al., 1999; Fig. 1B). In comparison, *S. thermophilus* phage Sfi21 showed the gene order: *attP-int-imm* (superinfection exclusion)-*mpm* (metalloproteinase motif protein)-*cl-cro-ant* (antirepressor) (Fig. 1A).
**Modular exchanges in the lysogeny modules of lactococcal phages**

Apart from the three temperate *S. thermophilus* phages (Neve et al., 1998), only the lysogeny modules from the two *Lactococcus lactis* phages r1t and BK5-T could be aligned at the nucleotide level (Fig. 1D). As previously demonstrated for *S. thermophilus* phages (Neve et al., 1998), the characteristic feature of this alignment was the alternation of regions with high DNA sequence identity with DNA segments showing only low or no sequence identity (Fig. 1D). The int gene, part of the repressor gene and a possible ant gene homologue (see below) were highly conserved at the DNA level.

Except for the repressor gene, the transition zones of DNA conservation to non-conservation coincided precisely with gene borders (Fig. 1D and 2). Over the right border of the lysogeny module, the phage r1t DNA can formally be regarded as a product of recombination between phages BK5-T and TP901-1 and at least one further phage (Fig. 2). Three possible crossover points were identified in a 2 kb stretch of phage r1t DNA. The very high DNA sequence similarity indicated that BK5-T and TP901-1 are very near to the authentic phages which contributed DNA to r1t.
The lysogeny modules are located between the lysis and the DNA replication modules

For nine out of the twelve Siphoviridae presented in Fig. 1A sequence information was provided for the regions flanking the lysogeny modules. Downstream of the integrase gene, separated by zero to six orfs a lysin gene was identified by the different authors (Fig. 3, see Table 3 for the original references). On the other border, the lysogeny module was flanked by a likely DNA replication module. One to eight orfs downstream of the cro-like repressor gene the different researchers identified genes which predicted proteins that showed bioinformatic links to DNA replication functions (Fig. 3). The genes encoded a putative helicase (g1e), a putative replication initiation and DnaC-like protein (rft), a putative single stranded DNA binding protein (TP901-1) or proteins with similarity to predicted proteins from the DNA replication modules of phage Sfi21 (Desiere et al., 1997, Foley et al., 1998) or phage rft (van Sinderen et al., 1996) (BK5-T, PVL).

Figure 3. The lysogeny module of Streptococcus (O1205, Sfi21), Lactobacillus (g1e, adh), Lactococcus (BK5-T, rft, TP901-1), Bacillus (phi-105) and Staphylococcus (PVL) phages are flanked by the lysis and the DNA replication module. The putative lysogeny related genes are shown with open arrows. Genes attributed by bioinformatic analysis to the lysis module (H: holin, L: lysin) are shown with black arrows, genes attributed to likely DNA replication functions (see text) are shown as striped arrows. Genes which cannot be attributed to a functional module are shown with gray arrows. I: integrase, M: metalloproteinase motif protein, R: repressor; C: Cro-like repressor, A: antirepressor.
Sequence similarities between the lysogeny modules from Siphoviridae of low GC content Gram-positive bacteria

At the protein level many sequence similarities linked the lysogeny modules from Siphoviridae infecting low GC content Gram-positive bacteria. Where significant matches (P values < 10^{-3}) to the database were found, the best matches were always to proteins from other members of this phage group.

All but one of the phage integrases were evolutionary related. A phylogenetic tree was previously presented (Bruttin et al., 1997a). The integrase from phage TP901-1 (Christiansen et al., 1996) belongs to a different class of enzymes (resolvases), its closest relative was found in the Bacillus cereus phage TP21 and phage Mu.

Proteins with the metalloproteinase motif were detected in Siphoviridae from Streptococcus, Lactococcus, Bacillus and Lactobacillus (data not shown). The sequence similarities were not limited to this motif, but were found over the entire protein sequence.

The sequence similarity defined two subtypes of CI-like repressor proteins (Table 2). One subtype was represented by the phage TPJ-34 repressor, the other by the phage TP901-1 repressor. The repressor from phage O1205 linked both subgroups, as well as two outliers (Table 2). A multiple alignment was possible over the entire length of the three repressor proteins.

<table>
<thead>
<tr>
<th></th>
<th>adh</th>
<th>Sf21</th>
<th>TPJ34</th>
<th>A2</th>
<th>O1205</th>
<th>TP901</th>
<th>BK5-T</th>
<th>r1t</th>
<th>296</th>
<th>PVL</th>
<th>g1e</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td>dh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f21</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P34</td>
<td>7</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1205</td>
<td>*</td>
<td>5</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P901</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K5-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>109</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>5</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>32</td>
<td>8</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>1e</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Sequence relationships between the putative repressor proteins of the indicated Siphoviridae from low GC content Gram-positive bacteria. Note: See Table 2 for a key to the phage strains. The figures are the negative decadic logarithms of the P values from Blast searches (e.g. 6 means P=10^-6). * 25 identical out of 102 aa positions, ** P=0.07, *** 31 identical out of 129 aa positions.
The Cro-like repressors tended to be more diverse with relatively distinct Sfi21, O1205 and BK5-T phage proteins (data not shown). Downstream of the cro gene, a putative antirepressor gene (Neve et al., 1998) was detected in three streptococcal phages (Sfi21, TP-J34, 296) and the staphylococcal phage PVL. At a corresponding position a distinct gene was detected in two Lactococcus phages (BK5-T, r1t) and one Lactobacillus phage (LL-H), while a hybrid gene was seen in Lactobacillus phage A2 that linked both subgroups. In fact, the A2 protein shared its N-terminal half with the BK5-T protein and its C-terminal half with the TP-J34 anti-repressor (Fig. 4).

Figure 4. Comparison of the putative antirepressor genes detected in phages from Streptococcus (TP-J34), Staphylococcus (PVL), Escherichia (coliphages P1, N15, H-19B), Lactococcus (BK5-T), Lactobacillus (A2, a stop codon was changed) or prophages (Haemophilus influenzae flu; B. subtilis SPBc2, Streptococcus pyogenes 296). Gene products are depicted by boxes and drawn to scale. Domains having amino acid similarity are shown by identical shading within the boxes.

This observation suggests a two-domain structure for these proteins and the possibility of recombination between these protein domains. Proteins which shared either the N- or the C-terminal halves of these hypothetical anti-repressors were detected in phages and prophages from Gram-negative bacteria (coliphages P1, N15, H-19B, Haemophilus influenzae prophage φflu) (Fig. 4). Overall these phage proteins defined four distinct C-terminal domains (TPJ-34-like and BK5-T-like [linked by the B. subtilis prophage SPBc2 domain], φflu-like, N15-like domain) and four clearly distinct N-terminal domains (TPJ-34-like, BK5-T-like, H-19B-like, P1-like). Notably, the evolutionary
distribution of one domain was even wider: proteins with sequence similarity to the N-terminal half of the BK5-T protein were also found in several different families of eukaryotic invertebrate viruses (Baculoviridae: orf 2 protein from Autographa californica nuclear polyhedrosis virus, P=10^{-7}; Poxviridae: MSV194 protein from Melanoplus sanguinipes entomopoxvirus, P=10^{-22}; Iridoviridae: 011L protein from Chilo iridescent virus, P=10^{-4}). The proteins belong to a new uncharacterized gene family of insect viruses (Afonso et al., 1999). The entomopoxvirus protein shared an astonishing 40 % aa identity with the N-terminal domain of the BK5-T protein (Fig. 5).

In comparison the degree of aa identity between proteins from temperate Siphoviridae infecting different genera of low GC content Gram-positive bacteria did not exceed 33 % for the integrases, 40 % for the Cl and Cro proteins and 56 % for the Ant proteins.

Figure 5. Multiple alignment of the N-terminal half of the antirepressor proteins of Siphoviridae from Lactococcus lactis phage BK5-T, Lactobacillus casei phage A2. E. coli phage N15 prophage flu from Haemophilus influenzae (for all see Fig. 4) with two insect viruses (Baculovirus from Lymantria dispar [orf 161 gp] and Poxvirus from Melanoplus sanguinipes [MSV194 gp]). Amino acid positions which are identical in at least three proteins were shaded.
Comparison with the lysogeny module from P2-like Myoviridae

The proteins encoded by the lysogeny modules of the P2-like genus of Myoviridae were also linked by many sequence similarities. *Haemophilus influenzae* phages HP1 and S2 are highly related. Their lysogeny modules could be aligned resulting in regions with high DNA sequence identity interspersed with DNA segments showing only low or no sequence identity (Skowronek and Baranowski, 1997). Their Int and CI proteins showed also significant similarity to *E. coli* phage P186 and *V. cholerae* phage K139, two further *Myoviridae* of the P2 group. The latter two phages shared a second type of Cox and CII protein.

The gene upstream of the integrase showed no similarity between the four P2-like phages. The same is also true for temperate *Siphoviridae* of low GC content Gram-positive bacteria. Interestingly, the S2 phage protein showed detectable aa similarity with the corresponding protein from r1t (P=10^-4) and in both cases a hydrophobic N-terminus was predicted. Recently, Nesper et al. (1999) have revealed a structural similarity between the genes upstream of the integrase in the *Myoviridae* K139 and the *Siphoviridae* TPJ-34. This gene encodes both in the Myovirus K139 (Nesper et al., 1999) and in the Siphovirus Sfi21 a superinfection exclusion function (Bruttin et al., 1997a). In addition, all integrases of the P2 phage group shared weak sequence similarity (about 25% aa identity over the C-terminal halves, P=10^-5) with the integrases of *Siphoviridae* from low GC content Gram-positive bacteria.

Discussion

Temperate *Siphoviridae* from low GC-content Gram-positive bacteria share not only related morphogenesis modules, but as demonstrated in the current communication also related lysogeny modules. In fact, all currently known temperate *Siphoviridae* from this evolutionary related group of bacteria showed an identical overall genome organization with the following modular organization: DNA packaging → head morphogenesis → tail morphogenesis → tail fiber morphogenesis → lysis → lysogeny → DNA replication and an unattributed, possibly regulatory module (Desiere et al., manuscript in preparation). This common overall organization and the peculiar organization of the lysogeny-related genes differentiates this group of temperate *Siphoviridae* from the two currently
established genera of temperate Siphoviridae (the \(\lambda\)-like genus and the L5-like genus, Maniloff and Ackermann, 1998). Taxonomically, one could propose a new genus ("Sfi21-like phages") to the current ICTV scheme of Siphoviridae.

The aim of taxonomy is the introduction of order into the biological diversity according to rational criteria. The ultimate aim is that this taxonomical order reflects the natural phylogenetic relationships between the biological objects. Bacteriophages present a dilemma with respect to this goal. The modular theory of phage evolution predicts exchanges of functional gene segments between phages that differ otherwise in many respects (Botstein, 1980). On the basis of the remarkable recombinogenic properties of phages, it has been doubted that phylogenetic relationships can be established for phages. It is thus interesting to note that temperate Siphoviridae from a group of evolutionary related bacteria are extensively linked by sequence relationships over two modules that are under different evolutionary constraints. Actually, a preliminary analysis of further genome regions (lysis, DNA replication modules) demonstrated additional sequence relationships between temperate Siphoviridae from low GC content Gram-positive bacteria. Apparently, this group of phages shares, at least to a certain extent, a common evolutionary history. It will now be important to extend the comparative analysis of phage genomes to other groups of bacterial viruses. Do temperate Siphoviridae from high GC content Gram-positive bacteria or from Gram-negative bacteria also constitute groups of phages with a common genetic organization suggestive of shared evolutionary histories?

