Doctoral Thesis

Ecology and evolution of Streptococcus thermophilus bacteriophages in industrial milk fermentations

Author(s):
Desiere, Frank

Publication Date:
1999

Permanent Link:
https://doi.org/10.3929/ethz-a-003822899

Rights / License:
In Copyright - Non-Commercial Use Permitted
Ecology and Evolution of *Streptococcus thermophilus* bacteriophages in industrial milk fermentations

Thesis
Submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH (ETHZ)
For the degree of
DOCTOR OF NATURAL SCIENCES

By

FRANK DESIERE
Dipl. Biol. (Biotechnology) University of Stuttgart, Germany
Born on July 31<sup>st</sup>, 1970
in Jugenheim a.d.B., Germany

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

Lausanne 1999
Thanks and Acknowledgements

How time flies! I still remember when I started my PhD and was beginning with the collection of bacteriophages at the EFAM in Liebefeld for 2 weeks and then 4 weeks at the Nestlé Regional Laboratory in Cergy, Paris. After another stop at the ETH in Zurich I had also the opportunity to compare our bacteriophages to Australian phages at the University of Melbourne to the end of my project. However, I am fortunate to have this opportunity to work on this project and in such a pleasant environment at the Nestlé Research Center in Lausanne. My first and warmest acknowledgement belongs to Dr. Harald Brüssow, who provided this opportunity, helped me through every stage for the completion of my thesis and I thank him also for the numerous scientific discussions. I very much appreciated the time he took for teaching me on how to become a real researcher and his "engagement" for the phage project.

I am also grateful to Prof. Dr. Michael Teuber, my thesis promoter, who "adopted" me for this project. I would like to extend my thanks to everybody in the department who helped to facilitate the project. I would like to thank especially Sacha Lucchini for the shared project, Anne Bruttin, Josette Sidoti and Sophie Foley for their pleasant way and all the students at the NRC. Furthermore I would like to thank Jaques-Edouard Germond for useful discussions between the door, Spiros Konteles for philosophising over Greek philosophy and life in general.
I would also like to express my thanks to our superiors at the Nestlé Research Center especially Dr. Andrea Pfeifer and Dr. Jean-Richard Neeser for the support of this project [RE-003101.09, 003102.08].
This study was financially supported by grants from the Swiss National Science Foundation in the framework of the SPP Biotechnology module (Food Biotechnology), grant 5002-44545 and AU-054586.
Last but really not least I thank my parents, my sister Bettina and my brothers Dirk, Robert, Axel and Eric for their support and help during the years of my studies.
Gratefully,
Frank Desiere
fdesiere@yahoo.com, October 1999
In considering the Origin of Species, it is quite conceivable that a naturalist, reflecting on the mutual affinities of organic beings, on their embryological relations, their geographical distribution, geological succession, and other such facts, might come to the conclusion that each species had not been independently created, but had descended, like varieties, from other species. Nevertheless, such a conclusion, even if well founded, would be unsatisfactory, until it could be shown how the innumerable species inhabiting this world have been modified so as to acquire that perfection of structure and co-adaptation which most justly excites our admiration.

Charles Darwin (1809–82), English naturalist. On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life. H. M. S. Beagle’s Voyage round the world. London: John Murray, Albemarle Street, 1859

The next morning [after the experiment], on opening the incubator, I experienced one of those rare moments of intense emotion which reward the research worker for all his pains: at the first glance I saw the broth culture, which the night before had been very turbid, was perfectly clear: all the bacteria had vanished, they had dissolved away like sugar in water. As for the agar spread, it was devoid of all growth and what caused my emotion was that in a flash I had understood: what causes my clear spots was in fact an invisible microbe, a filterable virus, but a virus parasitic on bacteria.

Felix d’Hérelle, French Epidemiologist.
On his discovery of bacteriophages in 1916.
# TABLE OF CONTENTS

1 Abstract 1

2 Kurzfassung 2

3 Introduction 3

3.1 Bacteriophages: general properties 3

3.2 Bacteriophages of Lactic Acid Bacteria 5

3.2.1 Phages as genetic tools 7

3.2.2 Bacteriophage Resistance in Lactic Acid Bacteria 8

3.3 Streptococcus thermophilus bacteriophage Sfi21 9

3.4 Bacteriophage ecology 9

3.5 Bacteriophage evolution 10

3.6 Aims of the study 11

4 Molecular Ecology of Streptococcus thermophilus Bacteriophage Infections in a Cheese Factory. 13

5 Characterization of the Lysogeny DNA Module from the Temperate Streptococcus thermophilus Bacteriophage Sfi21 21

6 Comparison of the Lysogeny Modules from the Temperate Streptococcus thermophilus Bacteriophages TP-J34 and Sfi21: Implications for the Modular Theory of Phage Evolution. 35

7 A Highly Conserved DNA Replication Module from Streptococcus thermophilus Phages is Similar in Sequence and Topology to a Module from Lactococcus lactis Phages 49

8 Evolution of Streptococcus thermophilus Bacteriophage Genomes by Modular Exchanges Followed by Point Mutations and Small Deletions and Insertions 61

9 Comparative Sequence Analysis of the DNA Packaging and Morphogenesis Genes in the Temperate cos-site Streptococcus thermophilus Bacteriophage Sfi21 75

10 Packaging and morphogenesis genes from Lactobacillus johnsonii and E.coli phages reveal different modes of evolution of Siphoviridae 87
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 The Relationship of Streptococcus thermophilus Bacteriophage Sfi21</td>
<td>103</td>
</tr>
<tr>
<td>with Lactococcus lactis subsp. cremoris Bacteriophage BK5-T:</td>
<td></td>
</tr>
<tr>
<td>Evidence for Horizontal Gene Transfer and a Common Ancestor</td>
<td></td>
</tr>
<tr>
<td>12 Summary and Conclusions</td>
<td>113</td>
</tr>
<tr>
<td>12.1 Ecology of Streptococcus thermophilus phages</td>
<td>113</td>
</tr>
<tr>
<td>12.2 The genome of Streptococcus thermophilus bacteriophage Sfi21</td>
<td>114</td>
</tr>
<tr>
<td>12.2.1 Lysogeny module</td>
<td>115</td>
</tr>
<tr>
<td>12.2.2 The lysis cassette</td>
<td>117</td>
</tr>
<tr>
<td>12.2.3 The conserved replication module</td>
<td>118</td>
</tr>
<tr>
<td>12.2.4 Structural Gene Cluster</td>
<td>118</td>
</tr>
<tr>
<td>12.3 Relationship of Sfi21 to other bacteriophages</td>
<td>119</td>
</tr>
<tr>
<td>12.4 Bacteriophage evolution</td>
<td>119</td>
</tr>
<tr>
<td>12.5 Genome organization of tailed bacteriophages</td>
<td>122</td>
</tr>
<tr>
<td>13 References</td>
<td>124</td>
</tr>
<tr>
<td>14 Curriculum Vitae</td>
<td>131</td>
</tr>
<tr>
<td>15 Publication list</td>
<td>132</td>
</tr>
</tbody>
</table>
1 Abstract

Bacteriophage attack of Lactic Acid Bacteria has always been a major problem in the dairy industry causing slow fermentation or even complete starter failure with consequent loss of product which implies high financial losses. The development of phage-resistant starter cells is thus an industrial priority. Since only few plasmid-encoded resistant factors are known for *Streptococcus thermophilus*, an important thermophilic dairy starter, it was decided to investigate the genome of their phages for phage inhibitory DNA elements. Temperate phages encode factors that prevent superinfection by many phages. I therefore analysed the prototype *Streptococcus thermophilus* bacteriophage Sfi21 in molecular detail. I determined its complete genomic nucleotide sequence, identified the genes and analysed them with bioinformatic tools. Furthermore I studied the relationship of virulent and temperate *Streptococcus thermophilus* phages as well as their relationship with bacteriophages of an evolutionary wide range of bacterial hosts. Questions of horizontal gene transfer, co-evolution of phages with their hosts and the short-term evolution of phages in a longitudinal factory phage ecology study were addressed. The main point in the thesis are:

1.) A factory survey determined that the diversity of the resident bacteriophage population arose from environmental sources and not from genetic changes of the resident phages in the factory. Distinct phage types were observed and no point mutations in multiple isolates of the same phage type were found. After the introduction of a defined starter into the factory, new phage types were observed in the factory while the resident phage population was diminished but not eliminated. Raw milk was the source of these new phages.

2.) The temperate *Streptococcus thermophilus* phage Sfi21 has a genome of 40,739 bp in size with 15 nt long cohesive ends (cos) with a 3' overhang that reconstituted a cos-site with rotational symmetry. All but 4 orfs, that encoded lysogeny-related functions, were encoded on the same DNA strand. Functionally related genes were clustered resulting in a modular genome structure: DNA packaging, head, tail, tail-fiber morphogenesis, lysis, lysogeny, DNA replication and a regulatory module.

3.) Comparative genome analysis of several bacteriophages from low GC-content Gram-positive bacteria demonstrated that the diversification of bacteriophage genomes during evolution was achieved by two distinct mechanisms: DNA recombination events and accumulation of point mutations. The units of genetic exchange were either large, comprising entire functional modules, or small, consisting of one or two genes or even segments of a gene which possibly represent protein domains. Recombinational hot spots for insertions and deletions were identified.

4.) *S. thermophilus* phages showed sequence similarity with many other bacteriophages and *S. thermophilus* plasmids. The similarity with other bacteriophages showed a hierarchy of relatedness that correlated with the evolutionary distance that separated the bacterial hosts. The observation could imply that the dsDNA tailed phages share a common ancestry.
2 Kurzfassung


2) Der temperente S. thermophilus Phage Sfi21 hat ein Genom von 40739 Basenpaaren und enthält überstehende Enden (cos), die aus einem 3', 15 Nucleotiden langen DNA-Einzelstrang mit Rotationssymmetry bestand. 4 Gene, die für die Lysogenie des Phagen mitverantwortlich sind, waren in entgegengesetzter Richtung kodiert. Funktionell verwandte Gene lagen gruppiert vor, was auf die folgende modulare Genomstruktur schliessen liess: DNA-Verpackung, Kopf- und Schwanz-Morphogenese, Lysis, Lysogenie, DNA-Replikation und Regulation.

3 Introduction

3.1 Bacteriophages: general properties

The bacteriophages were discovered independently by the microbiologists Frederick William Twort (1915) and Félix d'Hérelle (1917) (d'Herelle, 1917; Twort, 1915). They were the last of the three major classes of viruses to be discovered. Plant viruses were discovered in 1892 by Ivanowski and animal viruses in 1902 by Loeffler and Froesch. Twort and d'Hérelle were the first to recognize viruses which infect only bacteria, which d'Hérelle called bacteriophages (eaters of bacteria). In the 1930s and subsequent decades, pioneering virologists such as Luria, Delbrück and many others utilized these viruses as model systems to investigate many aspects of virology, e.g. including virus structure, morphogenesis, genetics and replication. Phage therapy of bacterial diseases received wide attention in this age without antibiotics (Eaton and Bayne-Jones, 1934). Nearly 4,500 different double-stranded DNA (dsDNA) phages capable of infecting a large diversity of bacterial hosts have been described until today (Ackermann and DuBow, 1987). Bacteriophages were instrumental in the development of our understanding of all types of viruses, including those of man, which were propagated in cell culture only much later. The phages have been much used in the study of bacterial genetics and cellular control mechanisms largely because the bacterial hosts are so easily grown and infected with phage in the laboratory. They are still a paradigm for many areas of biology and bacteriophage lambda is the best investigated biological system (Campbell and Botstein, 1983).

Bacteriophages, are very common in many natural environments, like in soil (Ogunseitan et al., 1992), aquatic systems (Bergh et al., 1989; Fuhrman, 1999) and sewage. They influence the bacterial abundance and thus indirectly biogeochemical processes. Phages mediate genetic transfer between bacteria and contribute thus to the evolution of bacteria.

Bacteriophages, or phages, have a head composed of protein, an inner core of nucleic acid, either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), and a hollow protein tail (Ackermann, 1999). In most cases, a particular phage infects only a single bacterial species or particular strains within that species. For example, coliphages are viruses that infect only the bacterium Escherichia coli. A virus infects a bacterial cell by first attaching to the bacterial cell wall by its tail. In coliphages the tail is a complex protein structure consisting of a hollow (non)-contractile sheath, with a plate at the base that contains long protein fibers. The tail fibers fix the base plate to the specific receptor site on the bacterial cell wall, and the tail sheath contracts like a syringe, forcing the DNA that is inside the virus through the cell wall and cell membrane. The entire virus protein coat remains outside the bacterium. The injected nucleic acid makes use of the bacterium's chemical energy and biosynthetic machinery to produce viral proteins that an uninfected bacterium does not make, as well as more phage...
nucleic acid. The viral proteins and nucleic acid molecules within the bacterial host assemble spontaneously into up to a hundred new phage particles. Eventually the bacterium lyses, and the particles are released. Lysis can be readily observed in bacteria growing on a solid medium, where groups of lysed cells appear as clear areas, or plaques.