Interestingly, distinct evolutionary affinities were revealed for the different modules from temperate Siphoviridae infecting low GC content Gram-positive bacteria. Over the morphogenesis module their closest relatives were \(\lambda\)-like Siphoviridae, while over the lysogeny module their closest relatives were P2-like Myoviridae. As in the case of the genetic similarities between the Siphovirus \(\lambda\) and the Podovirus P22, the relationship between Sfi21-like Siphoviridae and the P2-like Myoviridae blurs the taxonomical distinction of phage families distinguished primarily on morphological grounds.

A further contentious issue of phage genomics is the relationship of bacterial viruses to other biological entities. A previous analysis of gene sequences from theE. coli phage T4 revealed relationships to both bacteria and eukaryotes (Bernstein and Bernstein, 1989). What can be deduced from our comparative sequence analysis of S.
thermophilus phage genomes? With very few exceptions (e.g. repressors) S. thermophilus phage genes demonstrated no sequence similarities with bacterial genes. In striking contrast, clear links were observed between S. thermophilus phages and plasmids of the same bacterial species. Previously we reported that two non-coding phage regions (cos-site and origin of replication) showed sequence similarity to a cryptic S. thermophilus plasmid (Lucchini et al., 1999a). A recently sequenced S. thermophilus plasmid showed another link: its putative primase showed 28% aa identity over its entire length with a putative S. thermophilus phage primase (P=10^{-48}, accession number AJ242479.1). In contrast to T4, no S. thermophilus phage genes showed closer links to eukaryotic than to prokaryotic genes. Apparently, S. thermophilus phages derive their genes from other phage genomes and perhaps S. thermophilus plasmids. Interestingly, some weak, but significant links were also observed between S. thermophilus phages and eukaryotic viruses, e.g. a virus infecting the green algae Chlorella (Desiere et al., 1997). In the current report we describe a highly significant link between the hypothetical anti-repressor of staphylococcal, lactococcal and lactobacillar phages and at least two morphologically distinct classes of insect viruses. The current database of phage, algal and invertebrate viruses is too small to assess whether these links between bacteriophages and eukaryotic viruses are chance observations or reflect actual genetic relationships (common ancestry, ancient modular gene transfers).

We have identified only about 35 complete phage genome sequences in the database (excluding prophage sequences from bacterial genome projects). This number is clearly insufficient to decide whether the evolutionary history of individual phage modules or even entire phage genomes can be deduced from comparative sequence analysis. With the current ease of DNA sequence acquisition, an international consortium interested in bacteriophage evolution should direct efforts to obtain phage sequences from a set of phages representing a wide evolutionary distribution of cultivatable bacteria. Especially revealing could be random sequencing of DNA extracted from uncultivated tailed phages purified from environmental sources like ocean water and freshwater.
Material and Methods

All phage DNA sequences analyzed in the present report were retrieved from the database. Table 3 provide a list of the phage identifications, database accession numbers and references. Nucleotide and predicted amino acid sequences were compared to those in the databases (GenBank, Release 102; SWISS-PROT, Release 34 using FastA (Lipman and Pearson, 1985) and BLAST (Altschul et al., 1990) programs. Sequence alignments were performed using the CLUSTALW 1.6 method (Thompson et al., 1994), the Multalign program (Corpet, 1988) and the SIM alignment tool (Duret et al., 1996; Huang and Miller, 1991).

<table>
<thead>
<tr>
<th>Phage</th>
<th>Host (Phage Family)</th>
<th>Acc. no</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sf21</td>
<td><em>Streptococcus thermophilus</em> (S)</td>
<td>X95646</td>
<td>Lucchini et al., 1999</td>
</tr>
<tr>
<td>TPJ-34</td>
<td><em>Streptococcus thermophilus</em> (S)</td>
<td>AF020798</td>
<td>Neve et al., 1998</td>
</tr>
<tr>
<td>1205</td>
<td><em>Streptococcus thermophilus</em> (S)</td>
<td>U88974</td>
<td>Stanley et al., 1997</td>
</tr>
<tr>
<td>296</td>
<td><em>Streptococcus pyogenes</em> (S)*</td>
<td>Unfinished</td>
<td>Genome Project</td>
</tr>
<tr>
<td>TP901-1</td>
<td><em>Lactococcus lactis</em> (S)</td>
<td>Y14232</td>
<td>Madsen and Hammer, 1998</td>
</tr>
<tr>
<td>BK5-T</td>
<td><em>Lactococcus lactis</em> (S)</td>
<td>L44593</td>
<td>Boyce et al., 1995</td>
</tr>
<tr>
<td>r1t</td>
<td><em>Lactococcus lactis</em> (S)</td>
<td>U38905</td>
<td>Van Sinderen et al., 1996</td>
</tr>
<tr>
<td>adh</td>
<td><em>Lactobacillus gasseri</em> (S)</td>
<td>Z97974</td>
<td>Engel et al., 1998</td>
</tr>
<tr>
<td>A2</td>
<td><em>Lactobacillus casei</em> (S)</td>
<td>Y12813</td>
<td>Alvarez et al., 1998</td>
</tr>
<tr>
<td>g1e</td>
<td><em>Lactobacillus spec.</em> (S)</td>
<td>X98106</td>
<td>Kodaira et al., 1997</td>
</tr>
<tr>
<td>PVL</td>
<td><em>Staphylococcus aureus</em> (S)</td>
<td>AB009866</td>
<td>Kaneko et al., 1998</td>
</tr>
<tr>
<td>105</td>
<td><em>Bacillus subtilis</em> (S)</td>
<td>AB016282</td>
<td>unpublished</td>
</tr>
<tr>
<td>L5</td>
<td><em>Mycobacterium smegmatis</em> (S)</td>
<td>Z18946</td>
<td>Hatfull and Sarkis, 1993</td>
</tr>
<tr>
<td>K139</td>
<td><em>Vibrio cholerae</em> (M)</td>
<td>AF125163</td>
<td>Nesper et al., 1999</td>
</tr>
<tr>
<td>S2</td>
<td><em>Haemophilus influenzae</em> (M)</td>
<td>Z71576</td>
<td>Skowronek and Baranowski, 1997</td>
</tr>
<tr>
<td>HP1</td>
<td><em>Haemophilus influenzae</em> (M)</td>
<td>U06847</td>
<td>Esposito et al., 1994</td>
</tr>
</tbody>
</table>

Table 3: List of phages used for the comparative genome analysis. Notes: Phage family (S), Siphoviridae; (M), Myoviridae. * prophage sequence from bacterial genome project (J. Ferretti University of Oklahoma). # Incomplete phage sequence

Reference List


2.5 A short noncoding viral DNA element showing characteristics of a replication origin confers bacteriophage resistance to *Streptococcus thermophilus* 

(Virology 250, 377-387 (1998))
A Short Noncoding Viral DNA Element Showing Characteristics of a Replication Origin Confers Bacteriophage Resistance to *Streptococcus thermophilus*

Sophie Foley, Sacha Lucchini, Marie-Camille Zwahlen, and Harald Brüssow

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

Received May 21, 1998; returned to author for revision August 4, 1998; accepted August 19, 1998

A 302-bp noncoding DNA fragment from the DNA replication module of phage øSfi21 was shown to protect the *Streptococcus thermophilus* strain Sfi1 from infection by 17 of 25 phages. The phage-inhibitory DNA possesses two determinants, each of which individually mediated phage resistance. The phage-inhibitory activity was copy number dependent and operates by blocking the accumulation of phage DNA. Furthermore, when cloned on a plasmid, the øSfi21 DNA acts as an origin of replication driven by phage infection. Protein or proteins in the øSfi21-infected cells were shown to interact with this phage-inhibitory DNA fragment, forming a retarded protein-DNA complex in gel retardation assays. A model in which phage proteins interact with the inhibitory DNA such that they are no longer available for phage propagation can be used to explain the observed bacteriophage resistance. Genome analysis of øSfi19, a phage that is insensitive to the inhibitory activity of the øSfi21-derived DNA, led to the characterisation of a variant putative phage replication origin that differed in 14 of 302 nucleotides from that of øSfi21. The variant origin was cloned and exhibited an inhibitory activity toward phages that were insensitive to the øSfi21-derived DNA.

INTRODUCTION

After virus adsorption and DNA injection, DNA replication is a critical step in the virus life cycle and is frequently targeted in strategies developed to block viral infection. Bacteria possess several native defence mechanisms that protect against bacteriophage infection. During the past decade, molecular approaches have provided a number of genetically defined resistance mechanisms directed against mesophilic lactococcal phages (reviewed by Garvey et al., 1995b). Many of the resistance mechanisms are based on defence strategies native to *Lactococcus* and are frequently encoded by plasmids, the majority of which are self-transmissible. The resistance mechanisms characterised to date can be grouped into four main categories based on their mode of action: adsorption inhibition, DNA penetration blocking, restriction/modification (R/M), and abortive infection (Abi). A number of the Abi systems characterised from the lactic acid bacteria *Lactococcus* operate either directly or indirectly by blocking phage DNA replication (e.g., abiA, Hill et al., 1991; abiF, Garvey et al., 1995a). However, the exact mechanisms by which the phage DNA replication is inhibited remain to be determined (for a review, see Garvey et al., 1995b). Hill et al. (1990) successfully exploited lactococcal bacteriophage ø60 DNA in the development of an abortive type of phage resistance. The protective DNA included a 500-bp region rich in secondary structure, which represented the phage origin of replication (ori).

Our laboratory is interested in bacteriophages of *Streptococcus thermophilus*, a thermophilic lactic acid bacterium. Little information is available concerning antiphage mechanisms in this species (Larbi et al., 1992). However, Moineau et al. (1995) successfully protected several industrial *S. thermophilus* strains from phage infection by the introduction of an R/M determinant from a lactococcal plasmid. The fact that few phage defence mechanisms have been reported or developed for *S. thermophilus* reflects not only the fact that the molecular characterisation of *S. thermophilus* and its phages has started relatively recently but also that, in contrast to *Lactococcus*, plasmids are conspicuously scarce in *S. thermophilus* (Mercenier et al., 1990).