It became increasingly clear already during the 1930s and 1940s that some phages, called temperate phages, have a more complex relationship with the host than simple lysis. There were in fact two quite fundamentally different groups of bacteriophages. Lytic phages always have to infect from outside, reprogram the host cell and release a burst of phage through breaking open, or lysing, the cell after a relatively fixed interval. Temperate phages, on the other hand, have another option. They can actually integrate their DNA into the host DNA, leading to virtually permanent association as a prophage with a specific bacterium and all its progeny. The prophage directs the synthesis of a repressor, which blocks the reading of the rest of its own genes and also those of any closely-related phages -- a major advantage for the bacterial cell. Many prophages further protect the cell against various unrelated, lytic phages. Occasionally, a prophage escapes from regulation by the repressor, excises its DNA out of the genome by a site-specific recombination system, produce progeny phage and lyse open the cell. A lysogenic culture, i.e., a bacterial culture infected with temperate phages, can be treated with radiation or mutagens, either of which induces the cells to begin producing viruses and lyse. Sometimes the cutting-out process makes mistakes and a few bacterial genes get carried along with the phage DNA to its new host; this transduction process, plays a significant role in bacterial genetic exchange. Such lysogenic phages could induce resistance and can potentially lead to transfer of genes involved in bacterial pathogenicity (Betley and Mekalanos, 1985). Functional as well as defective prophages may be responsible for genetic variation and may contribute to the variability of bacterial populations in their natural environment (Ramirez et al., 1999).

Viruses are classified according to their genome and their mode of replication. According to the Baltimore classification, viruses can be classified in seven arbitrary groups:

- I: Double-stranded DNA (Adenoviruses; Herpesviruses; Poxviruses, etc)
- II: Single-stranded (+) sense DNA (Parvoviruses)
- III: Double-stranded RNA (Reoviruses; Birnaviruses)
- IV: Single-stranded (+) sense RNA (Picornaviruses; Togaviruses, etc)
- V: Single-stranded (-) sense RNA (Orthomyxoviruses; Rhabdoviruses, etc)
- VI: Single-stranded (+) sense RNA with DNA intermediate (Retroviruses)
- VII: Double-stranded DNA with RNA intermediate (Hepadnaviruses)

Furthermore, viruses are classified in a hierarchical virus classification system:

Order - Family - Subfamily - Genus - Species - Strain/Type

Within this classification, features such as morphology (size, shape, enveloped, unenveloped), physicochemical properties (molecular mass, buoyant density, pH, thermal, ionic stability), genome (RNA, DNA, segmented sequence, restriction map, modifications, etc.), macromolecules (protein composition and function), antigenic properties, biological properties (host range, transmission tropism, etc.) are all considered (Cann A.J., 1997).
With an estimated total number of \( \sim 10^{30} \) phages in the world (Bergh et al., 1989; Whitman et al., 1998), tailed phages represent the largest of the viral groups with more than 4500 isolated phage species to date. The apparent relatedness of tailed phages has led to their new classification as the order of "Caudovirales" (the name stems from the Latin cauda for "tail") (Ackermann, 1999). The order Caudovirales is subdivided into 3 families according to their structure: Myoviridae with contractile tails, Siphoviridae with long non-contractile tails and Podoviridae with short tails (Ackermann and DuBow, 1987).

The monophyletic origin of tailed phages has been proposed several times (Ackermann and DuBow, 1987) and the 11th international Congress of Virology 1993 in Glasgow introduced this new order because most of the tailed phages are of similar morphology and have a linear double-stranded DNA with a genome of about 50 kbp. However, some genomes are less than 20 kbp and phages from the T4-like group have a genome size of 160 kbp, and bacteriophage G is the largest with a genome of 650 kbp; bigger than the smallest bacterial genome (Donelli et al., 1975).

### 3.2 Bacteriophages of Lactic Acid Bacteria

Lactic Acid Bacteria (LAB) are widely used as bacterial starters in many food and feed fermentations and the production of foods is one of the oldest manifestations of biotechnology (Josephsen and Neve, 1998). Through their ability to ferment milk sugar (lactose) into lactic acid, the concomitant drop in pH inhibits growth of many spoilage and pathogenic bacteria and provides therefore a "natural" preservation of cheeses, yogurts, breads, wines, pickles and meats. LAB, which includes the genera Streptococcus, Lactococcus, Leuconostoc, Lactobacillus and Pediococcus play furthermore an important role in the development of flavor and texture but also other nutritional attributes of the products. LAB are found as part of the natural flora of milk but also in plant material and some are members of the normal flora of mucosal membranes (oral cavity, intestine and vagina) of mammals (Teuber, 1993). In these biotopes they are usually non-pathogenic. However, there are pathogenic species like hemolytic streptococci (e.g. Streptococcus pneumoniae), which are potent pathogens (Szmigielski et al., 1999).
The genome of LAB has typically a GC-content of less than 50% and they contain a chromosome within the size range of 1.8 to 3.4 Mbp (Davidson et al., 1996). LABs are Gram-positive, non-sporo-forming, acid tolerant, non-respiring rod or cocci and according to 16S and 23S rRNA sequence analysis they belong to the *Clostridium* branch of Gram-positive bacteria (low GC-content branch) (Schleifer and Ludwig, 1995). The species *Streptococcus thermophilus*, which is a member of the oral Streptococci (Hardie and Whiley, 1995), is most important in the production of yogurt, Italian- and Swiss-type cheese which are made at elevated temperature e.g. 42°C. They are usually associated with thermophilic lactobacilli (Teuber, 1993). *S. thermophilus* is naturally found in milk and decaying plant material. Due to their wide application in the food industry for the production of fermented food products, the physiology and genetics of LABs have been extensively studied (Davidson et al., 1996); (Fitzgerald, 1997; Kuipers et al., 1997; Vogel and Ehrmann, 1996).

Phage attack has always been a major threat to industrial milk fermentations in the dairy industry and they have remained the major cause of fermentation failure ever since they were first described in 1935 (Whitehead and Cox, 1935). Infection of the bacterial starter with bacteriophages results in fermentation failure and unacceptably low production of lactic acid and flavor compounds along with reduced proteolysis and in extreme cases may lead to loss of the product (Josephsen and Neve, 1998). Due to a constant risk of economic losses, control of phage is a major area of concern in handling lactic acid bacteria as starter cultures (Daly et al., 1996). Many LAB strains are lysogenic which poses a constant threat to the fermentation process releasing infectious particles in low numbers (Davidson et al., 1990).

All bacteriophages of *Streptococcus thermophilus* known today belong to the *Siphoviridae* family with isometric heads and long non-contractile tails and belong to group B as defined by Bradley (Brüssow et al., 1998; Brüssow, 1999). All virulent and temperate phages belong to one DNA homology group with a genome ranging in the size from 31 to 45 kbp. Only about 1% of the phage isolates are temperate phages. Extensive cross-hybridization between individual fragments of the prototype temperate phage Sfi21 and a large collection of virulent phages suggests their close relationship (Brüssow and Bruttin, 1995). Based on their protein profiles and their degree of homology they were classified into two or three subgroups (Benbadis et al., 1990; Brüssow et al., 1994a; Le Marrec et al., 1997; Neve et al., 1989).

Brüssow et al. have characterized 81 individual lytic phages covering 30 years of industrial fermentation mainly in France and in Italy. Many different restriction pattern were observed and the genome size was estimated to a mean of 35 kbp (Brüssow et al., 1994a). Two lytic groups could be distinguished on the basis of their host range, serology, protein analysis and DNA hybridization. Le Marrec et al., 1997, divided *Streptococcus thermophilus* phages into two groups on the basis of their packaging mechanism (cos-site/pac-site phages). The cos-containing phages possessed two major structural proteins in contrast to the pac-containing phages, which possessed three major structural proteins.
### Table 1: Examples of Lactic Acid Bacteriophages other than *S. thermophilus*.

<table>
<thead>
<tr>
<th>Host species</th>
<th>Morphology</th>
<th>Classific. in</th>
<th>Genome size [kbp]</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1t (van Sinderen <em>et al.</em>, 1996)</td>
<td></td>
<td></td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>sk1 (Chandy <em>et al.</em>, 1997)</td>
<td></td>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>c2 (Lubbers <em>et al.</em>, 1995)</td>
<td></td>
<td></td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>TP901-1 (Johnsen <em>et al.</em>, 1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BK5-T (Mahanivong <em>et al.</em>, 1996)</td>
<td></td>
<td></td>
<td>40</td>
<td></td>
</tr>
<tr>
<td><em>Lactobacillus delbrückii</em></td>
<td>Siphoviridae</td>
<td>4 unrelated homology classes</td>
<td>34-80</td>
<td>(Forsman, 1993; Sechaud <em>et al.</em>, 1988)</td>
</tr>
<tr>
<td>LL-H (Mikkonen <em>et al.</em>, 1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mv4 (Lahbib-Mansais <em>et al.</em>, 1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>phigte (Kodaira <em>et al.</em>, 1997)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adh (Raya <em>et al.</em>, 1989)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lactobacillus casei</em></td>
<td>a2 (Herrero <em>et al.</em>, 1994)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Oenococcus mesenteroides</em></td>
<td>Siphoviridae</td>
<td>6 unrelated classes, Lysogeny is widespread</td>
<td>24-70</td>
<td>(Poblet-lcart <em>et al.</em>, 1998)</td>
</tr>
<tr>
<td><em>Oenococcus oenos</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L10 (Sutherland <em>et al.</em>, 1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 3.2.1 Phages as genetic tools

Phages have also been studied to develop genetic tools which can be utilized for biotechnological applications in LABs. By exploiting their genetic and metabolic capacity, LAB can be used to generate a variety of products from different sugars (lactose, etc.). Applications include flavor and texturing compound production (De Vuyst and Degeest, 1999) and stress tolerance.

Novel expression systems can be constructed for LAB by using genetic material derived from their phages based on the regulatory systems of the temperate *Lactococcus lactis* phage r1t (Nauta *et al.*, 1996; Nauta *et al.*, 1997). Of the derivatives that exhibited temperature-sensitive phenotypes, one was shown to hold promise for both fundamental and industrial applications that require the controlled production of (heterologous) proteins in *L. lactis*. It has its industrial applications in accelerated cheese ripening through controlled expression of the lytic genes which induce lysis of the *Lactococcus lactis* starter strains for facilitated release of intracellular enzymes involved in flavour formation (de Ruyter *et al.*, 1997).

Another application is the construction of site-specific integration systems. Such systems allow to integrate foreign DNA into the bacterial chromosome into a defined place which is known not to interrupt any of the essential bacterial genes. These systems have been developed for *Lactococcus* (for review see Davidson *et al.*, 1996), for *Lactobacillus* (Alvarez *et al.*, 1996; Auvray *et al.*, 1997) and for *Streptococcus thermophilus* (Bruttin *et al.*, 1997). The site-specific recombination system of *S. thermophilus* bacteriophage Sfi21 has been identified (Bruttin and Brüssow, 1996) as well as in many other bacteriophages. Nonreplicative plasmids containing the bacteriophage A2-specific recombination cassette integrate into different lactobacilli but also into unrelated Gram-positive bacteria such as *Lactococcus lactis* and even into *Escherichia coli* (Alvarez *et al.*, 1998).

Furthermore, phages have been used for transduction of either chromosomal or plasmid DNA between strains (Teuber and Lembke, 1983) and for transfection
experiments to evaluate the potential of transformation of LAB (Mercenier, 1990; Teuber and Lembke, 1983).

3.2.2 Bacteriophage Resistance in Lactic Acid Bacteria

Bacteriophages have the ability to infect and destroy bacteria and pose a particularly serious threat to dairy fermentations that can result in high economic losses. The problem has led to attempts to limit the damage caused by bacteriophages and substantial progress has been made unraveling the genetics and molecular biology of LAB phages and their complex interactions with host strains. These developments have been reviewed by a number of excellent publications (Allison and Klaenhammer, 1998; Daly et al., 1996; Klaenhammer and Fitzgerald, 1994). Despite the progress in the field, bacteriophage attack continues to challenge industrial fermentations mainly due to the recombinogenic character of bacteriophages and their capacity to respond and overcome any anti-phage measures.

Starter culture management: Over time the dairy industry has implemented various methods to control phages. These methods include hygienic measures (e.g. separation of the starter culture room from the production area, heat treatment of the media, use of closed vats, and regular disinfecting of equipment), strain rotation strategies (from unknown mixed starters to rotation of defined starter strains) and the use of phage inhibitory media (PIM). Despite these efforts, phages have not been eliminated from milk fermentations.

Genetic strategies for improving the phage resistance have been fueled by the knowledge of natural phage systems. Lactococci for example, have developed a variety of methods to interfere with phages development and currently four distinct types of bacteriophage insensitivity have been described.