The increasing amount of genome sequence information available for *S. thermophilus* phages should facilitate the development of bacteriophage resistance mechanisms for the protection of industrial starter cultures. Sequence analysis of the *S. thermophilus* øSfi21 genome, combined with Southern blot hybridisation and analysis of restriction profiles, led to the definition of a DNA replication module of ~5 kb (Desiere et al., 1997). Based on DNA hybridisation experiments, this DNA module is present in ~70% of *S. thermophilus* phages (Brüssow et al., 1994). DNA sequence analysis of this region from øSfi1 identified several open reading frames (orf5) putatively coding for enzymes, including primase and helicase, implicated in the initiation of phage DNA rep-
Identification of bacotrophage resistance determinant

The clone pMZ23c, consisting of an 824-bp EcoRI fragment of φSfi21 cloned in the shuttle vector pNZ124 (Fig. 1A), protected the plasmid-free laboratory strain S. thermophilus Sfi1 from infection by a large number of phages (Table 1). Of 25 phages that are capable of propagating on Sfi1, 17 are sensitive to the presence of pMZ23c, as demonstrated by a dramatic reduction in the efficiency of plaquing (EOP). In fact, for all 17 phages, no plaques could be detected. This is equivalent to a >6-log reduction in phage titres, which compares favourably with the 4-log reduction observed for the phage-encoded resistance (Per) mechanism derived from a lactococcal phage (O’Sullivan et al., 1993). When introduced in the S. thermophilus strain Sfi9, pMZ23c also protects against infection by the lytic group II phages φSfi11, φSfi9, φSfi12, and φSfi2 (data not shown).
TABLE 1
Titration of S. thermophilus Phages on Sfi1 Transformed with the Indicated Plasmids

<table>
<thead>
<tr>
<th>Phage</th>
<th>pNZ124*</th>
<th>pMZ23c</th>
<th>pSF23e</th>
<th>pSF23f</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(pfu/ml)</td>
<td>(pfu/ml)</td>
<td>(pfu/ml)</td>
<td>(pfu/ml)</td>
</tr>
<tr>
<td>Sfi21</td>
<td>$9.7 \times 10^7$</td>
<td>&lt;10$^{26}$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi9</td>
<td>$1.7 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi9</td>
<td>$1 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>S3</td>
<td>$2.3 \times 10^7$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi9</td>
<td>$3 \times 10^7$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi9</td>
<td>$7.5 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>S3</td>
<td>$2.6 \times 10^9$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi44A</td>
<td>$3.2 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi1</td>
<td>$1 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi6</td>
<td>$6 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi44</td>
<td>$6 \times 10^6$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi42</td>
<td>$8 \times 10^6$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi4-31</td>
<td>$4.8 \times 10^7$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi4-15</td>
<td>$2.7 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi40</td>
<td>$2 \times 10^5$</td>
<td>&lt;10$^3$</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi18</td>
<td>$6 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi20</td>
<td>$3 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi19</td>
<td>$2.8 \times 10^9$</td>
<td>$1 \times 10^9$</td>
<td>1.4 $\times 10^9$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi25</td>
<td>$2.5 \times 10^8$</td>
<td>$3 \times 10^7$</td>
<td>$3 \times 10^7$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi33</td>
<td>$3 \times 10^8$</td>
<td>$1 \times 10^7$</td>
<td>$7 \times 10^7$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi28</td>
<td>$9 \times 10^9$</td>
<td>$1 \times 10^8$</td>
<td>$8 \times 10^8$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi17</td>
<td>$6 \times 10^9$</td>
<td>$6 \times 10^8$</td>
<td>$5 \times 10^8$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi19</td>
<td>$1 \times 10^8$</td>
<td>$2 \times 10^7$</td>
<td>$2 \times 10^7$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi7</td>
<td>$4 \times 10^8$</td>
<td>$2 \times 10^7$</td>
<td>$4 \times 10^7$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi30</td>
<td>$3 \times 10^8$</td>
<td>$1 \times 10^7$</td>
<td>$3 \times 10^7$</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

* Phage-sensitive vector control.

b <10$^2$ indicates that no plaques were detected on a lawn of bacterial growth.

n.d., not determined.

To define further the DNA sequence responsible for mediating the observed bacteriophage resistance, the EcoRI insert fragment of pMZ23c was subcloned (Fig. 1A), and the various subclones were tested for their ability to protect the strain Sfi1 from phage infection. The insert fragment of pMZ23c contains three orfs: orf 61, the oppositely oriented orf 143, and orf 50 (Fig. 1A). The three subclones pSF23d (deletion of orf 50 and orf 61), pSF23e (deletion of orf 50), and pSF23f (deletion of orf 143) protected Sfi1 from infection by $\phi$Sfi21, as determined by the inability to detect plaques (Table 2). Therefore, all three orfs present in pMZ23c could be deleted without loss of the resistance phenotype, suggesting that the observed resistance is not mediated by a phage protein.

The intergenic region between the divergently oriented orf 61 and orf 143 was then tested. A 302-bp DNA fragment, covering the $\phi$Sfi21 sequence showing similarity to the putative pSI23 sso of replication, was PCR amplified and cloned in pNZ124, generating the construct pSF23g (Fig. 1A). The presence of pSF23g in S. thermophilus Sfi1 protected the bacteria from phage infection and was just as effective as the parent clone pMZ23c in terms of reduction in the EOP and range of phages inhibited (Table 1). From these observations, we can conclude that the observed bacteriophage resistance in pMZ23c is mediated by a noncoding DNA sequence. Furthermore, because the noncoding sequences in pSF23g and pMZ23c are in opposite orientations with respect to each other, the inhibitory effect of the 302-bp noncoding sequence must be independent of transcription originating in the vector.

Interestingly, both of the clones pSF23d and pSF23f, in which all of the DNA to the left and right, respectively, of the Basi site is deleted, mediate resistance against phages $\phi$Sfi21, $\phi$Sfi9, and $\phi$Sfi9 (Table 2). This implies that pMZ23c contains, in fact, two genetic elements located on either side of the Basi site, each of which independently is capable of protecting the host cell from bacteriophage infection. A closer examination of the nucleotide sequence of pSF23g highlighted a number of sequences repeated to the left and right of the Basi site (Fig. 1B). It should, however, be noted that the protection provided by pSF23g is much stronger than that of pSF23d or pSF23f in terms of the reduction in EOP, and/or the range of phages affected (Table 2). These

TABLE 2
Analysis of pMZ23c Subclones for Bacteriophage Resistance

<table>
<thead>
<tr>
<th>Phage</th>
<th>pNZ124</th>
<th>pSF23d</th>
<th>pSF23e</th>
<th>pSF23f</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(pfu/ml)</td>
<td>(pfu/ml)</td>
<td>(pfu/ml)</td>
<td>(pfu/ml)</td>
</tr>
<tr>
<td>Sfi21</td>
<td>$9.7 \times 10^4$</td>
<td>&lt;10$^{26}$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi9</td>
<td>$9.9 \times 10^8$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>S3</td>
<td>$2.3 \times 10^7$</td>
<td>&lt;10$^2$</td>
<td>&lt;10$^5$</td>
<td>&lt;10$^5$</td>
</tr>
<tr>
<td>Sfi4-31</td>
<td>$2.6 \times 10^3$</td>
<td>&lt;10$^2$</td>
<td>8.8 $\times 10^7$</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sfi4-15</td>
<td>$3.2 \times 10^3$</td>
<td>&lt;10$^2$</td>
<td>2.3 $\times 10^7$</td>
<td>n.d.</td>
</tr>
<tr>
<td>S96</td>
<td>$6 \times 10^3$</td>
<td>&lt;10$^2$</td>
<td>2.6 $\times 10^3$</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

Note. All of the above plasmids were tested in strain Sfi1.

a <10$^2$ indicates that no plaques were detected on the lawn of bacterial growth.

n.d., not determined.
differences probably reflect different binding requirements of the corresponding phage proteins.

Effect of copy number on bacteriophage resistance phenotype

pSF28 is a nonreplicative integration vector for *S. thermophilus* based on the integration functions of the phage <\(\phi\)Sfi21 and the cat gene from pC194 as a selectable marker (Bruttin et al., 1997b). To examine whether the 824-nt <\(\phi\)Sfi21 fragment from pMZ23c is capable of interfering with phage infection when present as a single copy in the bacterial cell, it was cloned as an XbaI-PstI fragment in pSF28. This cloning was performed in *Lactococcus lactis* ssp. *lactis* MG1383 because the phage integrase gene present in pSF28 is toxic in *Escherichia coli*. *S. thermophilus* Sfi1 was subsequently transformed with the resulting construct, pSF20, and the integrants obtained were confirmed by PCR analysis. The bacteriophage sensitivity of this integrant Sfi1:pSF20 was tested by plaque assays. For <\(\phi\)Sfi21, a 100-fold reduction in the EOP was observed together with a reduced plaque size. However, no reduction in the EOP or plaque size was observed for the two phages <\(\phi\)S69 and <\(\phi\)S89.

The 824-nt <\(\phi\)Sfi21 fragment from pMZ23c was subsequently cloned in the low-copy-number vector pNZ121, yielding the construct pSF24. Of the four phages (<\(\phi\)Sfi21, <\(\phi\)S69, <\(\phi\)S89, and <\(\phi\)Sl13J) tested by plaque assays, pSF24 protected Sfi1 from infection by <\(\phi\)Sfi21 only (i.e., no plaques observed). This clearly indicates that to be effective against a larger range of phages, the <\(\phi\)Sfi21-derived DNA fragment must be provided at a high copy number.

Effect of bacteriophage resistance determinant on plasmid DNA replication

The relative amount of <\(\phi\)Sfi21 DNA was determined throughout the phage lytic cycle by comparing the total DNA isolated from phage-infected cells harvested at different times after infection. Using DNA probes specific for the phage or the vector pNZ121, it was possible to simultaneously monitor both phage and plasmid DNA. The high-copy-number clone pMZ23c had a dramatic effect on <\(\phi\)Sfi21 as demonstrated by the amount of <\(\phi\)DNA accumulated in the strain Sfi1(pNZ124) compared with strain Sfi1(pMZ23c) after phage infection. In the phage-sensitive strain Sfi1(pNZ124), <\(\phi\)Sfi21 DNA was detectable 20 min after infection, increasing to a maximum at 40 min and followed by a reduction at 60 and 80 min (Fig. 2B). For Sfi1 containing the pMZ23c clone, <\(\phi\)DNA was detected 20 min after infection, but in contrast to Sfi1(pNZ124), no further accumulation of <\(\phi\)DNA was observed (Fig. 2B). A similar pattern was observed for Sfi1(pSF23g) (data not shown). When <\(\phi\)DNA was monitored in Sfi1(pSF24), a pattern similar to the phage-sensitive control Sfi1(pNZ124) was observed, but the amount of DNA accumulated 40 min after phage infection was reduced with Sfi1(pNZ124) (Fig. 2A).

Plasmid replication driven by bacteriophage infection

To examine the effect, if any, of phage infection on the replication of a plasmid carrying the <\(\phi\)Sfi21-derived DNA fragment, Sfi1 containing the low-copy-number clone pSF24 was chosen. A significant increase in the pSF24 copy number was observed when the quantity of pSF24 DNA present in uninfected Sfi1(pSF24) was compared with that of <\(\phi\)Sfi21-infected cells, 40 min after infection (Fig. 3). Although <\(\phi\)Sfi21 infection apparently boosted pSF24 plasmid DNA replication, infection with the heterologous phage <\(\phi\)S69 had no observable effect on pSF24 copy number (Fig. 3). This correlates with the plaque assay results that indicated the high-copy-number clone pMZ23c had an inhibitory effect on a large number of phages, including <\(\phi\)Sfi21 and <\(\phi\)S69, whereas the low-copy-number clone pSF24 was effective against only <\(\phi\)Sfi21.

The simultaneous monitoring of phage (Fig. 2) and plasmid (Fig. 3) DNA indicated that although the accumulation of phage DNA was inhibited, the replication of a plasmid containing the <\(\phi\)Sfi21-derived insert was stimulated on phage infection. Apparently, the cloned <\(\phi\)Sfi21 DNA serves as an ori on the plasmid driven by phage-encoded proteins. The experimental data therefore suggest a working hypothesis wherein this noncoding <\(\phi\)Sfi21-derived DNA serves as a binding site(s) for phage proteins which, when multiple copies are present, are capable of titrating essential phage proteins such that they are no longer available for phage DNA replication.