- **Absorption inhibition**: Carbohydrate moieties and receptor proteins on the cell surface are implicated in phage adsorption and cell-infection. Spontaneous bacteriophage insensitive mutants (BIMs) can result from any mechanisms that change the characteristics of the cell surface or the phage receptors. The cell wall and membrane receptors are however poorly characterized (Allison and Klaenhammer, 1998).
- **Inhibition of DNA injection**: Following adsorption, injection of the phage DNA through the cell wall and membrane are essential for successful infection and propagation of bacteriophages in a bacterial host. It has been proposed that specific host cell membrane proteins must be present in productive phage infections (Babu et al., 1995).
- **Restriction/Modification**: Restriction and Modification systems are present in bacteria and they act to protect the cell from foreign DNA such as phage DNA (Bickle and Kruger, 1993). First, a specific sequence on the foreign DNA must be recognized. The subsequent modification of the recognized DNA by a methylase leads to protection against the action of the restriction system and unmodified DNA is digested subsequently by a site-specific endonuclease.
- **Abortive infection (Abi)**: Abortive infection is a phage resistance mechanism that acts after phage adsorption and DNA injection and results in cell death. The Abi mechanisms that have been described until now, remain incompletely understood as to their mode of action. It has not been resolved, if these natural Abi system are designed to kill the host cell or halt phage development (Daly et al., 1996).
Historic techniques like selection of phage resistant starter cultures have been largely empirical based on trial and error. Such systems were prone to the rapid development of new phage varieties which circumvent the natural resistance. With the development of molecular tools in microbiology and biotechnology a more rational approach has been employed. Now, the challenge is to develop new defense systems that exert a much longer protection against phage attack (Allison and Klaenhammer, 1998). Novel phage resistance mechanisms rely on the increased understanding of the structure and organization of bacteriophage genomes. One example is the cloning of an phage origin of replication in *Streptococcus thermophilus* which outcompetes the true phage origin for essential replication proteins and blocks bacteriophage propagation (Foley *et al.*, 1998).

**3.3 *Streptococcus thermophilus* bacteriophage Sfi21**

As attack of *Streptococcus thermophilus* bacteriophages is a common nuisance in milk fermentations, several laboratories around the world systematically collected bacteriophages from fermentation failures. 80 bacteriophage isolates originating from different European countries have been collected over the last 30 years by Nestlé. Classification of this collection defined 4 lytic groups (Brüssow *et al.*, 1994a). Only one temperate isolate could be obtained (Sfi21). This starter spontaneously released infectious Sfi21 particles at low level. The temperate cos-site phage Sfi21 was selected as the prototype phage. Within our phage collection, extensive cross-hybridization has been observed between individual restriction fragments from Sfi21 and a large panel of virulent bacteriophages. Phage Sfi21 showed close relationship with lytic bacteriophages from all lytic groups especially with lytic group I indicating a close relationship between temperate and virulent phages (Brüssow, 1999).

A restriction endonuclease map was constructed (Brüssow and Bruttin, 1995) and the genome consisted of a linear DNA of approximately 40kb with cohesive ends. The temperate phage Sfi21 differs from the lytic phages only in the possession of a single contiguous DNA fragment which we identified as the putative lysogeny module. All the bacteriophage isolates investigated, revealed the same morphological properties with isometric head (65 nm) and non-contractile tails (250 nm). *Streptococcus thermophilus* bacteriophage Sfi21 belongs to the family of *Siphoviridae*. Significant genetic instability was observed during serial lytic passages of phage Sfi21 (Bruttin and Brüssow, 1996). These deletions were localized in three different regions, did not affect the lytic cycle of the phage and were therefore apparently not essential regions. One deletion, which was located in the lysogeny region and truncated the integrase gene, abolished the lysogenic capacity of the phage.

**3.4 Bacteriophage ecology**

Currently, bacteriophages against *S. thermophilus* are such a growing problem in the Italian cheese industry that even the possibility of replacing *S. thermophilus* by other
lactic acid bacteria in mozzarella starter blends is being studied. Therefore, an understanding of the mechanisms creating phage diversity is not only of substantial theoretical interest, but also of practical importance for the food industry. Dairy fermentations remain susceptible to phage infection for different reasons: pasteurized milk is not sterile and may contain phages; contaminating phage are dispersed in milk; and repeated use of defined cultures under non-aseptic processing conditions provides a constant host for phage proliferation (Klaenhammer and Fitzgerald, 1994).

3.5 Bacteriophage evolution

The origin of bacteriophages is a matter of speculation (Ackermann and DuBow, 1987; Reanney and Ackermann, 1982). Three mutually exclusive theories have been proposed:

- Viruses are descendants of the earliest forms of life,
- They are degenerate complex organisms,
- They were once cellular components, normal cellular nucleic acids that gained the ability to replicate autonomously. DNA viruses came from plasmids or transposable elements. They then evolved coat proteins and transmissibility.

Analysis of virus evolution by comparative DNA sequencing of selected virus genes has become common place in contemporary virus research. A vast amount of such data exists for viruses such as the human immunodeficiency virus type-1 or the influenza virus, to quote only two animal viruses of medical importance. These analyses have led to phylogenetic trees for specific virus genes which has allowed some conclusions to be drawn with respect to the evolution of the specific virus group. Virologists have realized that certain ideas on virus evolution established for viruses of animals do not apply to those of bacteria. During the seventies a great deal of evidence has been accumulated suggesting that lambdoid phages are related in ways not easily accounted for by the standard ideas of evolution along branching trees of linear descent. On the basis of these data, Botstein has formulated a modular theory of bacteriophage evolution (Botstein, 1980). The essentials of this theory were stated as follows:

> The product of evolution is not a given virus but a family of interchangeable genetic elements (modules) each of which carries out a particular biological function; exchange of a given module for another occurs by recombination among viruses belonging to the same interbreeding population and finally, these viruses and their hosts can differ widely in many characteristics.

Thus evolution acts primarily at the level of the modules and not at the level of the intact virus. This theory turned out to be extremely helpful for understanding the genome organization of bacteriophages and has been accepted as a standard hypothesis for bacteriophage evolution in general. This modular theory has, however, been tested mainly by heteroduplex mapping of lambdoid coliphages. No sequence-based theory of bacteriophage evolution has been formulated.

Also the taxonomical classification of bacteriophages is problematic and no clear definition for the term "phage species" has been proposed. Many taxonomical criteria do not reflect the "natural" phylogenetic relationship of phages. Comparative phage genomics has the potential to provide a natural classification system for phages.
3.6 Aims of the study

The fermentation processes in the dairy industry (e.g. cheese and yogurt production) depend crucially on the acidifying capacities of Lactic Acid Bacteria (LAB). Bacteriophages attacking these LAB have always been a major problem in the industry. Contamination by virulent phages may result in the lysis of bacterial starter strains, causing slow fermentation or even complete starter failure with consequent loss of product which implies high financial losses. Up until now dairy-phage research has concentrated on phages attacking mesophilic starters based on Lactococcus lactis and the thermophilic starter strain Lactobacillus delbrueckii ssp. bulgaricus. In contrast, little information is available concerning bacteriophages of Streptococcus thermophilus, a thermophilic bacterium used as a starter culture for Italian- and Swiss-type cheese and yogurt.

Knowledge about the genetic diversity of Streptococcus thermophilus phages in the factory and also in their natural environment milk is essential for any anti-phage measures and the understanding of the phage-host interactions.

The existence of large S. thermophilus bacteriophage collections in many countries with long dairy tradition, is certainly an asset of research in this field. Although phage collections for other dairy phages e.g. lactococcal phages exist, S. thermophilus phages have the advantage to belong to a single DNA homology group. This genetic homogeneity will thus facilitate the analysis of their genetic properties and their evolution.

With the close relationship between lactococcal phages and streptococcal phages and the availability of DNA sequences from phages of different genera we were in a position to address questions concerning possible horizontal gene transfer or co-evolution of phages with their hosts.

The aim of this research projects was to study the ecology and the evolution of Streptococcus thermophilus bacteriophages. It was decided to investigate the temperate Streptococcus thermophilus bacteriophage Sfi21 in molecular detail by determining its complete genomic nucleotide sequence. Genes were predicted and their possible gene functions were investigated using advanced bioinformatic tools. Biological experiments were conducted to test the predictions. Our practical interest was to identify phage genes or DNA elements that could provide phage inhibitory functions when cloned into plasmid vectors.
Seite Leer / Blank leaf
4 Molecular Ecology of *Streptococcus thermophilus*
Bacteriophage Infections in a Cheese Factory.


Molecular Ecology of *Streptococcus thermophilus* Bacteriophage Infections in a Cheese Factory

ANNE BRUTTIN, FRANK DESIERE, NICOLA d’AMICO, JEAN-PIERRE GUÉRIN, JOSETTE SIDOTI, BORIS HUNI, SACHA LUCCHINI, AND HARALD BRUSSOW*

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

Received 6 March 1997/Accepted 8 June 1997

A mozzarella cheese factory using an undefined, milk-derived *Streptococcus thermophilus* starter system was monitored longitudinally for 2 years to determine whether the diversity of the resident bacteriophage population arose from environmental sources or from genetic changes in the resident phage in the factory. The two hypotheses led to different predictions about the genetic diversity of the phages. With respect to host range, 12 distinct phage types were observed. With two exceptions, phages belonging to different lytic groups showed clearly distinct restriction patterns and multiple isolates of phages showing the same host range exhibited identical or highly related restriction patterns. Sequencing studies in a conserved region of the phage genome revealed no point mutations in multiple isolates of the same phage type, while up to 12% nucleotide sequence diversity was observed between the different phage types. This diversity is as large as that between the most different sequences from phages in our collection. These observations make unlikely a model that postulates a single phage invasion event and diversification of the phage during its residence in the factory. In the second stage of our factory study, a defined starter system was introduced that could not propagate the resident factory phage population. Within a week, three new phage types were observed in the factory while the resident phage population was decreased but not eliminated. Raw milk was the most likely source of these new phages, as phages with identical host ranges and restriction patterns were isolated from raw milk delivered to the factory during the intervention trial. Apparently, all of the genetic diversity observed in the *S. thermophilus* phages isolated during our survey was already created in their natural environment. A better understanding of the raw-milk ecology of *S. thermophilus* phages is thus essential for successful practical phage control.

Phage attack has always been a major problem in industrial fermentation, especially in the dairy industry (11, 19). Contamination by virulent phages may result in the lysis of starter strains in the vat, causing slow fermentation or even complete starter failure with consequent loss of the product. *Streptococcus thermophilus*, a gram-positive thermophilic lactic acid bacterium, is used as a starter in the yogurt and cheese industry. All of the *S. thermophilus* bacteriophages isolated until now (1, 12, 15, 16, 18, 20; for a review, see reference 4) belong to group B as defined by Bradley ((2); Siphoviridae) and share DNA homology. During the last 30 years, about 80 bacteriophage isolates were collected by our company from batches in which fermentation failures occurred with *S. thermophilus* strains used in yogurt and cheese factories situated in various European countries. About 40 yogurt factory-derived phages were classified into two subgroups by serology, host range analysis, and hybridization with subgroup-specific DNA sequences (5), thus documenting rather limited genetic variability. In contrast, cheese-derived phages could not be classified into subgroups within 41 phages analyzed, 35 distinct restriction patterns and 34 distinct host ranges were detected (5). This apparent variability of *S. thermophilus* phages poses problems for practical phage control, especially for mozzarella fermentation. In fact, due to phage problems alternative lactic acid bacterial starters for mozzarella production in Italy are currently being explored (9).

Ecological studies that explore the origin of phage diversity are lacking. Therefore, we collected data from a single cheese factory during a longitudinal survey to determine whether the diversity of the bacteriophage population in the cheese factory arose from environmental sources or from genetic changes of the resident bacteriophage. A mozzarella factory which used a complex, undefined mixture of starter strains and open, continuous-fermentation tanks was chosen for this study. Such a factory should be very susceptible to invasion by phages from environmental sources. If a single phage invaded the factory and experienced genetic changes during its residence, one would predict a family of closely related phages with similar restriction patterns and one or a few overlapping host ranges. In addition, one would expect an accumulation of point mutations, deletion and recombination events as a consequence of rapid evolution of the phages, and changes in the restriction patterns during the survey period. Alternatively, if the diversity of phages arose before entry into the factory, one would predict a wide range of restriction patterns and clearly distinct host ranges. The factory phages should be genetically as different from each other as any other set of randomly collected phages. The data from the longitudinal survey, presented in the current communication, demonstrate little, if any, evidence in favor of genetic changes of a resident phage during the observation period.

Demonstration of identity between factory and raw-milk phages would be the clearest proof for the origin of phage diversity in the environment. Such proof is, however, difficult to provide, as the phage concentration in raw milk is likely to be very low and one will only see phages for which susceptible cells exist in the starter cultures. In addition, cheese factories tend to be infested with moderate to high titers of phages (5) and will therefore be refractory to invasion by new raw-milk phages. To circumvent these technical difficulties, we conducted an intervention study by replacing the complex, undefined starter system with a defined starter composed of bacterial strains which were unable to propagate the phage population.
resident in the factory. Three possible outcomes were considered. Under the strongly selective pressure of the new hosts, the resident phages could undergo rapid adaptation to the new hosts. This scenario would indicate a remarkable genetic flexibility of the resident phages. Alternatively, the resident phages could be wiped out and replaced by new phages. An ecological shift should then identify the likely source of the new phages (e.g., raw milk, starter cultures, operators, or airborne phage transmission). Finally, if the environmental phage load is low, the factory may run for an extensive time period without phages. The outcome of our intervention trial was the observation of rapid colonization of the factory by new phages which could be traced back to raw milk. This therefore indicates that the diversity of phages observed in a cheese factory reflects the diversity of phages normally found in their natural environment.

MATERIALS AND METHODS

Longitudinal survey. For the longitudinal survey, we chose a mozzarella cheese factory situated in northern Italy. A "madre naturale" starter system was used for cheese fermentation. In this artisan technique, the starter is derived from natural microbial flora of local raw milk samples. Milk was sterilized several cycles to select for thermophilic bacteria. The bacterial population is then further selected for proteolytic and acidifying properties. The starter is thus a complex population consisting mainly, but not exclusively, of an undefined mixture of \( \alpha \) thermophilus strains. Ten batches of such madre naturale starters, the strain compositions of which were unknown to the cheese maker, were used in rotation, with each batch being used for 1 day of fermentation. The raw-milk source and the processing procedure did not change over the study period. Therefore, phage variability could be studied solely as a function of time.

The study was conducted between December 1993 and February 1996. The factory processed about 500,000 liters of milk daily in a continuous and open fermentation system. Raw milk was introduced and processed in the factory. The milk was heat treated for 15 s at 74°C before transfer to the fermentation vat. Phages were titrated and isolated from cheese whey samples collected 2 h after addition of the starter strains.

Strain and phage identification. Individual bacterial colonies from the starter mixture were isolated on M17 agar (Difco) containing 5 g of lactose per liter and propagated at 40°C in M17 broth (23). For the analytical purpose of the longitudinal study, randomly selected colonies of \( \alpha \) thermophilus strains composing the madre naturale starter were first classified according to acidification properties.

Filter-sterilized cheese whey samples (0.1 ml) were tested for the presence of bacteriophages in 10 ml of M17 medium inoculated with 10\(^{6}\) CFU of colony-purified cells grown in milk. All of the resulting phage isolates were purified by two rounds of single-plaque isolation. Propagation of the phages was then carried out in 40°C M17 broth supplemented with 5 g of lactose per liter and 0.5 g of CaCl\(_2\) per liter.

In a second phase, bacteria not lysed by these phages were identified and characterized. These samples were screened for phages attacking these strains. Strain typing was repeated until we were able to type more than 95% of randomly selected colonies from the starter mixture.

Intervention study. Five defined starter cultures, each composed of two \( \alpha \) thermophilus strains, were chosen for the intervention trial (February to April 1996). None of the 10 strains, designated N to W, was susceptible to the phages isolated from the factory during the longitudinal survey. Eight of the strains could be induced by phages from our phage collection, but no overlapping susceptibility was observed. Each starter culture was used for 1 day of fermentation and then replaced by the next. For technological reasons (superior fermentation properties of strains from the previous starter population), 2 weeks after the intervention the number of starter mixtures was increased to nine, four of the eight additional strains (B, E, G, and K) were susceptible to phages from the longitudinal survey, and for three strains we had no homologous phages in our phage collection. Apart from alterations to the starter system, no other parameter (processing, raw-milk source) was changed. For 2 months, whey samples from the factory were screened daily for the presence of phages.

Biochemical methods. Phage purification, phage DNA extraction, restriction enzyme digestion, agarose gel electrophoresis, and Southern and blot hybridizations with individual PstI-XhoI restriction fragments labeled with the Random Primed DNA Labeling kit of Roehringer (Mannheim, Germany) were done as described previously (3). Chromosomal DNA was labeled from the starter cells as described by Slos et al. (22). The purified DNA was digested with restriction enzymes PstI, EcoRI, or HindIII, Southern blotted, and probed with radioactive \( 32^P \)ATP in accordance with the manufacturer's protocol. Primers 3' (\( 5'\)-GATACCTCGATAGTCC-3') and 4' (\( 5'\)-TTGGT AAGTTCACAGG-3') or variants of \( a \) were used for comparative sequencing.

PCR. PCR products were prepared by using synthetic oligonucleotides A and D, described previously (5), or primers 1 and 2 (\( 5'\)-GGTACTAATTGTA GC-\( 3'\), 2' (\( 5'\)-CACCATTGGTGTTATC-3'), purified phage DNA, and Super Tag Polymerase (Stehelin, Basal, Switzerland). PCR products were purified by using the Qiaquick Spin PCR purification kit (Qiagen) in accordance with the manufacturer's protocol.

Phages: biological properties. Systematic phage counts were done at regular intervals by using group A to M indicator cells defining the different lytic groups. Phages were found in all of the whey samples tested over the 2-year period. Phage titers ranged from \( 10^4 \) to \( 10^9 \) PFU/ml. Any given whey sample contained phages belonging to different lytic groups: the minimum was four different groups, while the maximum was eight. Marked daily variations in phage titers were observed which were due to the starter rotation. Each phage strain was named with a lowercase letter according to the host strain on which it was isolated, which was indicated with an uppercase letter. If more than one type of phage strains were isolated, a number was added as a suffix.

Phages attacking lytic group B and G cells were the most prevalent isolates: 35 and 23% of all isolates belonged, respectively, to these two groups. Phages attacking lytic group A, D, E indicator cells could be differentiated from group G and M cells possessing identical ribotypes. On EcoRI digestion, group E indicator cells could be differentiated from group G and M cells (data not shown). About 100 colonies from the different starters were tested for the presence of phage by the mitomycin C induction test, and no lysogenic starters were identified.

RESULTS

Longitudinal survey: host cells. A total of 13 different \( \alpha \) thermophilus strains, A to M, defined by their distinct phage susceptibility, were isolated from the undefined starters used during the longitudinal survey. Determination of the strain composition of the starters led to the observation that different starter cultures possessed common strains, thus compromising the idea of starter rotation. These 13 strains represented about 95% of the randomly selected colonies. The remaining 5% of the colonies could not be further differentiated by phage typing, since they were not attacked by phages from either the cheese whey or our phage collection. Four of the strains represented 85% of the randomly selected colonies (B, 36%; G, 19%; E, 16%; H, 14%). On PstI digestion of genomic DNA, all indicator cells showed distinct ribosomal DNA restriction patterns (ribotypes), with only lytic group E, G, and M indicator cells possessing identical ribotypes.

Phages: biological properties. Systematic phage counts were done at regular intervals by using group A to M indicator cells defining the different lytic groups. Phages were found in all of the whey samples tested over the 2-year period. Phage titers ranged from \( 10^4 \) to \( 10^9 \) PFU/ml. Any given whey sample contained phages belonging to different lytic groups: the minimum was four different groups, while the maximum was eight. Marked daily variations in phage titers were observed which were due to the starter rotation. Each phage strain was named with a lowercase letter according to the host strain on which it was isolated, which was indicated with an uppercase letter. If more than one type of phage strains were isolated, a number was added as a suffix.

Phages attacking lytic group B and G cells were the most prevalent isolates: 35 and 23% of all isolates belonged, respectively, to these two groups. Phages attacking lytic group A, D, E, and L cells were only occasionally isolated (each group representing less than 5% of all isolates). Phages growing on lytic group C, E, F, H, I, K, and M cells were found at intermediate frequencies. During the 2-year survey period, no significant shifts in the distribution of the phages among the different lytic groups were observed.

During the 2-year period of the longitudinal survey, several hundred phage isolates were plaque purified and their host ranges were investigated by using lysis of bacterial cells in liquid medium as the test. The phages isolated defined the 13 lytic groups A to M of starter strains described above. Phage assays confirmed that there were, with one exception, no over-
lapping host ranges. The exceptions were phages øh and øl, which could be grown on three strains (H, L, and our Sfil indicator cells [4]). However, the efficiency of plating was about $10^3$ to $10^4$ times higher on the last propagating host than on the other two strains. As øh and øl showed similar restriction patterns (Fig. 1A, right half) and as strains H and L possessed distinct ribotype patterns, the observed differences might indicate the presence of a restriction-modification system.

Molecular analysis of phages. With the exception of phages øc and øf and phages øh and øl, all of the phages showed distinct restriction patterns when tested with the restriction enzymes PvuII (Fig. 1) and EcoRI (data not shown). øf and øh could be distinguished with EcoRI but not with PvuII. Independent phage isolates attributed to the same lytic group showed either identical (Fig. 2A) or highly related (Fig. 2B) restriction patterns. For example, four and two variant øh and ød isolates, respectively, were observed. Each isolate differed from the other in only one (ød.1 and ød.2) or, maximally, two (øh.1 to øh.4) fragments. The molecular basis for the variant restriction pattern is unknown. The variant øh isolates, for example, were grown on their factory starter H, thus excluding effects of different restriction-modification systems. Major deletion events, as previously observed in S. thermophilus øS121 (6), are also unlikely. A panel of primers was used to amplify DNA regions of these phages by PCR. All of the PCR products obtained for the øh variants indicated that they were identical in size (data not shown). In addition, øh isolates with different PvuII restriction patterns did not differ when digested with EcoRI (data not shown).

All of the phages isolated during the survey belonged to one DNA homology group based on the observation that when øg (or, alternatively, øh, øe, or ød) DNA was used as a radiolabeled probe in Southern blot hybridization, many PvuII restriction fragments of phages belonging to other lytic groups gave a hybridization signal (Fig. 1). All hybridization experiments were done under high-stringency conditions (3). A high degree of overall DNA homology has already been described for S. thermophilus øS121 (6).
which differed at 56 positions from the prototype sequence, and evey (3). All of the factory phages except èg and èk hybridized with thermophilus phages from our collection (4). To obtain better Vol. 63, 1997 MOLECULAR ECOLOGY OF S. THERMOPHILUS PHAGE INFECTIONS 3147

To obtain better phages in our collection isolated over more than 30 years from other at only 3 positions. PCR products from nine independent phages attacking starter strains A, B, C, E, G, H, I, L, and M were detected, with the highest titer being obtained for phages attacking starter strain G (maximal titer, 10^9 PFU/ml; Fig. 4). The whey samples contained two to seven distinct types of phages with a minimal phage titer of 10^3 PFU/ml. During the first 2 weeks, only starters which were unable to propagate the resident phage population were used. As a consequence, the resident phage contamination level fell: the maximal phage titer was 10^3 PFU/ml, and 18 of 26 whey samples had no detectable phages. This is in stark contrast to the situation of the 2-year survey period during which a phage-free whey sample was never observed. However, only 5 days after the introduction of the new starter system, two "new" phage types were observed (öö and öb; Fig. 4). By now we mean that phages of this host range were not detected in whey samples taken 1 week, 1 year, and 2 years before the intervention. Seven days after the introduction of the new starter system, a third new phage type was detected (öö). The phage titre rose quickly: a titer of 10^7 PFU of öp per ml was observed on day 10 (Fig. 4).

In the second phase of the intervention trial, eight further starter strains were added to the new starter system. This was done for technological reasons. Four of these strains were susceptible to attack by four distinct phage types (öb, öc, ög, and ök) of a phage population previously resident in the factory. This gave us the opportunity to test for recombination of the factory by what remained of the previous phage population. During this second phase, we observed regularly and at high titers two of the new phage types, öo and öp, which apparently got established in the factory (Fig. 4). öo was observed for only 1 week. No other new phages were isolated during the 50-day follow-up. All four strains susceptible to the previous resident phage population once again experienced phage infections. Phages attacking starter G were sampled before the intervention (day -3) and at three time points (days 18, 31, and 59) after the reintroduction of starter G into the factory. The ög isolates taken before and after the intervention did not differ in their EcoRI, XbaI, and HindIII restriction patterns (Fig. 5A).

Origin of the new phage types. The appearance of the new phage types gave us the opportunity to trace the origin of these phages. The restriction profile of up to six independent isolates of each of the new phage types (öö, öö, and öö) was analyzed. Each type presented a unique restriction pattern (Fig. 5B). The restriction patterns of the new phages showed no relatedness with each other or those of the previous resident bacteriophage population. The PCR fragments of the putative DNA primase genes of the new phages were sequenced. They differed from each other and from öö at a maximum of two nucleotide positions (Fig. 3). We postulated that the raw milk delivered to the factory during the intervention trial is the most likely source of these new phage types, since we have detected small amounts of phages in several raw-milk samples. Phages attacking starter cells belonging to lytic groups A to M were found in raw milk on only two occasions (10 and 100 PFU/ml against starters B and M, respectively). However, the raw-milk öö isolates differed in restriction pattern (Fig. 5B) and sequence analysis (data not shown) from the resident öö. In contrast, the phages attacking strains N, O, and P, which were detected with titers between 10 and 130 PFU/ml in raw-milk samples, possessed restriction patterns identical to those of new factory phages öö, öö, and öö (Fig. 5B). Phages from our phage

thermophilus phages from our collection (4). To obtain better resolution of their genetic relationship, we used individually labeled PvuII-XbaI (PX) restriction fragments from dSF21 in dot blot hybridization assays against the phages from the survey (3). All of the factory phages except öö and öö hybridized with fragment PX1, all hybridized with fragment PX2, and most hybridized with fragments PX4 and PX5, while only moderate levels of cross-hybridization were seen with fragments PX3 and PX6 (Table 1).

Previously, we described a conserved DNA module (5) localized in fragment PX1 from dSF21. Sequence analysis revealed a DNA replication module in this region (10). Within this DNA module, we targeted a 546-nucleotide-long region for PCR and comparative sequencing. This region was localized in an open reading frame coding for a putative DNA primase (10).