Bacteriophages have adopted a plethora of replication strategies that differ mainly in the method of priming and the dependence on host- and phage-encoded proteins. Among the phages of gram-positive bacteria, only the *Bacillus* phages <\(\phi\)SPP1 (Pedré et al., 1994) and <\(\phi\)29 (which uses protein-primed replication; Salas et al., 1995) have been extensively characterised in terms of their replication functions. In contrast, although considerable sequence data exist for phages of lactic acid bacteria, replication origins have been identified only for the lactococcal phages <\(\phi\)ek1 (Chandry et al., 1997), <\(\phi\)c2 (*Waterfield et al., 1996*), and <\(\phi\)e0 (*Hill et al., 1990; O'Sullivan et al., 1993*). The putative ori of the lactococcal <\(\phi\)31 has been functionally identified, although no sequence data are available (O'Sullivan et al., 1993). As yet, no information is available concerning the general mechanisms of replication or the phage and host proteins required. All of the putative origins of lactic acid bacteria phages are rich in direct repeat sequences, but there is no apparent correlation in the size, relative location, or sequence of the repeated elements between lactococcal phages and those of *S. thermophilus*. In this report, we describe a
VIRAL DNA ELEMENT CONFERS RESISTANCE TO S. thermophilus

A

![Agarose gel electrophoresis of total DNA isolated at various time points during phage \(\phi\)Sfi21 infection](image)

The cultures tested were \(\phi\)Sfi containing pNZ124 (lanes 6a–f), pMZ23c (lanes 2a–f), and pSF24 (lanes 4a–f). a–f represent total DNA isolated before phage infection, immediately after infection, and 20, 40, 60, and 80 min after infection, respectively. Lane 1 contains \(\phi\)Sfi21 DNA. Lanes 3, 5, and 7 contain pMZ23c, pSF24, and pNZ124 plasmid DNA, respectively. All DNA samples were digested with Xba1. The size of the \(\lambda\) HindIII DNA molecular weight marker (lane m) is indicated in kilobases.

B

![Hybridisation of the samples indicated in panel A using labeled A and \(\phi\)Sfi21 DNA as probes](image)

sequence and functional symmetry for the \(\phi\)Sfi21-derived origin fragment. The observed symmetry is, however, very different than that of the 500-bp \(\phi\)Phi origin of replication identified by Hill et al. (1990), which consists of a 236-bp direct repeat separated by 28 bp and for which only one phage-inhibitory determinant was described.

**S. thermophilus** plasmid pST1 does not mediate bacteriophage resistance

The selection pressure maintaining cryptic plasmids in lactic acid bacteria has not been defined. Because the \(\phi\)Sfi21-derived DNA described above shows 80% identity to the putative single-stranded origin of the cryptic S. thermophilus plasmid pST1 (Lanjzen et al., 1992), it may be interesting to examine whether a phage-inhibitory activity could be associated with the pST1 sso. For this purpose, the sso region of pST1 was PCR amplified and cloned in pNZ124. The presence of the cloned putative sso from pST1 in S. thermophilus \(\phi\)Sfi1 did not protect the host from infection by phages in our collection or those in Kiel, from which the pST1-containing S. thermophilus strain was obtained (H. Neve, personal communication).

**Binding of phage-specific complex to putative \(\phi\)Sfi21 replication origin**

The model described above for phage inhibition depends on a specific interaction between a phage-encoded protein(s) and the putative ori of \(\phi\)Sfi21. To test this, gel retardation assays were performed using a labeled 302-bp DNA fragment corresponding to the phage-inhibitory DNA from \(\phi\)Sfi21 and crude cell extracts prepared from \(\phi\)Sfi21-infected and uninfected \(\phi\)Sfi1 cells. \(\phi\)Sfi21 infection resulted in a DNA fragment of lower mobility in polyacrylamide gels (Fig 4). The specificity of the protein–DNA interaction was demonstrated by competition experiments in which the presence of a 500-fold excess of the unlabeled DNA fragment abolished the retardation (data not shown). Interestingly, when the same experiment was repeated using crude cell extracts...
FIG. 3. Hybridisation of XbaI-digested total DNA isolated from SF1(pSF24) without phage infection (A) and infected with \( \phi \)SF21 (B), \( \phi \)S69 (C), and \( \phi \)SF19 (D). The probe consisted of labeled \( \lambda \) and pNZ21 DNA. 1a-1f represent total DNA isolated at various time points: before phage infection, immediately after infection, and 20, 40, 60, and 80 min after infection, respectively. Lane 2 contains XbaI-digested pSF24 plasmid DNA. The size of the \( \lambda \)-HindIII DNA molecular weight marker (lane m) is indicated in kilobases. Measurements obtained of OD_{600nm} at various time points are also indicated.

Prepared from \( \phi \)SF19-infected SF1, no retarded protein-DNA complex was observed (Fig. 4). This observation is significant because the \( \phi \)SF21-derived DNA has no inhibitory effect on infection by \( \phi \)SF19.

Sequence comparison of the \( \phi \)SF19 and \( \phi \)SF21 DNA replication modules

Although pMZ23c is effective against a range of phages, eight of the phages tested in this study, including \( \phi \)SF19, are insensitive to the \( \phi \)SF21-derived DNA (Table 1). Comparative sequencing of a 17-kb region of \( \phi \)SF21 and \( \phi \)SF19 encompassing the lytic cassette and the structural gene cluster revealed that these two phages are similarly organised (Desiere et al., 1998). Therefore, it was of interest to extend this comparison to the region encompassing the DNA replication module (previously defined for \( \phi \)SF21 by Desiere et al., 1997) because it may facilitate the identification of the genetic determinant or determinants responsible for the insensitivity of \( \phi \)SF19 to the \( \phi \)SF21-derived DNA element. \( \phi \)SF19 revealed an identical organisation of ORFs and a high degree of sequence similarity to \( \phi \)SF21 (Table 3). An interesting observation was the near absence of silent point mutations (Table 3).

A sequence alignment of the inhibitory \( \phi \)SF21 DNA element with the corresponding intergenic region of \( \phi \)SF19 revealed an overall nucleotide identity of 95% (i.e., 14-nt differences over a stretch of 302 nt; Fig. 5). When this comparison was extended to eight additional phages, a clear division emerged in which phages sensitive or insensitive to the \( \phi \)SF21-derived DNA yielded a sequence resembling that of \( \phi \)SF21 and \( \phi \)SF19, respectively (Fig. 5). Exclusion of the polymorphisms within the \( \phi \)SF21 group identified nine nucleotide differences between the \( \phi \)SF21- and \( \phi \)SF19-type sequences. An extension of this sequence comparison to a broader range of phages may lead to the identification of the critical nucleotides determining the
VIRAL DNA ELEMENT CONFFRS RESISTANCE TO S. thermophilus

FIG. 4. Gel retardation assay using the 32P-labeled 302-bp phage-inhibitory DNA fragment from ϕSF21. Crude cell extracts were prepared from S. thermophilus Sfi1 infected with ϕSF21 (lane 2) and ϕSF19 (lane 3). Lane 1 is a control lane in which a crude cell extract prepared from uninfected Sfi1 cells was used.

specificity of the protein-DNA interactions at the phage putative replication origin.

The differences between the putative origins could reflect the molecular basis for the insensitivity of ϕSF19 to the ϕSF21 origin. To test this, the noncoding DNA region of ϕSF19, located downstream of orf 504, was PCR amplified and cloned in pNZ124, resulting in the construct pSF19. Sfi1(pSF19) was found to be resistant to ϕSF19 infection. Sfi1(pSF19) was then tested against the remaining seven phages which were insensitive to the ϕSF21 origin-mediated phage resistance. Paradoxically, the ϕSF19-derived fragment in pSF19 protects Sfi1 from infection by ϕSF21 and the other phages inhibited by the ϕSF21 DNA element, whereas in contrast, the ϕSF21-derived fragment does not protect against ϕSF19. Apparently, a ϕSF21 protein is capable of interacting with both types of origin, whereas the equivalent ϕSF19 protein is origin specific. In phages, DNA sequences targeted by proteins are frequently found directly in the vicinity of the gene coding for the protein in question. The genes preceding orf 504 in the DNA replication module are unlikely to code for the origin-binding protein because all of the amino acid replacements in ϕSF19 proteins (compared with ϕSF21) were also detected in proteins predicted for ϕSF11 or ϕSF18 (Table 3), both of which are susceptible to the ϕSF21-derived inhibitory activity. Interestingly, the orf 504 bp of ϕSF21 and ϕSF19 differ in only one amino acid position in which a C-terminal lysine residue is substituted by a glutamic acid residue in ϕSF19. The difference between a basic and acidic residue may have consequences for protein-folding. To explain the complex inhibitory activity of the different origins will require a sequence comparison (both nucleotide and amino acids) of a broader range of phages together with the elucidation of the specific DNA-protein interaction(s) and the identification of the components involved.

MATERIALS AND METHODS

Bacterial strains, bacteriophages, and media

The E. coli strain XL1-Blue was propagated in LB broth or on LB broth solidified with 1.5% (w/v) agar (Sambrook

<table>
<thead>
<tr>
<th>orf gp of ϕSF19</th>
<th>Size (aa) of equivalent orf gp in ϕSF11</th>
<th>No. of aa (n) changes between ϕSF21/ϕSF19</th>
<th>No. of aa (n) changes between ϕSF21/ϕSF11</th>
<th>No. of aa changes specific to ϕSF21/ϕSF19^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>157</td>
<td>157</td>
<td>2 (7)</td>
<td>2 (7)</td>
<td>0</td>
</tr>
<tr>
<td>233</td>
<td>233</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td>443</td>
<td>443</td>
<td>0 (0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>151</td>
<td>151</td>
<td>0 (0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>271</td>
<td>271</td>
<td>4 (5)</td>
<td>2 (3)</td>
<td>0</td>
</tr>
<tr>
<td>504b</td>
<td>504b</td>
<td>1 (2)</td>
<td>0</td>
<td>1^c</td>
</tr>
<tr>
<td>143</td>
<td>143</td>
<td>0 (0)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

^a ϕSF21 predicted gene products were compared to those of ϕSF11 and ϕSF18 in order to identify the amino acid (aa) replacements specific to the ϕSF21/ϕSF19 comparison.

^b Several start codons are possible for the putative primase gene. The one with the most suitable ribosome binding site (RBS) is UUG at position 3758 giving an orf. orf 504 as was selected for ϕK1205 (Stanley et al., 1997).

^c Comparison of ϕSF21 and ϕSF19 orf 504 gp revealed a Lys to Glu transition (Fig. 5).
<table>
<thead>
<tr>
<th></th>
<th>Lys/Glu</th>
<th>stop 504</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sfi11</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
<tr>
<td>Sfi17</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
<tr>
<td>Sfi12</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
<tr>
<td>Sfi19</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
<tr>
<td>Sfi18</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
<tr>
<td>Sfi1</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
<tr>
<td>Sfi21</td>
<td>GAAACCGAGG CCGAGAGGAG AGCAGTAGTG GTTATGTTA CTCAGTAACC GCAAGTTATT GCAACACGA</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 5. Nucleotide sequence alignment of the phage-inhibitory sequence derived from Sfi1 and transformants thereof with the equivalent sequence from additional S. thermophilus phages. Sequence differences between the Sfi1- and Sfi19-type sequences are in bold. The single amino acid difference (Lys/Glu) in orf 504 gp is indicated. The region corresponding to the 302-bp phage-inhibitory fragment is indicated by arrows.

et al., 1989) at 37°C under agitation. S. thermophilus strain Sfi1 and transformants thereof were routinely subcultured at 42°C in either LM17 (M17 supplemented with 0.5% lactose) (Terzaghi and Sandine, 1975) or Belliker (Elliker plus 1% beef extract) media. Lactococcus lactis MG1363 was propagated in GM17 (M17 supplemented with 0.5% glucose). Chloramphenicol was used when required at a final concentration of 3, 5, and 20 μg/ml, respectively, for S. thermophilus, Lactococcus, and E. coli.

The S. thermophilus phages used in this study were obtained from the Nestlé phage collection. The phages were propagated on their appropriate S. thermophilus strain in LM17 broth as described previously (Brüssow et al., 1989).
100

VIRAL DNA ELEMENT CONFERS RESISTANCE TO *S. thermophilus* 385

and Bruttin, 1995; Bruttin and Brüssow, 1996). Phage enumeration was achieved by plaque assay using LM17 agar supplemented with 0.25% (w/v) glycine and 0.1% (w/v) skim milk, to enhance plaque formation, and MRS top agar.