Alignment of these sequences demonstrated a series of graded relatedness (Fig. 3). Maximal sequence diversity was 65 nucleotide differences (12%), but many phage isolates showed much fewer or no sequence differences over this region. Phages öö and öö gave identical sequences which differed at only one base position from the prototype dSF21 DNA sequence. öö, öö, and öö differed at 15 to 17 positions from the prototype sequence, while they differed from each other only at two positions. Identical sequences were found for öö and öö, which differed at 56 positions from the prototype sequence. öö and öö differed from the reference DNA at 63 and 62 nucleotide positions, respectively, while they differed from each other at only 3 positions. PCR products from nine independent öö isolates, which were the predominant phages during the survey, yielded identical nucleotide sequences. DNAs from phages in our collection isolated over more than 30 years from different European countries (4) covered the same range of diversity as the phages recovered during 2 years from a single factory (data not shown).

<table>
<thead>
<tr>
<th>Phage</th>
<th>Degree of cross-hybridization* with fragment:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PX2</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
<tr>
<td>öö</td>
<td>++</td>
</tr>
</tbody>
</table>

* The phages investigated are listed in the first column. The second to seventh columns give the hybridization with the indicated labeled PvuII-XbaI restriction fragments of our prototype dSF21 DNA; the fragments are ordered according to their positions on the restriction map of dSF21 (3). Fragments PX2 and PX8 were not used due to their small size. A sequencing project has identified the following functions of the dSF21 fragments: PX1, DNA replication module (10); PX6, antirepression (7); PX3, lysozyme module, lysis (8); PX4, tail (7) morphogenesis; PX5, head (8) morphogenesis; PX2, cos (cohesive end) site, DNA packaging (7).
collection, which were isolated from a different cheese factory in northern Italy, that attacked strains O and P showed a restriction pattern distinct from those of the raw-milk and factory phages (Fig. 5B). Interestingly, one raw-milk ßm isolate, but not the resident factory ßm isolate, possessed a restriction pattern that was very similar, although not identical, to that of the new factory ßm isolate (Fig. 5B). All ßm isolates also yielded plaques on host strain M (efficiency of plating, \(10^{-3}\)), while the different ßm isolates showed an efficiency of plating of less than \(10^{-3}\) on host strain N. S. thermophilus strains susceptible to ßm were detected in raw milk, while we could not detect the host cells for ßo and ßp in raw milk during the intervention trial.

**DISCUSSION**

In general, S. thermophilus phages isolated from cheese factories show much greater diversity than those from yogurt fermentation. An explanation for this may lie in the differences of starter strains and fermentation conditions. Since phage propagation is dependent on the presence of suitable host bacteria and yogurt factories use a small number of well-defined starter cultures, a greater variety of phage types is expected in cheese factories, using mainly undefined and probably complex cultures, than in yogurt factories. In addition, due to different raw-milk heat treatment regimens and differences in factory design and product processing (closed versus open fermentation vats), yogurt factories are much less susceptible to invasion by phages from environmental sources than are cheese factories. The most critical process in cheese making is whey separation, which inevitably leads to aerosol-borne phages in the factory. As the production line in some cheese factories is open, it allows the possibility of long-term maintenance of phages in the factory.

The aim of the present study was to settle the question of whether the diversity of phages obtained from a cheese factory reflects the natural diversity present in the environment or phage diversification in the factory. For this purpose, a two-
stage ecological study was conducted. In the first stage, we monitored a cheese factory using an undefined starter system over a 2-year period. Based on host range analysis, 12 distinct phage types were observed. This large variety of phage types confirmed our initial observation of the diversity of phages detectable in cheese factories. According to Southern and dot blot hybridization studies, all of the factory phages were closely related. This observation does not, however, prove a single phage invasion event in the factory, since all of the *S. thermophilus* phages investigated until now and originating from different countries have shown extensive cross-hybridization (1, 3, 5, 12, 18, 20).

With two exceptions, phages defining different lytic groups showed clearly distinct restriction patterns while multiple isolates of phages attacking the same host, sampled over the 2-year observation period, showed identical or highly related restriction patterns. Four variants of 06 were observed. Even if the differences were relatively minor, in terms of genetic evolution the observation might be significant. The molecular basis for the differences between phages showing highly related restriction patterns is unknown. Similar differences between closely related raw-milk phages were found. Variant phages have been previously observed for our prototype *S. thermophilus* phage 85521 during serial passages in the laboratory. In that case, the differences in the restriction patterns reflected site-specific deletions in three regions of the phage genome (7).

Sequence analysis in a conserved region of the genome from multiple isolates of the same phage type revealed no point mutations. In contrast, up to 12\% nucleotide sequence diversity was observed between different phage types. This diversity was as great as that between the most different phages from our collection.

The observations of our survey do not support a model that postulates a single phage invasion event and subsequent diversification of the phage during its residence in the factory. On the one hand, such a rapid evolution of phages in a cheese factory requires a very high mutation rate in *S. thermophilus* phages, and on the other, a strong selection imposed by the starter system is required for new phage types to become detectable. It is not apparent how an undefined starter system like that used in our longitudinal survey could provide this selection.

In the second part of our factory study, intervention with defined starter cultures exerted strong selective pressure on...
the factory phage population: the new starter cultures could not propagate the resident phages. Therefore, any phages isolated on the new host cells must be either host range mutants of the resident phages or else new phages. Interestingly, the phage population changed quickly following the intervention. Within 5 to 7 days, three apparently new phages were isolated that attacked 3 of the 10 new starter strains. Similar data have been reported in ecological surveys of cheese factories using lactococcal starters (17). Phages φ4 and φ6 quickly became established in the factory. The restriction patterns of the new phages indicated that they were not derivatives of the previous resident phages. For all three new phages, raw milk was the most likely source, as phages with identical host ranges and restriction patterns were found in the raw milk delivered to the factory during the intervention trial. Phages in raw milk were apparently also the source for phage problems in cheese factories using lactococcal starters (13, 14, 17), demonstrating the general applicability of our results for cheese factories. The reports on lactococcal phages differ, however, in some important aspects from our data. In these reports, no proof of molecular identity between raw-milk and factory phages was provided. In addition, in two reports raw-milk lactococcal phages appeared to represent induced prophages from lysogenic strains (13, 14). Lysogenic S. thermophilus starters are relatively rare (5) and were not detected at all in our factory survey. In New Zealand cheese factories, the whey was sprayed on pastures where the dairy herds that provided the raw milk for the factory grazed (17). This procedure could contaminate the raw milk with factory phages. In our factory, the whey was concentrated in the factory and sent abroad for further processing, thus excluding recycling of the factory phage via raw milk.

Theoretically, the intervention should wipe our the resident phages from the factory unless they are able to mutate to new host ranges. The resident phage population was, however, not eliminated after 10 days of use of nonsusceptible starters. Retroduction of susceptible strains led to the rapid reappearance of the resident phage, indicating that phages can be maintained in the factory for at least 3 weeks in the absence of their host strain in the starter system.

In summary, data from our surveys identified the environment, specifically, raw milk, as the source of phages which contaminated the factory. We have no evidence for genetic changes in the factory phages during the survey period. The absence of recombination within the factory phage population appears surprising in view of laboratory experiments which reported frequent recombinations between phages after double infections of starter cells (12). This may be due to the lack of overlapping host ranges, resulting in a kind of reproductive isolation which could be very important in the design of starter systems in the cheese industry. Apparently, all of the genetic diversity observed in S. thermophilus phages during our survey was already present in the environment outside of the factory. Therefore, detailed knowledge of the genetic diversity of S. thermophilus phages and their hosts within their natural environment is essential for the success of any phage control measures in cheese factories.

ACKNOWLEDGMENTS

We thank S. Foley for proofreading of the manuscript and valuable discussion and the factory staff at Nestlé for their cooperation during the survey.

We thank the Swiss National Science Foundation for financial support of F. Desiere and S. Lucchini in the framework of its Biotechnology Module (grant 3100-044545/1).

REFERENCES

10. Desiere, F., S. Lucchini, A. Bruttin, M.-C. Zwahlen, and H. Brüssow. A highly conserved DNA replication module from Streptococcus thermophilus is similar in sequence and topology to a module from Lactococcus lactis phages. Virology, in press.
5 Characterization of the Lysogeny DNA Module from the Temperate *Streptococcus thermophilus* Bacteriophage Sfi21

Published in *Virology* 1997 Jun 23;233(1):136-48

Bruttin A, Desiere F, Lucchini S, Foley S, Brüssow H
Characterization of the Lysogeny DNA Module from the Temperate Streptococcus thermophilus Bacteriophage φSfi21

Anne Bruttin,1 Frank Desiere,1 Sacha Lucchini, Sophie Foley, and Harald Brussow2

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

Received December 28, 1996; returned to author for revision March 7, 1997; accepted April 15, 1997

Phage φSfi21, the only temperate Streptococcus thermophilus phage from our phage collection, showed extensive DNA homology with virulent phages from lytic group I. Southern blot hybridizations demonstrated that the φSfi21-specific DNA was clustered in an approximately 6.6-kb-long region, the putative lysogeny module. Sequence analysis and database research identified an integrase within this module: orf 203 with homology to an anonymous orf 258 from the temperate lactococcal phage BK5-T; orf 127 and orf 122 with weak homology to the N- and C-terminal parts, respectively, of the c-like repressor from lactococcal phages Tuc2009 and BK5-T; orf 75 with homology to a repressor protein from lambdoid phage 434 and an anti-repressor ant with homology to phage P1. The molecular arrangement of the predicted orfs in phage φSfi21 was very similar to that of the lactococcal phage BK5-T. The transition from φSfi21-specific DNA into DNA shared with virulent phages was abrupt and flanked at one side by notable DNA repeats. Sequence analysis identified a holin protein to the left of the lysogeny module. A site-specific deletion of 2.4 kb, which reproducibly transformed φSfi21 into a lytic phage, was localized in the lysogeny module. It was flanked at both sides by conspicuous DNA repeats. Sequence analysis identified a holin protein to the left of the lysogeny module. A site-specific deletion of 2.4 kb, which reproducibly transformed φSfi21 into a lytic phage, was localized in the lysogeny module. It was flanked at both sides by conspicuous DNA repeats. One repeat region reflected the DNA around the attP site, while the other reflected the putative genetic switch region between repressor and anti-repressor genes. S. thermophilus host Sf1 transformed with a plasmid containing int and orf 203 showed resistance to superinfection by heterologous phages, but not by the homologous φSfi21. Part of the int gene could be deleted without loss of this activity, while a deletion in orf 203 resulted in loss of the phage resistance. We speculate on the possibility of a bipartite immunity system for the control of lysogeny in φSfi21.

INTRODUCTION

An understanding of the evolution of bacteriophages is of both theoretical and practical value. We decided to address some of the questions on phage evolution using phages of Streptococcus thermophilus, a gram-positive lactic acid bacterium used in milk fermentation (Mercenier, 1990). Phage infection has always been a major problem in industrial fermentation, especially in the dairy industry (Peitersen, 1991). Contamination by virulent phages may result in the lysis of bacterial starter strains, causing slow fermentation or even complete starter failure with consequent loss of the product. S. thermophilus phage control measures have been hampered by the large variability of phages encountered in the environment (Benbadis et al., 1990; Brussow et al., 1994a; Neve et al., 1989; Prevots et al., 1989) and the lack of knowledge of their genome organization (Brussow and Bruttin, 1995). Analysis of phages from our phage collection covering 30 years of industrial fermentation (Brussow et al., 1994a) and from an ecological survey (Bruttin et al., 1997) indicated that very many, perhaps several hundred, distinct S. thermophilus phages may exist in the environment.

From work with phage λ it has been inferred that the various races of phage λ are related by recombination events. Careful analysis of lambdoid phages by sequencing and heteroduplex analysis revealed that homologous and heterologous genome segments were interspersed (Campbell and Botstein, 1983). Where sequence data locate the transitions between homology and heterology on the genetic map, the transition points frequently lie strikingly close to the boundaries of functional segments of the genome (Casjens et al., 1992). This observation gave rise to the concept of phage evolution by exchange of modules, where a module implies a stretch of genes of related function (Susskind and Botstein, 1978).

Some observations of S. thermophilus phages can be interpreted within the module exchange hypothesis (Brussow et al., 1994b). Systematically, dot blot hybridization studies (Brussow and Bruttin, 1995) showed a close genetic relationship between temperate phage Sfi1 and lytic group I phages of S. thermophilus in that the temperate phage Sfi1 differed from these lytic phages only in the possession of a single contiguous DNA fragment. The φSfi21-specific DNA is flanked at both sides by conserved DNA sequences of lytic group I phages (Brussow and Bruttin, 1995). Furthermore, the φSfi21-specific DNA fragment was also found in another temperate S. ther-

1 Both authors contributed equally to this work
2 To whom correspondence and reprint requests should be addressed. Fax: 0041 21 785 8925. E-mail: Harald.Brussow[a]NESTR.DCH.
mophilus phage, TP-J34 (Neve et al., 1997), but, in contrast, the rest of the genome of this phage showed little or no homology with φSfi21 (Brüssow and Bruttin, 1995). The temperate phage Sfi21 could thus have been created through the acquisition of a hypothetical lysogeny module by the genome of a lytic phage.

Interestingly, upon serial passages of phage φSfi21 we repetitively observed a 2.4-kb deletion which transformed the temperate phage into a lytic phage (Bruttin and Brüssow, 1996). As the deletions occurred at exactly the same nucleotide positions in several independent mutants, one may suspect an enzymatic mechanism behind this deletion process. We envisioned two hypotheses. First, the deletion process is the result of a hypothetical site-specific recombinase responsible for exchanging genome segments in the process of module shuffling. Alternatively, the deletion process resulted from an error of the phage-encoded integrase which used the phage attP site and, instead of the bacterial attB site, a secondary att site on the phage DNA for prophage integration. Such an accidental integration process has already been demonstrated for phage λ (Davis and Parkinson, 1971).

To settle this question and to further our understanding of the genetic organization of the lysogeny module from S. thermophilus phages we decided to sequence the φSfi21-specific DNA.

**MATERIALS AND METHODS**

**Phages, strains, and media**

Phage Sfi21 was propagated on S. thermophilus strain Sfi1 in lactose M17 broth as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995). Escherichia 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. Lytic phages from our phage collection (Brussow et al., 1994a) were used for comparative Southern blot hybridization.

**DNA techniques**

Phage purification, DNA extraction, agarose gel electrophoresis, Southern blot hybridization, and labeling of plasmid DNA were done as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995). General DNA techniques were performed as described by Sambrook et al. (1989). 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.

**Cloning**

Phage Sfi21 DNA was cut with restriction enzymes EcoRI, HindIII, NsiI, or XbaI and cloned into pUC19 vector or the E. coli/lactococcal/streptococcal shuttle vector pNZ124 (Plattetew et al., 1994). In addition, total phage Sfi21 DNA was cut with restriction enzyme Sau3A and used for shotgun cloning into pUC19.

**Sequencing**

DNA sequencing was started with the universal forward and reverse primers of pUC19 or pNZ124 and continued with synthetic oligonucleotide (17-mer) primers (Microsynth, Switzerland). Cloned DNA was sequenced on both strands by the Sanger method of deoxy-terminated 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° for 30 sec, 50° for 30 sec, and 72° for 1 min.

pUC19 clones containing Sau3A-digested phage Sfi21 DNA were sequenced by 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. No Sau3A clones were obtained for the middle of the XbaI fragment 3.

**PCR**

PCR was used to prepare a template for sequencing the middle part of XbaI fragment 3 that could not be obtained by cloning. PCR products were prepared using the indicated synthetic oligonucleotide pair (5′-3′ primer 1, CGTTTCAAGGCGTGGGC, and primer 2, CGCTCGCGCTTTAGGTT), 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 97.0; EMBL (Abridged), Release 48.0; PIR-Protein, Release 50.0; SWISS-PROT, Release 33.0; PROSITE, Release 13.0) using the FastA (Pearson and Lipman, 1988) and BLAST (Altschul et al., 1990) programs. Sequence alignments were performed using the CLUSTAL W 1.6 method (Thompson et al., 1994) and the BLOCKMAKER program (Henikoff et al., 1995). The integrase tree was built with the program Alignment of GeneWorks (V2.0) from IntelliGenetics Inc.
RESULTS

Cloning of DNA fragments affected by the 2.4-kb deletion

Phage Sfi21 DNA was cut with a number of restriction enzymes which had been mapped and the resulting fragments were cloned into plasmid vectors. No clones were obtained for the left and right end of the genome containing the cos sites and a central part of the genome in which the type 2 deletion was localized (Bruttin and Brussow 1996). To bridge the central gap, sequence information from the left and rightmost available clones bordering the gap were obtained and the unclonable DNA fragment was amplified by PCR. Partial sequencing of the PCR product identified a BglII site in the middle of the unclonable DNA fragment. A BglII/XbaI fragment of the phage DNA, representing the right half of this unclonable area and covering the DNA deleted in the type 2 deletion mutants (Bruttin and Brussow, 1996), was then successfully cloned into the vector pNZ124, yielding plasmid pPX3R (for BglII/XbaI fragment 3, right part) and sequenced. In contrast, the 3.1 kb of phage DNA to the left of the BglII restriction site still could not be cloned and therefore a PCR product was sequenced. The localization of the holin gene in this DNA region (see below) might explain why this DNA was unclonable.

FIG. 2. Definition of the lysogeny module in φSfi21 by Southern blot hybridization against a panel of 26 virulent phages. (A) Left, a representative restriction pattern (BamHI/HindIII) of S. thermophilus phages (listed below) as revealed by ethidium bromide fluorescence. Right, Localization of the hybridization probes A to D on the map presented in Fig. 1. (B to D) Southern blots of the EcoRV/XbaI (B, left) and EcoRV (B, right, C and D) digests of the indicated phages hybridized with probes A (B, left), B (B, right), C (C) and D (D). Phage (lane) coding: φSfi21 (1), φSfi3 (7), φSfi18 (11); φS5 (15), φSfi9 (2), φSfi9 (17), φS4 (20), φS6 (21); φST3 (6), ST12 (23), ST44A (3), ST128 (16), ST130 (27); φ8 (5), C (13), E (18), F (26), H (6), I (19), L (24), M (4); φL-A3 (9), L-A4 (10), L-A9 (22), L-A12 (12), L-F1 (14).
The deletion is flanked by DNA repeats

PCR using primers located to the right and left side of the deletion yielded an approximately 5-kb DNA product for the wild-type phage and an amplification product, smaller by about 2 to 3 kb, for the phage deletion mutants D and E (Bruttin and Brussow, 1996). Partial sequencing of these PCR products from several independent phage mutants revealed identical deletion sites: nucleotide (nt) position 3418/19 (left deletion site) and nt position 5809/10 (right deletion site; Fig. 1). The deletion was flanked at both sides by conspicuous direct and inverted DNA repeats. Within the 150-nt region surrounding the left deletion site four 5-nt repeats (TAAAA) are followed by a further four 5-nt repeats (TTTTA), which are the inverted repeats of the first. In addition, one 10-nt-long inverted repeat showing one mismatch was identified (Fig. 1). Within 300 nt of the right deletion site two distinct direct repeats, one 10 and another 16 nt long, were observed. In addition, two different 11-nt-long inverted repeats were detected. No homology could be found between the DNA sequences, including the repeats, at the left and right deletion sites.

Within the analyzed DNA fragment, such a concentration of DNA repeats was rather specific to the deletion sites. Two other areas of the analyzed phage DNA possessed notable DNA repeats: 20- and 26-nt-long direct repeats preceded by 12- and 13-nt-long inverted repeats, respectively, were detected between nucleotide positions 1430 and 1535, and 1995 and 2417 (Fig. 1).

Prediction of open reading frames and homology searches

The sequenced DNA and the adjacent XbaI fragment reported previously (Bruttin and Brussow, 1996) were investigated for open reading frames (orf) longer than 50 codons and using ATG start codons. Nineteen orfs potentially coding for proteins ranging from 54 to 359 aa were detected (Fig. 1). The orfs were labeled according to the number of aa in the predicted protein. Only 7 orfs were preceded by a standard S. thermophilus ribosomal binding site (GAG; Guédon et al., 1995) in appropriate spacing with respect to the start codon. It should be noted that the GAG consensus is based on relatively few sequences of S. thermophilus, but not of its bacteriophages. Even within the eps gene cluster of S. thermophilus only 9 of 14 genes are preceded by the standard GAG sequence (Stingle et al., 1996). Analysis of the orfs revealed that between nt positions 1 and 3250 the majority of the predicted orfs pointed to the right, between nt positions 3250 and 6100 all but one predicted orf pointed to the left (i.e., they were transcribed from the opposite strand), while all orfs between nt positions 6200 and 8100 pointed again to the right.

The orfs were translated into a protein sequence and the databases were screened for homologous proteins with the BLAST P program. Possible functions could be attributed to several orfs (for details see below under the specific subheadings). orf 359 demonstrated a high degree of homology to the integrase family of proteins. The adjacent orf 203 showed homology to an anonymous orf 258 from the temperate lactococcal phage BK5-T. Notably, orf 258 was also found adjacent to the int gene from BK5-T (Boyce et al., 1995). orfs 122, 127, and 75 demonstrated weak to moderate homology with several phage and bacterial repressors, while orf 287 showed a good homology to an antirepressor of phage P1 as reported previously (Bruttin and Brussow, 1996). In the left part of the analyzed DNA fragment orf 87a showed homology with holins and orf 288 showed weak homology with orf 259 from lactococcal φBK5-T.

Definition of the module

According to the DNA homology searches the contiguous DNA fragment between nt positions 3400 and 7200 represents functions essential for the regulation of the lysogenic life style of φSfi21. As the putative holin is a gene function common to lytic and temperate phages, we suspected the transition between the putative lysogenic and lysis modules somewhere between nt positions 500 and 3400. To define the borders of the lysogeny module, Southern blot hybridizations were done using defined φSfi21 DNA fragments as probes against 26 lytic phages covering the whole range of our phage collection. When the XbaI fragment 4 (nt positions 5590 to 8130) was used as a probe, only φSfi21, meaning none of the 26 lytic phages hybridized (Fig. 2B), indicating that ot 87 b (ending at nt 7830) belonged to the lysogeny module. According to comparative Southern blot hybridizations the left side of the φSfi21-specific DNA extended to nt 110 (starting at nt 1759) (Fig. 2B). The transition from φSfi21-specific DNA to DNA shared between the temperate and lytic phages occurred between nt 288 and nt 110. The transition point was in fact very sharp as a probe covering nt positions 1269 to 1564 hybridized with 22 of 26 lytic phages (Fig. 2C), while a probe covering nt positions 1564 to 2219 hybridized exclusively with phage φSfi21 (data not shown). The left border of the lysogeny module was preceded by a 20-nt-long direct repeat and an 11-nt inverted repeat.

Fig. 3. Multiple sequence alignment of the deduced product of φSfi21 orf 359 and various integrases. Amino acids identical to the φSfi21 integrase are in bold. The integrases shown are from Staphylococcus aureus phages phi-11 (Accession No. M34832), phi-42 (U01872), and phi-13 (X82312); Lactobacillus gasseri phage adh (M26679), Lactobacillus delbrueckii phage m4 (P20710), Lactococcus lactis phage r11 (U38906) and transposon Trn22 (L27649), Enterococcus faecalis transposon Tri16 (M37184). Amino acids that are conserved in at least 8 of the 9 sequences are boxed. Domains I and II represent conserved motifs of integrase and resolvase proteins.
**Mycobacterium smegmatis** phage L5

**Staphylococcus aureus** phage phi-11

**Streptomyces ambofaciens** (int)

**Streptococcus pneumoniae** Tn1545

**Enterococcus faecalis** Tn916 (rec)

**Bacillus thuringiensis** Tn4430 (res)

**Klebsiella pneumoniae** (int)

**Serratia marcescens** (int)

**Weeksella zoohelcum** (int)

**Bacillus subtilis** (codV)

**Lactobacillus leichmanii** (xerC)

**Escherichia coli** (xprB)

**Haemophilus influenzae** (int)

**E. coli λ integrase**

**Lactobacillus delbrueckii** phage adh

**Lb. delbrueckii** phage mv4

**Staphylococcus aureus** phage L54a

**Lc. lactis** Tn5276

**Staphylococcus aureus** phage phi-13

**Staphylococcus aureus** phage phi-42

**Lc. lactis** phage BK5-T

**Lc. lactis** phage Tuc 2009

**Lc. lactis** phage phi-LC3

**Lc. lactis** phage r1t

**S. thermophilus** phage SFi21

---

**FIG. 4.** Tree analysis of ψSFi21 integrase and indicated integrases of various bacterial origin. The tree was built using the ALIGNMENT program of GeneWorks (V2.0) from Intelligentic Inc. The numbers show the cost of the alignment. Lb, *Lactobacillus*; Lc, *Lactococcus*; S, *Streptococcus*.

A further DNA probe covering orf 87 (holin) and 288 hybridized with 13 of 26 lytic phages (Fig. 2D). With this probe only phage SFi18 showed an identical hybridization pattern to phage SFi21.

**Integrase**

Alignment of the ψSFi21 int homologue (translation of orf 359) with different members of the int gene family.
LYSGENY DNA MODULE FROM S. thermophilus

(Fig. 3) showed a close relationship with the integrases of Lactococcus lactis bacteriophage r1 (30.6% aa identity; van Sinderen et al., 1996) and Staphylococcus aureus bacteriophage phi-42 (26.5%; Carroll et al., 1995). More than 20% aa identity was found with integrases from Lactobacillus delbrueckii subsp. bulgaricus bacteriophage mv4 (22%; Dupont et al., 1995). Lactobacillus gas-seri bacteriophage adh (23.4%; Raya et al., 1992) and Lactococcus lactis transposon Tn5276 (20.3%). 17.5% aa identity was found in comparison with integrases from the Enterococcus faecalis transposon Tn916 and the Staphylococcus aureus bacteriophage phi-11. Homology was not restricted to the conserved domains 1 and 2, but was scattered throughout the entire protein sequence (Fig. 3).

Tree analysis of integrases showing homology with φSfi21 Int defined three major branches represented by Mycobacterium smegmatis phage L5 (Lee et al., 1991), phage λ, and a branch to which φSfi21 Int belonged (Fig. 4). A subbranch of the third group was composed exclusively of Lactococcus lactis phage integrases and the integrase from our temperate S. thermophilus phage.