**DNA techniques**

Phage purification, DNA extraction and purification, agarose gel electrophoresis, Southern blot hybridisation, and DNA labelling were done as described previously (Brüssow and Bruttin, 1995; Bruttin et al., 1997a; Bruttin and Brüssow, 1996). General DNA techniques were performed as described by Sambrook et al. (1989). The Qiaprep plasmid kit (Qiagen) and the Jetstar Plasmid Maxi-kit (Genomod) were used for the rapid isolation of plasmid DNA from *E. coli*. Restriction enzymes and T4 DNA ligase were obtained from Boehringer-Mannheim and used according to the supplier’s instructions. *E. coli* was electrotransformed as outlined in the BioRad instruction manual. *Lactococcus* and *S. thermophilus* were electrotansformed using the procedures described by Holo and Nes (1989) and Slos et al. (1991), respectively.

**PCR**

DNA samples were amplified in a Perkin-Elmer thermal cycler programmed for 30 cycles, each consisting of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min. Synthetic primers were designed according to the established *φSfi19* and *φSfi21* DNA sequences and used together with the relevant DNA template and Taq polymerase Fermentas. PCR products were gel-purified using Ultrafree-MC Centrifugal Filter Units (Millipore) and according to the manufacturer’s instructions.

**DNA sequencing and analysis**

DNA was sequenced on both strands by the Sanger method of dideoxy-mediated chain termination using the *fmo1*™ DNA Sequencing System of Promega (Madison, WI). The sequencing primers were end-labeled with [γ-33P]ATP according to the manufacturer’s instructions. 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. The sequence obtained was analysed as previously described by Desiere et al. (1998). The relevant accession numbers for *φSfi19* and *φSfi21* sequences are AF004379 and AF077306, respectively.

**Construction of plasmids**

The cloning vectors used in this study were as follows: the high-copy-number *E. coli*-lactococcal-streptococcal shuttle vector pNZ124 (Platteauw et al., 1994), the low-copy-number shuttle vector pNZ121 (de Vos and Simons, 1994), and the *S. thermophilus* integration vector pSF28 (Bruttin et al., 1997b).

pMZ23c consists of an 824-bp *EcoRI* fragment from the *φSfi21* DNA replication module cloned in pNZ124. The pMZ23c subclones pSF23d, pSF23e, and pSF23f were constructed by deleting the *PvuII-MscI* (isochizomer of *BamH I*), *PvuII-BstI*107I, and *MscI-EcoRI*136II fragments, respectively, from pMZ23c.

pSF23g consists of a 302-bp *EcoRI-BgII* PCR fragment cloned in the high-copy-number vector pNZ124. The PCR fragment was generated using pMZ23c as the template DNA and primers (5'-'GCG AAT TCA GCA GTA GTG GTT ATG G-3' and 3'-'GGA GAT CTA AGT ATT GAA TTG AAC C-3') containing *EcoRI* or *BglII* restriction sites (underlined).

The low-copy-number clone pSF24 was constructed by ligating the 876-bp *BglII-XhoI* fragment from pMZ23c to *BglII-SalI*-digested pNZ121 DNA. For chromosomal integration, the 834-bp *XbaI-PstI* fragment from pMZ23c was cloned in the equivalent sites of the *S. thermophilus* integration vector pSF28, thereby generating the construct pSF20.

pSF191 consisted of a 352-bp *EcoRI-BgII* digested PCR fragment cloned in the equivalent sites of the high-copy-number vector pNZ124. The PCR fragment was generated using *φSfi19* DNA as a template and primers (5'-'GCG AAT TCA GGA CGA AAA CGA GGC GG-3' and 3'-'GCA GAT CTC ATT AGG TTC GTG TTC TTG-3') containing *EcoRI* or *BglII* sites (underlined) to facilitate cloning.

**Analysis of intracellular phage DNA**

Bacterial cultures were grown at 40°C in Belliker broth to an OD of 0.2. CaCl₂ was added to a final concentration of 10 mM, followed by the addition of the relevant bacteriophage at a multiplicity of infection (m.o.i.) of >2. Samples (1 ml) were removed at regular intervals: before infection, immediately after infection, and at 20, 40, 60, and 80 min after infection. The samples were rapidly centrifuged, and the pellets were frozen by immersion in ultracold ethanol (<−50°C). After thawing, the pellets were resuspended in 1 ml of lysis buffer (6.7% sucrose, 50 mM Tris, 1 mM EDTA, pH 8.0) containing 10 mg/ml lysozyme and incubated for 1 h on ice. SDS was added (125 μl of 10% stock solution) together with proteinase K (50 μl of 20 mg/ml stock) and the lysates were incubated at 65°C for 30 min. RNase A (50 μg/ml) was subsequently added, and incubation was continued at 65°C for an additional 30 min. Two phenol-chloroform extractions were performed on the resulting lysates, followed by an ethanol precipitation. The DNA pellet obtained was washed in 70% ethanol and finally resuspended in 50 μl of distilled water, of
which 25 μl was used for restriction analysis. Restricted DNA samples were electrophoresed, and by hybridisation of Southern blots, the amount of phage DNA was assessed at various time points during infection.

Gel retardation assay

An overnight culture of *S. thermophilus* SfI1 was inoculated (2%) in Belliker broth and grown to an OD<sub>600</sub> of 0.2. CaCl<sub>2</sub> was then added (final concentration, 10 mM), and the culture was aliquotted in 50-ml volumes. The cultures were subsequently infected with the relevant phage at an m.o.i. of 2 and incubated at 38°C for 30 min. The cells were harvested by centrifugation and washed, and the pellets resuspended in 750 μl of Tris buffer (10 mM pH 8, 4°C). The crude cell extracts were prepared as described by Foley et al. (1996).

A 302-bp BglII-digested PCR-generated DNA fragment corresponding to the phage DNA insert in pSF23g was end-labeled with the Klenow fragment of DNA polymerase I (New England Biolabs) in the appropriate buffer and in the presence of 30 μCi of [α<sup>32</sup>P]ATP. The labeled DNA was phenol extracted and ethanol precipitated. The binding conditions used were as described by Foley et al. (1996) and included 11 μl of crude cell extract (~5 μg of total protein). After a 10-min incubation at room temperature, 5 μl of 50% glycerol was added to each reaction before immediate loading on a 4% non-denaturing polyacrylamide gel containing 2.6% glycerol. Electrophoresis was performed at room temperature for 4 h at 10 W/cm. The gel was vacuum dried onto 3MM Whatman paper and autoradiographed for 36 h.

**ACKNOWLEDGMENTS**

We thank Josette Sidoti for invaluable technical assistance and the Swiss National Science Foundation for financial support of Sophie Foley and Sacha Lucchini in the framework of its Biotechnology Module (Grant 5002-044545/1). We thank Beat Mollet and David Pridmore for important contributions in the initial phase of the work.

**REFERENCES**


Bruttin, A., Foley, S., and Brüssow, H. (1997b). The site-specific integra-
and Hermosa, J. M. (1999). Protein-nucleic interactions in bacterio-
Spring Harbor, New York.
characterisation of chromosomal promoters of *Streptococcus salivarus 
Stanley, E., Fitzgerald, G. F., Le Marec, C., Fayard, B., and van Sinderen, 
D. (1997). Sequence analysis and characterization of fd01205, a 
temperate bacteriophage infecting *Streptococcus thermophilus 
813.
Waterfield, N. R., Lubbers, M. W., Polzin, K. M., Le Page, R. W. F., and 
2.6 Construction of food-grade *Streptococcus thermophilus* starters with potent, broad range phage resistance phenotypes
Construction of Food-Grade Streptococcus thermophilus Starters with Potent, Broad Range Phage Resistance Phenotypes

Sacha Lucchini and Harald Brüssow

Nestlé Research Centre, Nestec Ltd., Vers-Chez-Les-Blanc, CH-1000 Lausanne 26, Switzerland

Introduction

*S. thermophilus* is a thermophilic starter used in the production of various dairy products such as yogurt, Italian- and Swiss-type cheese. Bacteriophages represent the main cause of slow fermentation or even complete starter failure. To cope with the phage problem, methods currently used are rigorous sanitation of the dairy plant, use of multiple strain starters in rotation of cultures and closed production lines (12). In *Lactococcus lactis*, the major mesophilic starter of the dairy industry, natural phage-resistance mechanisms are abundant and usually encoded on conjugative plasmids (1). These have been used to construct food-grade phage resistance mechanisms into important industrial lactococcal starters (22).

In *S. thermophilus* little is known about natural phage resistance mechanisms (14,19). This is partially the consequence that plasmids are scarce in *S. thermophilus* (17). Spontaneous phage-insensitive mutants (BIM) can be selected by phage challenge. This approach has been applied to *S. thermophilus*. However, such mutants are usually slow acid producers and can also revert to phage sensitivity (4). Thus, it was decided to exploit molecular methods to create phage-resistant strains. One approach based on the random cloning of phage DNA segments on a plasmid has been successful. One fragment was shown to interfere with phage DNA replication and to provide effective protection against a broad range of phages (10).

To increase the range of strategies available to obtain phage-resistant strains, we tested two further approaches. First, we targeted bacterial host genes for inactivation. The rationale was that we postulated the presence of bacterial genes essential to phage development, but dispensable for bacterial growth in milk. In fact, bacteriophages depend on host factors in many steps of their life cycle, e.g. adsorption, DNA injection, replication and morphogenesis. Disruption of one of these factors will lead to a phage-resistant cell. Having no knowledge of the *S. thermophilus* genome we opted for a random gene inactivation approach relying on the temperature sensitive plasmid pG*host*9ISS (16). The method is self-selective:
after phage challenge nonresistant cells will be eliminated and mutants leading to decreased bacterial fitness will be outnumbered.

The second approach relied on the protection provided by the Sfi21 prophage (8). Temperate phages usually code for superinfection immunity genes, which are quite effective in protecting the lysogen from superinfection by both temperate and lytic phages. In the case of phage Sfi21 the superinfection control is apparently mediated by two distinct genetic elements: orf 203, a superinfection immunity gene and orf 127, the phage repressor. The drawback is that Sfi21 lysogens continuously release infectious phage particles in the media, which would contaminate the factory and possibly prevent a later introduction of valuable starters susceptible to phage Sfi21. It has also been shown that temperate phages can be the source of lytic derivatives (3;7;23). Therefore, we decided to create targeted deletions in the Sfi21 prophage inserted in the lysogenic starter Sfi1cl6 to obtain a lysogen that would retain superinfection control, but has lost its capacity to produce infectious virions.

Here we demonstrate that both approaches provide powerful phage-resistant food-grade starters for industrial milk fermentation.

**Materials and methods**

**Strains, media, plasmids, and culture conditions.** The E. coli strain 101 was propagated in LB broth or LB broth solidified with 1.5% (W/V) agar at 37°C. Liquid cultures were grown under agitation (240 rpm). S. thermophilus strains Sfi1, its lysogenic derivative Sfi1cl6 (containing the Sfi21 prophage) and their transformants were cultivated at 42°C either in M17 supplemented with 0.5% lactose (LM17), Belliker media or MSK. Erythromycin was used when required at a final concentration of 2 and 150 μg/ml for S. thermophilus and E.coli, respectively. The phages used in this study were obtained from the Nestlé collection and propagated on their appropriate S. thermophilus strain in LM17 broth as described previously (5;7). Phage enumeration was achieved as described in (10). For random insertion mutagenesis and directed mutagenesis, plasmid pG’host9ISSf1 and pG’host9 have been used (16), respectively.

**DNA techniques.** Phage purification, DNA extraction and purification, agarose gel electrophoresis, Southern blot hybridization, and DNA labeling were executed as described previously (5;7;8). The Qiaprep plasmid kit (Qiagen) was used for the rapid isolation of plasmid DNA from E. coli. Restriction enzymes and T4 ligase were purchased from Boehringer-Mannheim and used according to the supplier’s instructions. E. coli was electrotransformed as outlined in the BioRad instruction manual. S. thermophilus was electroporated using the procedure described by Slos et al. (25). The analysis of intracellular phage DNA, PCR and DNA sequencing have been performed as described previously (10).

**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, release 56; PIR-Protein, release 57; SWISS-PROT, release 36; and PROSITE, release 15.0) with FastA(15) and BLAST programs (2). Prediction of transmembrane domains was performed using the TMpred program (11).

**Construction of plasmid for site directed integration.** To create a deletion in the Sfi21 prophage, the thermosensitive plasmid pG\(^+-\)host9AB1560, a derivative of pG\(^+-\)host9 has been created. Two fragments of approximately 500 bp, A and B, were chosen in the orf1560 on the prophage sequence at a distance of 2.4 kb so that a deletion of the same size would be created by homologous recombination. Fragment A was generated by PCR using phage Sfi21 DNA as a template and primers (5'-AAC TGC AGT CTC AGC TCA AAG GGA C-3' AND 5'-GGA ATT CTA GCC GTG ATG TTT TTG-3') containing PstI and EcoRI restriction sites (underlined) Fragment B was generated by PCR using primers (5'-GGA ATT CGA CGC AAT TAA AGA CCC-3' AND 5'-CCA TCG ATC TGC TTC CAA AAT CTC G-3') containing EcoRI and ClaI restriction sites. Both clones were then cloned into pG\(^+-\)host9, so to be adjacent, generating the construct pG\(^+-\)host9AB1560. The construct was first generated into *E. coli* 101, then transformed into *S. thermophilus* Sfi1cl6.

**Transposition of pG\(^+-\)host9ISS1 and pG\(^+-\)host9AB1560 in the *S. thermophilus* chromosome.** *S. thermophilus* Sfi1 and Sfi1cl6 containing pG\(^+-\)host9ISS1 and pG\(^+-\)host9AB1560, respectively, were grown overnight in LM17 medium supplemented with 2 µg/ml of erythromycin. The saturated cultures were diluted 100-fold in LM17 medium containing 1 µg/ml of erythromycin and incubated 2 h at 30°C. The cultures were then shifted to 42°C to eliminate free plasmids and grown to saturation. The frequency of integration per cell was estimated as the ration of the number of Em\(^R\) cells at 42°C to the number of viable cells at 30°C. Integration of the plasmids was checked by Southern blot hybridization. To excise the transposed vectors, serial passages have been performed in LM17 broth without antibiotic.

**Selection of phage-resistant mutants.** The culture containing the original population of Sfi1::pG\(^+-\)host9ISS1 integrants was diluted 100-fold in LM17 supplemented with 2 µg/ml of erythromycin and challenged with lytic phage Sfi19 at a M.O.I. of 5. The culture was then grown to saturation. The experiments were considered unsuccessful when no growth was observed after 48h.

**Phage adsorption test.** *S. thermophilus* cultures were grown in LM17 until an OD\(_{600}\) of 0.6 was reached. Then phages were added at a M.O.I. of 1 and the cultures incubated at room temperature. Probes were taken immediately after phage addition and after 30 min. These probes were then filtered to remove the bacterial cells from the cultures. Phages left were then enumerated. The adsorption was calculated as phage counts at time 30 divided by the phage counts at time zero.
Results

Random mutagenesis of *S. thermophilus* Sfi1. Starter strain Sfi1 is our best indicator cell. It is susceptible to about 25 of the 100 phages from our collection. This allows testing mutant starters against a broad range of phages.

Transposition of the plasmid could be achieved with a very high integration frequency, about 50% (see materials and methods). Integrants were randomly chosen on agar-plates and grown in liquid cultures. Their total DNA was then extracted to check by Southern blot hybridization the integration of the plasmid into the chromosome. Hybridization was performed on *EcoRI* digested chromosomal DNA using labeled pG+host9ISS1 as probe. All integrants showed signals corresponding to distinct chromosomal fragments (data not shown), confirming the integration of the plasmid and the randomness of the transposition.

Selection of phage-resistant mutants. Integrants were challenged with the lytic phage Sfi19 (M.O.I.=5) to select for phage-resistant mutants. In presence of phage Sfi19 the Sfi1::pG+host9ISS1 culture showed a delayed growth, an OD_{600} of 0.85 was only reached after 8 to 12 h of incubation in comparison with 3 h for the transformants in the absence of challenge phage. In contrast, no growth was observed for the parental Sfi1 starter after phage challenge (OD_{600} < 0.02 after 48 h). In fact, in LM17 medium we consistently failed to obtain phage resistant mutants of starter Sfi1 after challenge with numerous phages.

![Figure 1](image-url)

**Figure 1.** Southern blot of total DNA of phage resistant integrants. Lanes 1-5: phage resistant Sfi1::pG+host9ISS1. Lane 6: Sfi1. M: DNA marker (λ-DNA x HindIII). The blot was probed with radioactively marked pG+host9ISS1 plasmid and λ-DNA. Note: this figure is the combination of two independent blots (lane 1 has been added).

Individual phage-resistant colonies were isolated; their total DNA was extracted, and digested with EcoRI, and Southern blots were performed using the plasmid DNA
as probe. Three different hybridization patterns were identified, corresponding to the Sfi1-R7 (lane 3), Sfi1-R24 (lane 1) and Sfi1-R71 (lanes 2,4,5) insertion mutants (Fig. 1).

Characterization of the phage-resistant mutants. The 3 mutants were tested for phage resistance against 15 different S. thermophilus phages. In no case, phage plaques were observed suggesting in some cases efficiency of plating of < 10^-7 (Table 1).

<table>
<thead>
<tr>
<th>Phage</th>
<th>Sfi1 (pfu/ml)</th>
<th>Sfi1-R7 (pfu/ml)</th>
<th>Sfi1-R24 (pfu/ml)</th>
<th>Sfi1-R71 (pfu/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sfi21</td>
<td>6 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S69</td>
<td>7.2 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>Sfi3J</td>
<td>2 X 10^4</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S44A</td>
<td>1 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S19</td>
<td>2 X 10^4</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S96</td>
<td>4 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S44</td>
<td>1 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S40</td>
<td>2 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>Sfi18</td>
<td>2.7 X 10^7</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>Sfi19</td>
<td>1.1 X 10^6</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S25</td>
<td>3 X 10^6</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S17</td>
<td>1 X 10^5</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>S17</td>
<td>6 X 10^5</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>H</td>
<td>4.5 X 10^3</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
<tr>
<td>F</td>
<td>1 X 10^4</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
<td>&lt;10^2</td>
</tr>
</tbody>
</table>

Table 1

To have some indications at what stage the phage development is blocked, phage-adsorption and phage DNA-replication tests were performed. The adsorption test was performed using Sfi19 on Sfi1 and the three mutants. After 30', 90% of the input phages adsorbed to Sfi1, 94% to Sfi1-R7, 91% to Sfi1-R24 and 87% to Sfi1-R71. This indicates that the mutation did not affect phage adsorption.

DNA replication of phage Sfi19 in Sfi1 and the mutants was analyzed by determining the relative amount of intracellular phage DNA at different times during phage infection. In Sfi1 phage DNA was readily detected 20 min after infection (Fig. 2). The amount of DNA increased to a maximum after 40 min. Then a decrease was observed after 60 and 80 min, probably because of phage induced cell lysis. Phage DNA replication in the Sfi1-R24 mutant was delayed and decreased.
Phage DNA was detected only after 40 min reaching a maximum at 60 min. In the case of Sfi1-R7 and Sfi1-R71 no phage replication was observed. A very low and invariable level of phage DNA was detected by hybridization. This could represent the injected DNA of the infecting input phages or DNA from uninjected, but adsorbed input phages. The growth of the mutants was compared to that of the wild type strain in milk. This was done by measuring changes in impedance of the culture media (Rapid Automated Bacterial Impedance Technique, Don Whitley Systems). This system measures indirectly the transformation of a weak electrical conductor like the polar, but uncharged lactose into the electrically charged lactic acid during bacterial growth. The resulting curve (time vs. conductivity) can be correlated with both bacterial growth and acidification of the culture. All three mutants did not show significant differences in growth (and acidification) to the parental strain Sfi1 (Fig. 3).
Determination of the site of integration. The fragments adjacent to the insertion point were obtained by plasmid rescue as described by Maguin et al. (16) for the Sfi1-R7 and Sfi1-R24 mutants. The approach was unsuccessful for Sfi1-R71. The regions flanking the pG+host9ISS7 insertion point were sequenced for the two mutants (Fig. 4).

In Sfi1-R7 the plasmid integrated in the last 4 amino acids of a putative chorismate mutase chain A gene (PheA), orf 90. This site of integration could also function as promoter region for the downstream orf 394. Orf 394 gene product shows similarities (26% identity) to a hypothetical conserved protein of Methanococcus jannashii (Accession number MJ0305). Orf 394 gp may be an integral protein since 9 strong transmembrane helices were predicted. Orf 115 gp encoded by a gene located directly downstream of orf 394, showed significant similarity to the ribosomal protein L19 from a number of bacteria (Haemophilus influenzae, Salmonella typhimurium, Serratia marcescens, Synechocystis sp.). A gene almost identical to a tRNAArg gene (anticodon CCU) follows this gene from E. coli and interestingly the phage attachment site attB for prophage integration. Upstream of orf 90 a gene we identified orf 486 (see Fig. 4), which encodes a further possible membrane protein. The deduced protein for this orf shows significant similarity (50% over 427 aa) to a hypothetical Cl-channel-like protein from E. coli (Accession number P37019).

Figure 4: Sites of pG+host9ISS7 insertion in Sfi1-R7 and Sfi1-R24. Open arrows indicate the predicted open reading frames (orf). The numbers indicate the number of codons for each orf. When possible, a tentative function has been attributed to the predicted gene products.
In Sfi1-R24 the pG\textsuperscript{+}host9ISS\textsuperscript{1} insertion point was in a likely oxidoreductase involved in fatty acid biosynthesis (orf 269). The predicted protein showed 42% amino acid identity with an \textit{E. coli} 3-oxoacyl-[acyl-carrier protein] reductase. Interestingly, one fragment of an R-subunit of a type I restriction enzyme is located downstream orf 269. No database matches were found for the orfs upstream of orf 269.

**Plasmid excision.** To obtain food-grade starter strains it is necessary to remove the erythromycin resistance gene present in the transformants. ISS\textsuperscript{1} undergoes replicative transposition which leads to an integrated plasmid vector flanked by two IS elements. Therefore, homologous recombination between this two copies of the IS element will entirely remove the plasmid and leave only a single copy of the IS element. To favor recombination, serial passages of the transformants were done in absence of antibiotic selection. Sfi1-R24 lost the Em\textsuperscript{R} phenotype after 29 passages (now Sfi1-R24e), Sfi1-R71 after 40 (Sfi1-R71e). Excision was easier with Sfi1-R24, because of the lack of tandem repeats (Fig. 1). Sfi1-R7 Em\textsuperscript{S} (Sfi1-R7e) could be obtained after 60 passages. Then, phage resistance was checked for the Em\textsuperscript{S} mutants. In 2 cases (Sfi1-R24 and Sfi1-R71) the loss of the plasmid resulted in full reversion to phage sensitivity while Sfi1-R7e did retain its full phage resistance phenotype. This derivative lacked the plasmid sequence except for one ISS\textsuperscript{1} sequence. Sequencing of the flanking sequences of the remaining ISS\textsuperscript{1} showed in addition a 69 bp deletion which started at the IS transposition site and removed 3'-end of orf 90 and the first 15 codons of orf 394 (Fig. 5). Since two independently obtained Sfi1-R7e mutants showed the same deletion it is likely that the deletion has already occurred in the parental Sfi1-R7 mutant.