Repressor

The predicted protein from orf 127 showed weak homology (15% aa identity, P=0.0029) with the cl-like putative repressor of the temperate lactococcal phages Tuc2009 (van de Guchte et al., 1994). Orf 127 showed two clear repressor motives. We identified a helix-turn-helix motif showing homology with phage λ cl repressor (Pabo, 1992; Fig. 5A). In addition, a motif showing similar-

FIG. 5. Alignment of the helix-turn-helix motif (left) and the RecA cleavage site (right) found in the φSfi21 cl homologue (REP, upper line in bold) with cl-type repressors from Lactococcus lactis phages BK5-T (L44933) and Tuc2009 (L26219). Bacillus subtilis catabolite repressor CcpA (L47838), E. coli src open reading frame (P24242), and transcriptional repressor of Pseudomonas aeruginosa (L06240). Bottom left, the corresponding sequence from the cl protein of phage λ is shown. Amino acids identical to the φSfi21 cl homologue are shown in bold. *Conserved in all sequences; + conserved in four other sequences; **conserved in three other sequences. Amino acid positions conserved in all but one sequence were boxed.
ity with the RecA cleavage site (Sauer et al., 1982) homology region of the lactococcal phage repressors was identified, although the Ala/Gly cleavage site was not preserved (Fig. 5B). One should note, however, that the E. coli and B. subtilis repressors also did not possess this Ala/Gly site. Interestingly, the S. thermophilus phage cl homologue was half the size of its lactococcal homologue (Tuc2009, 286 aa) or of the phage Λ cl repressor (236 aa).

Cro equivalent and antirepressor

The lytic/lysogenic life cycle decision in phage Λ is realized by a competitive interplay between the cl repressor and the Cro protein for operator binding. The operators are situated between the cl and cro genes, which are adjacent to each other, but are transcribed in opposite directions (Gussin et al., 1983). A similar control circuit is found in lactococcal phage r11 (Nauta et al., 1996). In phage SF121 the topological homologue of cro is orf 75 (Fig. 1). The predicted protein product of orf 75 showed homology to the N-terminal part of the cl repressor protein of the lambdoid phage 434 (Fig. 6), namely the helix-turn-helix motif with which this protein binds DNA (Modragon et al., 1989). A similar degree of homology was seen with a putative DNA binding transcription repressor of a Pseudomonas aeruginosa phage (Fig. 6). The observation of direct repeats between orf 127 and orf 75 (Fig. 1), and the prediction of DNA binding activity for both proteins, may identify this DNA region by analogy with other phage systems as the genetic switch region of phage SF121. The topological equivalent of cro, orf 75, overlaps orf 287, which is in a different reading frame. The predicted protein product of orf 287 showed good homology to the antirepressor of phage P1 (Bruttin and Brussow, 1996) and thus may also be involved in the lytic/lysogenic life cycle decision.

Orf 203 and phage resistance

The predicted protein product of orf 203 showed 21.7% aa identity with that of orf 258 from lactococcal phage BK5-T (Boyce et al., 1995; Fig. 7). This homology was further underlined by an identical position in the phage genome since in both cases the putative gene was situated between the int and the putative cl genes. A gene of unknown function situated between int and cl genes was also described for the lactococcal phage rlt (van Sinderen et al., 1996). Orf 203 was of interest to us since we have observed a phage resistance phenotype associated with the plasmid pPX3R, possessing the phage SF121 int gene and orf 203.

When our indicator S. thermophilus strain Sfi1, which is susceptible to 21 distinct phage strains, was transformed with the plasmid pPX3R we observed a significant reduction in the efficiency of plaquing (e.o.p. \( \geq 10^{-3} \)) for 12 of the 21 phages compared to the vector control (Table 1).

In some cases the exact titer reduction was difficult to evaluate as the phage plaque size was substantially reduced on Sfi1 cells transformed with plasmid pPX3R. The pattern, but not the strength, of phage resistance observed by this plasmid was similar to that conferred by the prophage SF121 in the lysogenic Sfi1 host (Table 1). Phages ΦST28, ST25, ST44A, ST33, ST17, Sfi3, SF19, and S17 were only incompletely or not at all inhibited by the SF121 prophage in the lysogenic Sfi1 host, which thus defined phages belonging to a different immunity group. Interestingly, phages ΦST28, ST25, ST44A, ST17, Sfi3, SF19, and S17 were also incompletely inhibited (e.o.p. \( \geq 10^{-3} \)) by the plasmid pPX3R. A notable difference to the lysogen-mediated immunity function was that the plasmid pPX3R did not protect Sfi1 cells against infection with homologous ΦSF121 or its deletion derivative, ΦS3. It should be noted that a totally different pattern of phage resistance was observed when an anonymous 4-kb DNA fragment from the lytic phage ΦS17 was introduced into the Sfi1 host (data not shown). To demonstrate further the specificity of the phage resistance conferred by the plasmid pPX3R, several other DNA fragments from the temperate ΦSfi121 were introduced into the Sfi1 host and failed to confer a protection against phage infection (data not shown).

In order to pinpoint the factor responsible for the ob-
TABLE 1
Titration of the Indicated S. thermophilus Phages on Normal SF11 Host and Its Lysogenic Derivative (SF11.cl6) and SF11 Cells Transformed with Plasmids pNZ124 and pPX3R

<table>
<thead>
<tr>
<th>Phage</th>
<th>SF11</th>
<th>SF11.cl6</th>
<th>pNZ124</th>
<th>pPX3R</th>
<th>pM227</th>
</tr>
</thead>
<tbody>
<tr>
<td>phi4M-15</td>
<td>3 x 10^4</td>
<td>&lt;10</td>
<td>5 x 10^4</td>
<td>&lt;10</td>
<td>1 x 10^5</td>
</tr>
<tr>
<td>phi4M-31</td>
<td>4 x 10^4</td>
<td>&lt;10</td>
<td>1 x 10^4</td>
<td>&lt;10</td>
<td>7 x 10^5</td>
</tr>
<tr>
<td>phiST23</td>
<td>9 x 10^4</td>
<td>9 x 10^4</td>
<td>2 x 10^6</td>
<td>5 x 10^5</td>
<td></td>
</tr>
<tr>
<td>phiST25</td>
<td>9 x 10^4</td>
<td>3 x 10^4</td>
<td>2 x 10^6</td>
<td>2 x 10^5</td>
<td></td>
</tr>
<tr>
<td>phiST42</td>
<td>10^5</td>
<td>&lt;10</td>
<td>2 x 10^3</td>
<td>1 x 10^4</td>
<td></td>
</tr>
<tr>
<td>phiST44</td>
<td>2 x 10^5</td>
<td>&lt;10</td>
<td>2 x 10^3</td>
<td>1 x 10^4</td>
<td></td>
</tr>
<tr>
<td>phiST44A</td>
<td>8 x 10^5</td>
<td>4 x 10^4</td>
<td>2 x 10^6</td>
<td>5 x 10^4</td>
<td></td>
</tr>
<tr>
<td>phiST33</td>
<td>2 x 10^5</td>
<td>6 x 10^4</td>
<td>3 x 10^4</td>
<td>1 x 10^5</td>
<td></td>
</tr>
<tr>
<td>phiST17</td>
<td>3 x 10^5</td>
<td>3 x 10^4</td>
<td>7 x 10^3</td>
<td>2 x 10^4</td>
<td></td>
</tr>
<tr>
<td>phiST3</td>
<td>10^6</td>
<td>&lt;10</td>
<td>2 x 10^7</td>
<td>&lt;100</td>
<td></td>
</tr>
<tr>
<td>phiS96</td>
<td>2 x 10^6</td>
<td>&lt;10</td>
<td>2 x 10^7</td>
<td>&lt;100</td>
<td></td>
</tr>
<tr>
<td>phiSF131</td>
<td>2 x 10^6</td>
<td>4 x 10^6</td>
<td>9 x 10^4</td>
<td>8 x 10^6</td>
<td></td>
</tr>
<tr>
<td>phiSF11</td>
<td>2 x 10^6</td>
<td>2 x 10^6</td>
<td>3 x 10^4</td>
<td>2 x 10^6</td>
<td></td>
</tr>
<tr>
<td>phiS89</td>
<td>5 x 10^6</td>
<td>&lt;10</td>
<td>2 x 10^7</td>
<td>3 x 10^7</td>
<td></td>
</tr>
<tr>
<td>phiS19</td>
<td>4 x 10^6</td>
<td>&lt;10</td>
<td>6 x 10^7</td>
<td>2 x 10^6</td>
<td></td>
</tr>
<tr>
<td>phiS69</td>
<td>2 x 10^6</td>
<td>40</td>
<td>6 x 10^7</td>
<td>4 x 10^7</td>
<td></td>
</tr>
<tr>
<td>phiS118</td>
<td>9 x 10^5</td>
<td>&lt;10</td>
<td>2 x 10^6</td>
<td>6 x 10^8</td>
<td></td>
</tr>
<tr>
<td>phiS3</td>
<td>2 x 10^6</td>
<td>&lt;10</td>
<td>8 x 10^6</td>
<td>&lt;10^9</td>
<td></td>
</tr>
<tr>
<td>phiSF121</td>
<td>4 x 10^6</td>
<td>&lt;10</td>
<td>2 x 10^7</td>
<td>4 x 10^7</td>
<td></td>
</tr>
<tr>
<td>phiSF12</td>
<td>8 x 10^6</td>
<td>&lt;10</td>
<td>4 x 10^6</td>
<td>5 x 10^6</td>
<td></td>
</tr>
<tr>
<td>phiS17</td>
<td>6 x 10^6</td>
<td>5 x 10^6</td>
<td>4 x 10^4</td>
<td>&lt;10</td>
<td></td>
</tr>
</tbody>
</table>

Note. The phage titrations (given as PFU/ml) on SF11 and SF11.cl6 on the one hand and SF11 host transformed with plasmids pNZ124 and pPX3R on the other hand were done at different times with different phage stocks. Titers reductions by a factor of 1000 are underlined.

* pinpoint plaques at the limit of detection.

Observed bacteriophage resistance, deletions were generated in the int gene and orf 203. The 498-bp BfiI–EcoRV fragment was deleted from pPX3R, yielding the clone pSF101, in which the int gene was disrupted. Subsequent digestion of pSF101 with ClaI and NspI, followed by religation, resulted in the construct pS145, which possessed a deletion of 212 bp within orf 203. When these clones were tested in the SF11 bacterial host we could assign the phage inhibitory action to orf 203 (Table 2). The strength of this inhibitory action was also tested by chromosomal integration of the PX3R DNA fragment in SF11. The resulting strain was bacteriophage sensitive (Table 2), indicating that orf 203 is not effective as a single copy.

Holm

Orf 87a at the 5’ end of the sequenced area showed significant homology with holins from two lactococcal phages, Tuc2009 (Arendt et al., 1994) and phi-LC3 (Birkeland, 1994) (34.5% aa identity). The aa identity was even higher (36.8%) when compared with Staphylococcus aureus phage phi-11 (Weerakoon and Jayaswal, 1995) (Fig. 8). A very lysine-rich N-terminus was observed for the putative phiS21 holin but this was not the case for the other holins.

The adjacent orf 288 showed homology to orf 259 from lactococcal phiBK5-T: over a 35-aa stretch in the N-terminal half of the predicted phiBK5-T protein 12 identical positions were seen in the protein predicted for phiS21 (Fig. 9). In several lactococcal phages the lysin gene was found adjacent to the holin gene (Arendt et al., 1994; Birkeland, 1994); however, orf 288 showed no homology to phage lysins.

DISCUSSION

Since loss and gain of modules are a major driving force for phage evolution (Casjens et al., 1992; Susskind and Botstein, 1978), elucidation of the underlying mechanisms will be of substantial interest. As the 2.4-kb site-specific deletion in phage phiS21 was flanked at both sites by conspicuous repeats, we tested whether the deletion process could be the result of a site-specific recombinase responsible for module shuffling, which in this case effects the transformation of a temperate into a lytic phage. If the observed deletion is a consequence of a hypothetical recombinase implicated in module-shuffling, one might expect the deletion to define a functional module. The definition of the lysogeny module and a closer examination of the DNA repeat sequences should elucidate this.

The data presented in this report demonstrated that the putative lysogeny module comprises at least 6.6 kb of genetic information. The length estimation of the lysogeny module was based on Southern blot hybridization experiments which differentiated phiS21-specific sequences (approximate nt positions 1550 to 26100) from sequences shared with lytic phages (nt positions 1 to 1150). phiS21-specific DNA sequences do not necessarily define lysogenic functions, but a minimal extension of the lysogeny module is given by the attribution of likely functions to the identified orfs by database searches. Between nt positions 3400 and 7200, sequence analysis identified genes which could potentially encode for an

TABLE 2
Titration of S. thermophilus Phage S89 on Normal SF11 Host and SF11 Cells Transformed with the Indicated Plasmids

<table>
<thead>
<tr>
<th>Host</th>
<th>Phage titer (PFU/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF11</td>
<td>3 x 10^8</td>
</tr>
<tr>
<td>SF11 pNZ124</td>
<td>7 x 10^7</td>
</tr>
<tr>
<td>SF11 pPX3R</td>
<td>&lt;100</td>
</tr>
<tr>
<td>SF11 pSF101</td>
<td>&lt;100</td>
</tr>
<tr>
<td>SF11 pSF45</td>
<td>1.6 x 10^8</td>
</tr>
<tr>
<td>SF11 2.8'</td>
<td>4 x 10^9</td>
</tr>
</tbody>
</table>

* Sf 2.8 is Sf1 containing an integrated copy of PX3R DNA.