---

**Deletion of orf 1560 from the Sfi21 prophage.** Integration of pG\textsuperscript{+}host9AB1560 into the prophage was readily obtained (see materials and methods). Integration in the right site was checked by PCR. More than 50% of the integrants were shown to

---

**Figure 5:** Deletion in Sfi1-R7e. The sequence deleted in Sfi1-R7e is in bold. The arrow indicates the pG\textsuperscript{+}host9ISS\textsuperscript{1} integration site.
give a signal corresponding to the predicted size for homologous recombination events. Plasmid excision was obtained after 20 serial passages. Because the second homologous recombination can lead both to the wild type or the deletion mutant, PCR was used to test the Em<sup>6</sup> clones. As expected about 50% of the clones were deletion derivatives (Sfi1cl6Δ1560).

**Characterization of Sfi1cl6Δ1560.** To test if we inactivated an essential gene of the phage we induced the prophage of both Sficl6 and Sfi1cl6Δ1560 with 2 μg/ml of Mytomicin C. Sficl6 released 10<sup>6</sup> pfu/ml, Sfi1cl6D1560 released no detectable infectious particles. Subsequently we checked whether the superinfection exclusion phenotype was preserved in the derivative lysogen.

<table>
<thead>
<tr>
<th>Phage</th>
<th>Sfi1 (pfu/ml)</th>
<th>Sfi1cl6 (pfu/ml)</th>
<th>Sfi1cl6Δ1560 (pfu/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sfi18</td>
<td>7.4 X 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>&lt; 10&lt;sup&gt;2&lt;/sup&gt;</td>
<td>&lt; 10&lt;sup&gt;9&lt;/sup&gt;</td>
</tr>
<tr>
<td>S96</td>
<td>9.8 X 10&lt;sup&gt;6&lt;/sup&gt;</td>
<td>&lt; 10&lt;sup&gt;2&lt;/sup&gt;</td>
<td>&lt; 10&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sfi21</td>
<td>8.3 X 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>&lt; 10&lt;sup&gt;2&lt;/sup&gt;</td>
<td>&lt; 10&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sfi19</td>
<td>1.6 X 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>8.8 X 10&lt;sup&gt;5&lt;/sup&gt;</td>
<td>1.7 X 10&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td>ST25</td>
<td>1.1 X 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>1.6 X 10&lt;sup&gt;7&lt;/sup&gt;</td>
<td>2.2 X 10&lt;sup&gt;7&lt;/sup&gt;</td>
</tr>
<tr>
<td>ST17</td>
<td>6.5 X 10&lt;sup&gt;8&lt;/sup&gt;</td>
<td>1.2 X 10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>1.3 X 10&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Table 2

In fact, plaque assays demonstrated an identical phage exclusion phenotype in both with respect to suppression of phage infectivity expressed as efficiency of plaquing as well as in the range of different phages which were inhibited (Table 2). Growth curves of the lysogen and its deletion derivative were identical (Fig. 6).

Figure 6. Growth curve of *S. thermophilus* Sfi1cl6Δ1560 (2) compared to Sfi1cl6 (1).
Discussion

Two independent strategies led to increased phage resistance of *S. thermophilus* starter strains. The first strategy relied on random insertion of an integrative plasmid into the chromosome of *S. thermophilus*. Thereby, we obtained highly phage resistant (e.o.p. < 10^6) mutants.

The broad range of anti-phage activity of the Sfi1-R7 mutation suggests that the 15 phages, which were tested, have similar requirements for host factors; i.e. orf 90 and orf 394 gps. At the current level of analysis we can not exclude effects of the IS element and/or the adjacent deletion on the transcription of adjacent bacterial genes. Phage structural proteins interact with bacterial structures at least for phage adsorption and DNA injection. The apparent lack of intracellular phage DNA in the Sfi1-R7 and Sfi1-R71 mutants may be an indication of interference with phage DNA injection. In phages from Gram-positive bacteria initial phage adsorption and phage DNA injection seem to follow a two step process (13;20;24;26;27). First, phages adsorb reversibly to carbohydrate components of the cell wall, then appears to follow an irreversible interaction with the plasma membrane and ejection of DNA. For bacteriophage c2 it has been shown that a membrane protein, PIP, was responsible for irreversible phage adsorption to the host (18). Phage c2 was unable to infect cells with defective PIP proteins. Whether PIP is also responsible for phage DNA injection or other structures are needed remains an open question. For example, bacteriophage λ requires a host protein (Pel) in addition to the LamB receptor for an effective DNA injection. Pel is apparently not involved in phage adsorption since λ binds tightly to *E. coli pel* strains (21). It is thus possible that in the case of the Sfi1-R7 and Sfi1-R71 mutants, proteins involved in phage irreversible adsorption and/or DNA injection have been modified. In fact, the sequence information available for Sfi1-R7 indicates that the insertion of pG*host9ISS1* may have blocked the expression of a putative integral membrane protein.

The phage resistance phenotype of Sfi1-R24 resembles that of an abortive infection mechanism. The phage DNA can enter the cell and phage DNA replication takes place, but it is delayed and diminished. We suspect therefore that the insertion event affects the expression of a host factor involved in a phage multiplication step preceding DNA replication. Blockage of this step apparently affects DNA replication
negatively. In the case of Sfi1-R24, the oxidoreductase is not the host factor whose inactivation causes the delay in DNA replication since its interruption by the IS element after plasmid excision fails to provide phage inhibition. Apparently, the transcription of adjacent genes is disturbed by the integrated plasmid. A gene encoding a subunit of a restriction endonuclease was the only bioinformatic hit in the targeted region. However, restriction and modification mechanisms are excluded since no traces of phage DNA degradation are detected in the Southern blot (Fig. 2). Direct interference with phage DNA replication is unlikely since the inhibited phages have apparently different DNA replication modules (6) and are thus likely to depend on different host replication functions.

The Sfi21 prophage could be inactivated by targeted food-grade phage gene inactivation. No release of infectious particles by the Sfi1cl6\1560 lysogen could be detected after prophage induction. Importantly, the inactivated prophage retained its protective effect against superinfection with a broad range of temperate and lytic S. thermophilus phages. Consequently, inactivated lysogens can now be considered as valuable starters. Their protective power exceeds that obtained with starter strains containing phage-inhibitory plasmids. For example, the superinfection immunity gene of Sfi21 (orf 203) was cloned in a high copy number plasmid and tested for protection against phages (8). It was quantitatively and qualitatively less complete than that mediated by the prophage. In addition, the plasmid was not any longer phage inhibitory when integrated as a single copy into the bacterial chromosome. Therefore, the use of orf 203 has to rely on its presence on a high copy number plasmid, with the problem of high metabolic costs and plasmid instability. In contrast, the integrated prophage seems to be of low metabolic cost to the cell since lysogenic and non-lysogenic S. thermophilus starters have identical growth properties (unpublished results). Inactivated prophages can be used to increase the phage resistance of other strains than Sfi1 since a number of valuable industrial S. thermophilus starters can be lysogenized with phage Sfi21 (9).

Reference List


3 Summary and conclusions

*S. thermophilus* phage diversity. To address the question of phage evolution and diversity *S. thermophilus* phages were initially compared by Southern blot and dot blot (3,4). All *S. thermophilus* phages, both temperate and lytic, were highly related and belonged to a single DNA-hybridization group. However, regions of high sequence similarity were interspersed with unrelated DNA segments. An extremely conserved 2.2 kb DNA region (>99% identity), probably involved in phage DNA replication, was shared by phages differing in many characteristics (4). This was interpreted as the result of recent genetic exchange within this group of phages. Evidences of genetic exchange between related phages were previously observed for the lambdoid phages, too (7). A sequencing project was then started to determine the complete DNA sequence of several *S. thermophilus* phages covering differences in life styles (lytic and temperate), DNA packaging mechanism (cos- and pac-site phages) and structural protein pattern (2 major structural proteins vs. 3), host range, serotype (1 and 2), susceptibility to prophage control (homo- or hetero-immune to phage Sfi21) and inhibition by phage Sfi21 encoded replication origin.

The complete sequences of five *S. thermophilus* phages are now available (Sfi11, Sfi19, Sfi21 (11), DT1 (38) and O1205 (36)). In addition, the right half of the genome of phage Sfi18 was sequenced. Pairwise and multiple alignments of these sequences supports the theory of modular evolution of bacteriophages (1): genes are clustered into functional units (modules), all *S. thermophilus* phages have a similar genomic organization (DNA-packaging followed by head morphogenesis, tail morphogenesis, host lysis, lysogeny module, DNA replication and finally a region possibly involved in phage development control). When phage DNA sequences were aligned, sharp transitions between regions of high sequence similarity and regions of low or no relatedness were observed. Individual *S. thermophilus* phages appear to result from recombination of exchangeable genetic elements drawn from a common gene pool. This gene pool is apparently restricted, since no evidence of recent horizontal gene transfer of host genes or genes from phages infecting other bacterial species have been found (defined by >70% nucleotide sequence similarity). An exception was provided by the high similarities (>80%) detected between the phage Sfi21 replication origin (12) and the noncoding region surrounding the cos-site of Sfi19 with the cryptic *S. thermophilus* plasmid pST1.

The units of genetic exchange were either entire functional modules (several adjacent structural genes), part of a module (individual genes in the lysogeny module) or even parts of a gene corresponding to likely protein domains (host range
determinant of the likely tail fibers, DNA binding domain of regulatory proteins). For each unit of genetic exchange, multiple alleles exist. In the lambdoid phages the frequency of the occurrence of the different alleles indicated that the total number in the population might be small (20). This seems also to be the case among the S.thermophilus phage population. For example, two different structural gene clusters were found; one corresponds to the cos-phages and one to the pac-phages. No evidence of a third allele could be found, all phages analyzed by Southern blot could be classified into one of these two groups (23;24). Also only two different DNA replication modules have been detected (39). The diversity within each allele was limited to point mutations (resulting in 15% sequence diversity) and small indels.

The genes involved in DNA replication or phage morphogenesis and DNA packaging are apparently exchanged as multigenic sequences. These proteins build a network of interactions with each other and the exchange of a single protein between unrelated alleles will probably result in nonproductive protein-protein interactions. The situation is different for the likely tail fiber genes. This region is clearly a recombination hot spot. Horizontal gene transfer events may be facilitated by the postulated elongated conformation of these proteins. In fibrous proteins individual protein segments make less contacts with other parts than in globular proteins allowing for more genetic flexibility.

The amino-end of the likely tail fibers are well conserved within the pac- or the cos-phages (>90%), whereas the carboxy-terminal part is the most diversified. This is similar to the situation in the tail fiber genes from \( \lambda \) and T4. In these two phages the highly conserved amino-end is thought to bind to the body of the virions, the more variable carboxy-end determines the host range (9). Transfer of segments of tail fiber genes seems to occur between cos- and pac-phages. This is reminiscent of what has been described for coliphages (18), where likely horizontal gene transfer of tail fiber genes occurs between phages having otherwise unrelated structural proteins. It has been postulated that the exchange of host range determinants is a mechanism used by phages to extend their host range.

A second recombination hot spot is the lysogeny module (30). In the comparison of three temperate S. thermophilus phages distinct multiple sharp transitions zones from highly similar to moderately similar or different DNA sequences were observed, suggesting multiple exchange events. The unit of genetic exchange could be as small as a single gene or even a segment of a gene. This remarkably recombination freedom could be rationalized by the characteristics of the encoded functions.
Proteins encoding integrative functions, superinfection immunity and transcription control do not depend on protein-protein interactions as structural virion proteins. Therefore, these functional units may be exchanged freely. Such exchanges within the lysogeny module have also been observed for lactococcal phages (25), and P2-like phages from *Haemophilus influenzae* (34). DNA/DNA-hybridization experiments showed that temperate and lytic phages of *S. thermophilus* are related (3). This has also been observed for *L. lactis* (5;13;31) and *Lactobacillus* (28;32;37) phages. Comparison of the entire genome of the temperate cos-site phage Sfi21 with the lytic cos-site phage Sfi19 confirmed their relationship. Both phages have a similar genomic organization and share extensive DNA sequence identity (>85%) along their entire genome except for the lysogeny region. The corresponding region in phage Sfi19 is much shorter (2.6 kb vs. 5.7 kb) and lacks sequence relatedness, both at the DNA and amino acid level.

Comparison of the Sfi19 lysogeny replacement module revealed a sequence similarity with the lysogeny modules of further temperate *S. thermophilus* phages (O1205, TP-J34). It is therefore likely that virulent *S. thermophilus* phages are the result of deletion events and gene rearrangements in temperate phages. In fact, a lytic derivative of the temperate phage Sfi21 could be readily isolated in the laboratory (6). A very similar virulent phage was also isolated from the field. All these observations raise the possibility that temperate and lytic phages of *S. thermophilus* are related to each other by simple gene insertion or deletion events. Sfi19-like lysogeny replacement modules were also found in Sfi11 and Sfi18. Their sequences were highly conserved (>95%). Many indels and frame shifts were detected, suggesting that several of the genes in that region are nonessential. Interestingly, all sequenced lytic phages conserved a cro-like gene.

Most of the sequenced *S. thermophilus* phages have related host lysis modules consisting of two holins followed by a lysin. These proteins are well conserved with aa similarities ranging from 78 to 99%. However, in phage DT1 (38) the first holin gene is missing. It has been shown that overexpression of only one of the holins is needed to have a lethal effect on *E. coli*. Further evidences that each gene of the host lysis module can be exchanged independently from the others have been reported by Sheehan et al. (33).

All the sequenced *S. thermophilus* phages have a very similar right genome half. The region directly downstream the lysogeny or lysogeny replacement module is involved in DNA replication. This is the most conserved part of the *S. thermophilus*
phage genomes (4;12). The region adjacent to the cos-site may have regulatory functions. Many deletions, insertions and gene replacements punctuated this region.

A detailed comparison of the DNA sequences from phages Sfi19 and Sfi18 indicates that point mutations may also play an important role in phage adaptation. Very few base pair changes seem to be under the escape of phage Sfi19 from the inhibition mediated by the Sfi21 encoded origin of replication.

**Relationship of S. thermophilus phages to Siphoviridae from the Gram-positive bacteria.** Classification of bacteriophages has always been a controversial issue. Since phages seem to exchange genetic material so frequently, phylogenetic relationships are expected to be blurred. The taxonomical criteria used currently to classify phages can not easily be used to establish phylogenetic relationships. We are convinced that any meaningful classification of phages has ultimately to be sequence based. Because of the frequent lateral gene transfers between phages, their evolutionary relationship has to be examined at the level of individual modules (1). Recently, data are accumulating indicating that lateral gene transfer events play also a prominent role in the evolution of bacterial genomes (14). For example, since *E. coli* diverged from the *Salmonella* lineage at about 100 million years ago, 18% new genes have been introduced into the *E. coli* genome (22).

To test whether it is possible to trace the evolutionary history of phages, we examined the genetic relationship between *S. thermophilus* phages and other phages focusing on modules that are under different evolutionary constraints, the structural gene cluster and the lysogeny module.

By comparing the structural genes of the pac-phage Sfi11 to the databases, we found multiple links to B1-Siphoviridae (small isometric head and long noncontractile tail) from Gram-positive bacteria. The phage sequence comparison revealed a hierarchy of relatedness, which correlated approximately with the evolutionary distance between their bacterial hosts to *S. thermophilus*: the more distant the hosts, the less similar were the phage structural genes. The observation of such a graded relatedness in phages is important, since it is the hallmark of a system undergoing evolutionary changes. Our data also indicate lack of horizontal gene transfer events between phages infecting distinct bacterial species. The highest similarity between a *S. thermophilus* phage and a phage infecting a different host species was found comparing the structural gene modules of Sfi21 and that of the lactococcal phage BK5-T. The comparison showed detectable DNA sequence similarity (50-60 %) along most of the structural gene module (11). This high identity may indicate that lateral
gene transfers between these phages may have occurred. However, DNA/DNA homology studies have shown that phages from these bacteria are genetically distinct (21) and probably, no more in genetic contact.

Alignment of the genetic maps of B1-Siphoviridae from Gram-positive bacteria revealed a striking conservation of gene order. This was true for both pac-site and cos-site B1-Siphoviridae from Gram-positive bacteria and extended even to temperate B1-Siphoviridae from Gram-negative bacteria like coliphage λ and an archaeavirus (11). On the basis of these observations we hypothesize that the morphogenesis module of B1-Siphoviridae shares a common ancestry. The conservation was not restricted to the gene order, but was seen in protein characteristics such as size and isoelectric point (10). A similar gene order may, at least in theory, favor the productive recombination between phages (8). Whether this fact is a selective force to maintain gene order is questionable for phages that are not any longer exchanging genetic material. This particular gene order could have been conserved because a correlation between gene arrangement and order of action of the gene products might be of selective advantage (8). The conservation of the gene order is even more striking when considering the fact that no sequence similarity can be detected between the structural proteins of λ and those of the B1-Siphoviridae from the Gram-positive bacteria. It is envisageable that these structural gene clusters separated in such a distant past that the protein sequences diverged beyond recognition.

At the moment the amount of available sequences is a limiting factor for the determination of sequence relationships between evolutionarily distant phages. Only about 35 complete sequences of phages are available and most of them come from low-GC content Gram-positive bacteria (n=19), followed by γ-proteobacteria (n=9) and high-GC content Gram-positive bacteria (n=4). The sequencing of phages from a broader phylogenetic range of hosts should give us the possibility to find more links between phages and thus, will provide the data for a sequence based theory of phage evolution (19). Recently, sequencing of the Streptomyces phage C31 showed that the head assembly proteins were conserved in phages coming from distantly related hosts; Streptomyces phage C31, coliphage HK97, staphylococcal phage PVL, two Rhodobacter capsulatus prophages and two Mycobacterium tuberculosis prophages (35). Similar links were demonstrated for head assembly proteins from S. thermophilus Sfi21 (11).
Alignment of the lysogeny modules of all currently sequenced temperate *Siphoviridae* from the low GC-content Gram-positive demonstrated related lysogeny modules. These modules showed not only a similar overall gene organization, but were also linked by many sequence similarities. The gene organization of the lysogeny module of this group of phages distinguishes it from the *Siphoviridae* infecting Gram-negative bacteria (e.g.: coliphages λ, 933W and P22) or high GC-content Gram-positive bacteria (mycobacteriophage L5). In contrast, similarly organized lysogeny modules were found in the temperate P2-like *Myoviridae* (29). Such a relationship is not in accordance with the current phage taxonomy. Other phage genome comparisons have revealed similar inconsistencies. The Podovirus P22 and the Siphovirus λ have a comparable genome map and still can form viable hybrids in the laboratory (2;17), but were classified into different phage families on the basis of morphological criteria. The dilemma of current phage taxonomy is thus too evident. In fact, a non-sequence-based taxonomy may miss many evolutionary links between phages.

At the moment, two genera of temperate *Siphoviridae* have been established (27): the λ-like and L5-like genera. Interestingly, all the currently sequenced temperate phages from the low-GC Gram-positive bacteria showed an identical overall genome organization (11;25) that differentiates them from these two established genera. Therefore, we suggest a new genus: the Sfi21-like phage genus. Because of their relatedness, we would also include lytic *S. thermophilus* phages in this genus, even if the difference in life style, according to the currently proposed taxonomy, would separate them in a different genus.

**Phage resistance.** Sequencing of several *S. thermophilus* phage genomes combined with large surveys of phage genomes by DNA/DNA hybridization provided insight into the genomic diversity of this phage group. There are regions of extreme diversity interspersed with highly conserved genome segments (4). The later regions are important in view to obtain strains that are resistant to a broad range of phages. The phage resistance encoded by the Sfi21 origin of replication (16) is a good example of an anti-phage mechanism providing protection against a large number of phages. In fact, the Sfi21-like replication module is widespread and highly conserved among phages from our collection (3;4). It has been shown that phages, e.g. Sfi19, having a Sfi21-like replication module could still escape the inhibition. Point mutations probably altered the specificity of some replication protein for the ori. Substitution of the Sfi21 replication origin with that of Sfi19 was shown to extend the
range of inhibitory activity of the origin (16). However, this strategy has one important limitation: to be efficient the replication origin has to be cloned in a high copy number plasmid. At the moment, food-grade vectors are not yet available for *S. thermophilus*. Attempts to obtain a food-grade vector based on selection exerted by key metabolic functions have encountered difficulties (15). Therefore, we decided to test two methods, which do not rely in the maintenance of a plasmid in the phage-resistant strains.

First, we targeted bacterial genes for inactivation by hypothesizing the presence of bacterial genes essential to the phage replication cycle. Should these genes be dispensable for bacterial growth in milk, their inactivation would lead to stable phage resistant cells. We used a temperature-sensitive integrative plasmid (pG’host9ISS7)(26) to randomly inactivate bacterial genes. After plasmid insertion we challenged the culture with phages to select for phage-resistant mutants. Using this approach we obtained three distinct Sfi1 phage-resistant insertion mutants. The growth of the starter cells was unaffected by the mutation. The phage protection level was very high (e.o.p. < 10^6) and each insertion mutant was protected against a wide range of phages, suggesting that the phages tested have similar cellular requirements. In two cases, phage development was likely blocked at the DNA-injection level. The third mutant showed a delayed and lowered phage-DNA replication. Excision of the integrative plasmid was then induced to obtain food-grade strains in which only one copy of the ISS1 element was left. One of the three mutants retained its phage resistance phenotype after the excision.

The second approach relied on the protection provided by the immunity functions of the Sfi21 prophage. The drawback of lysogenic starters is that they continuously release infectious phages into the factory environment. Because of this problem, the use of lysogens is avoided in the dairy industry. We decided to create a targeted deletion in an essential tail gene of the prophage to obtain a lysogen that would retain superinfection immunity, but has lost its capacity to produce infectious virions. We obtained the desired deletion derivative of the lysogen, which proved to be unable to release infectious phage particles (< 100 pfu/ml) but maintained intact superinfection control.

We showed that both methods provide powerful phage resistant and food-grade starters for industrial milk fermentation. Both approaches can be applied to a number of industrial starter strains. The field is now ready for the development of phage resistant *S. thermophilus* starters at a pilot plant level. Should they meet the technological requirements, a large-scale application of these starters could be envisioned in the dairy industry.
Reference List

15. Foley, S. (UnPub)
24. Lucchini, S. (UnPub)
35. Smith, M.C., R.N. Burns, S.E. Wilson, and M.A. Gregory. 1999. The complete genome sequence of the Streptomyces temperate phage straight phiC31: evolutionary relationships to other viruses. Nucleic Acids Res. 27:2145-2155.
39. van Sinderen, D. Personal Communication
Curriculum vitae


1976-1981  Primary education in Melide, TI.
1981-1985  Secondary education in Barbengo, TI.
1985-1990  Lyceum in Lugano, TI.
1990-1996  Study of Biochemistry (Abteilung XAe) at the Swiss Federal Institute of Technology in Zürich, ZH.
Publication list