LYSOGENY DNA MODULE FROM S. thermophilus 145
integrate, a cl-like repressor, a Cro-like protein, and an antirepressor. Furthermore, the distribution and orientation of ORFs in phage φSfi21 were remarkably similar to that in the lactococcal phage BK5-T (Boyce et al., 1995).

The noticeable repeats which flank the type 2 deletion are unlikely to represent recognition sites for a module-recognition recombinase. Immediately downstream of the left deletion site we detected the int gene with the 3' end of the gene pointing to the deletion site. Classically the attP site is found directly downstream of the int gene (Carroll et al., 1995). In bacteriophages from lactic acid bacteria the attP site is typically surrounded by a complicated array of direct and inverted repeats (Boyce et al., 1995; Dupont et al., 1995; Liliehaug and Birkeland, 1993; van de Guchte et al., 1994). The repeats at the left side of the deletion are thus not recognition elements for a hypothetical module recombinase, but likely recognition elements for prophage integration. The right deletion site was also flanked by conspicuous DNA repeats which are most likely genetic elements involved in the genetic switch between the lytic and lysogenic modes of phage infection (see below). The type 2 deletion in phage Sfi21 is thus probably due to an Int-mediated cross-over event between the phage attP site and a secondary att site in a nonessential region of the phage genome. For a lytic phage derivative of phage Sfi21 any DNA region within the lysogeny module becomes necessarily a nonessential region. Similar deletion mutants have been isolated for phage λ by Davies and Parkinson (1971). Many deletions of heat-resistant phage λ mutants had one end point at the phage attP site and another more variable endpoint in secondary att sites. Current work in our laboratory confirms that the left deletion site of the type 2 spontaneous phage deletion mutants is identical to one end of the core sequence shared between the attB and attP sites (manuscript in preparation).

A phage inhibitory activity was associated with plasmid pPX3R and its derivatives containing an intact orf 203 from φSfi21. The orf 203-mediated activity showed similarity to the resistance of the lysogenic host against superinfection with heterologous phages. As no cl-like gene was present on this plasmid we may have identified a second phage repressor system in φSfi21, which is independent of the cl-like repressor. This phage repressor is, however, not efficient against the homologous φSfi21 and is only effective against heterologous phages when present in the multicopy state.

In phage λ, the cl repressor is counteracted by the Cro protein, a small one-domain protein of about 70 aa in length with DNA binding activity. The cro gene is found adjacent, but in the opposite orientation to the cl gene. A set of operators to which cl and Cro bind is found between the cl and cro genes (Gussin et al., 1983). A similar control circuit was observed in the lactococcal phage r1 (Nauta et al., 1996). Orf 75 in φSfi21 is the topological equivalent of cro. As the gene products from both orf 127 and 75 showed DNA binding motifs and as a region with DNA repeats was situated between both

| hol  | DDEKEVY | 96     |
| LysA| ENK     |
| Tuc 2009 | ENK+   |
| phi-11 | DEGREE |

FIG. 8. Alignment of a 35-aa stretch predicted for orf 203 from φSfi21 with that predicted for orf 259 from lactococcal phage BK5-T (Accession No. L44593). Identical amino acids are indicated in bold; + represents amino acid similarity.
genes, we might have identified the genetic switch region of φSIII21. Orf 75 overlaps the ant gene, potentially coding for an anti-repressor, in the same orientation but in a different reading frame. Therefore one might speculate on the possibility of a bipartite immunity system for the control of lysogeny in φSIII21.

ACKNOWLEDGMENTS

We thank Josette Sidoti for valuable technical assistance and the Swiss National Science Foundation for financial support of Frank Desiere, Sacha Lucchini, and Sophie Foley in the framework of its Biotechnology Module (Grant 5002-04454/1).

REFERENCES


6 Comparison of the Lysogeny Modules from the Temperate Streptococcus thermophilus Bacteriophages TP-J34 and Sfi21: Implications for the Modular Theory of Phage Evolution.

Published in Virology. 1998 Feb 1;241(1):61-72.

Comparison of the Lysogeny Modules from the Temperate *Streptococcus thermophilus* Bacteriophages TP-J34 and Sfi21: Implications for the Modular Theory of Phage Evolution

Horst Neve,* Kirsten I. Zenz,* Frank Desiere,† Andreas Koch,* Knut J. Heller,* and Harald Brüssow†

*Institut für Mikrobiologie, Bundesanstalt für Milchforschung, Hermann-Weigmann-Straße 1, D-24103 Kiel, Germany; and †Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blancs, CH-1000 Lausanne 26, Switzerland

Received September 1, 1997; returned to author for revision October 23, 1997; accepted November 19, 1997

A 76-kb DNA segment covering the putative lysogeny module of the pac-site-containing temperate *Streptococcus thermophilus* bacteriophage TP-J34 was sequenced. Sequence alignment with the lysogeny module from the cos-site-containing *S. thermophilus* bacteriophage Sfi21 revealed areas of high sequence conservation (e.g. over the int gene), interspersed with regions of low or no sequence similarity (e.g. over the cro gene). Four of the six sharp transition zones from high to low sequence conservation were found within open reading frames coding for the CI repressor, the Anti repressor, the Immunity protein, and a protein of unknown function. The transition points in the cl and ant genes appear to separate gene segments coding for distinct functional domains of these proteins. In addition, these two transition points were located at or near the deletion sites observed in spontaneous phage Sfi21 deletion mutants, thus suggesting these transition points as recombinational hotspots. Furthermore, the sequence at the transition point in the cl gene resembles the attachment site of the phage, suggesting the involvement of the phage integrase in at least some of the exchange reactions. Contrary to the initial formulation of the modular theory of phage evolution the unit of the evolutionary exchange in streptococcal phages is not a group of functional genes, but can be as small as a single gene. Exchange reactions can also occur within genes, possibly between gene segments encoding distinct protein domains.

INTRODUCTION

About 3000 different bacterial viruses have been described in the scientific literature (Ackermann and DuBow, 1987). There is no doubt that the described isolates represent only a minute fraction of the bacterial viruses actually present on earth (Suttle et al., 1990). Further complexity is provided by the genomic diversity of phages belonging to any well-defined phage family (Casjens et al., 1992; Monod et al., 1997; Repoila et al., 1994). In fact, individual phage isolates that are indistinguishable from the well-studied "type isolates" are rare—bacteriophage λ, for example, has been found in nature only once (Casjens et al., 1992). Botstein (1980) proposed a model of phage evolution that could account for this surprising genetic variability of bacteriophages. According to this "modular theory" of bacteriophage evolution, new phage types are readily created by recombination processes. The theory was formulated on the basis of experimental data from the lambda phage system. Careful analysis of lambdoid phages by sequencing and heteroduplex analysis revealed that homologous and heterologous genome segments were interspersed (Campbell and Botstein, 1983). Where sequence data locate the transitions between homology and heterology on the genetic map, the transition points frequently lie strikingly close to the boundaries of functional segments of the genome (Casjens et al., 1992). Similar examples have also been found outside the lambdoid phages, e.g., in the closely related pair of phages T5 and BF23 (Mondigler et al., 1996).

The essentials of Botstein's theory were stated as follows (Botstein, 1980). The product of evolution is not a given virus but a family of interchangeable genetic elements (modules) each of which carries out a particular biological function; exchange of a given module for another occurs by recombination among viruses belonging to the same interbreeding population; and finally, these viruses and their hosts can differ widely in many characteristics. Due to its success in explaining the genetic diversity of lambdoid (Casjens et al., 1992) and T4 phages (Kutter et al., 1996; Monod et al., 1997; Repoila et al., 1994), the modular theory of phage evolution has been accepted as a standard hypothesis for the evolution of bacteriophages in general.

Also in extensively characterized phage systems the recombination systems which promote the exchange of modules are largely unknown (Casjens et al., 1992; Monod et al., 1997). It was speculated (Casjens et al., 1992) that modular exchanges might be mediated by homologous recombination using either highly con-
served motifs within coding sequences or short regions of homology (microhomology) in regulatory elements. Site-specific recombination could also be involved in shuffling modules. Finally, illegitimate recombination must be considered which fortuitously results in the joining of two nonhomologous DNAs in a manner that is functional.

In general, little is known about the natural populations and ecology of the few phages which have been genetically and biochemically well characterized. In contrast, phage systems of industrial importance for which ecological data are available have rarely been investigated for their evolutionary relationship. We decided to address some questions of phage evolution and phage population biology with phages from Streptococcus thermophilus, a Gram-positive bacterium (Mercenier, 1990) used industrially in milk fermentation (lactic acid bacterium). Several phage collections are available for this phage family (Benbadis et al., 1990; Brüssow et al., 1994a; Fayard et al., 1993; Larbi et al., 1992; Le Marrec et al., 1997; Neve et al., 1989; Prevots et al., 1989) and the ecology of phage infections has been studied (Bruttin et al., 1997a). Finally, a molecular characterization of the temperate S. thermophilus phage 63F21 was started (Brüssow et al., 1994b; Brüssow and Bruttin, 1995; Bruttin and Brüssow, 1996; Bruttin et al., 1997b,c; Desiere et al., 1997, 1998), making this phage system an acceptable subject for the study of phage evolution. This experimental system has already yielded two interesting observations. First, the temperate and the lytic S. thermophilus phages showed a close genetic relationship (Brüssow et al., 1994b; Brüssow and Bruttin, 1995; Desiere et al., 1998) except for a 6.6-kb region, the putative lysogeny module (Bruttin et al., 1997b). Second, sequence analysis of S. thermophilus phages revealed striking similarities to phages from other bacterial genera (Desiere et al., 1997, 1998). In the present communication we provide a functional and genetic characterization of the putative lysogeny module of an additional S. thermophilus phage, 6TP-J34 (Neve et al., 1998). We compare this sequence with that of 63F21 (Bruttin et al., 1997b) and analyze it with respect to the predictions of the 'modular theory' of bacteriophage evolution.

RESULTS

The lysogeny module of 6TP-J34

The attP site was physically located on the genome of 6TP-J34 (Neve et al., 1997). Phage DNA covering this genome region was cloned and a 7.6-kb DNA segment was sequenced. The sequenced DNA was investigated for open reading frames (orf) longer than 50 codons and using ATG start codons. Fourteen orfs potentially coding for proteins ranging from 51 to 359 amino acids (aa) were detected (Fig. 1). Four adjacent orfs pointed to the left, while the remaining 10 orfs pointed to the right (i.e., they were transcribed from the opposite strand).

Database search

The orfs were translated into protein sequences and the databases (excluding 63F21 sequences) were screened for similar proteins using BLASTP.

Orf 359 gene product (gp) demonstrated a high degree of similarity to the integrase family of proteins and was closest to integrases from Lactococcus lactis phages (phage r11, score = 217, P = 10^-17). None of the gps from orfs to the left of orf 359 showed a significant similarity (criterion: score ≥ 80 and P value ≤ 0.001) to proteins from the databases. In contrast, four proteins predicted from orfs to the right of orf 359 gave significant matches with predicted proteins from the databases.

Orf 122 gp shared 30 aa with orf 3 gp of unknown function from the L. lactis bacteriophage TP901-1 (Christiansen et al., 1996; 25% identity; score = 89, P = 0.00017). Both genes occupied a comparable topological position: they were located two genes upstream of the int gene. In addition both proteins showed the zinc-binding motif VxxHE(L)GH (Fig. 2) characteristically found in metalloproteinases (Jongeneel et al., 1999). However, metalloproteinases also share aa outside of this motif (Murphy et al., 1991), which was not the case for the orf 122 gp.

Orf 121 gp showed significant similarity to a number of bacteriophage and bacterial repressor proteins. Similarities with bacteriophage repressors included repressor protein c2 from the lambdoid phage P22, the PBSX repressor of a defective prophage from Bacillus subtilis, the putative repressors from the Lactobacillus phage 82, and the L. lactis phage Tuc2009 (all scores > 100, P < 10^-5) and the CI repressor of the lambdoid phage 434 (P = 0.0003). In a multiple alignment the N-terminal half from orf 121 gp showed clear similarity with these repressors (Fig. 3A). The similarity was with the DNA-binding region from the c2 repressor which showed a helix-turn-helix motif (Fig. 3A; Sevilla-Sierra et al., 1994). In addition, weaker similarity was seen over the C-terminal end from orf 121 gp with the same repressor proteins (data not shown).

Orf 67 gp showed a significant similarity to an anonymous protein from Methanococcus jannaschii with 22 shared aa (33% identity; score = 97, P = 10^-5). In addition, weak similarities (P = 0.02) to two phage proteins were found: orf 79 gp from L. lactis phage 8K5-T, located in the vicinity of the putative 8K5-T repressor (Boyce et al., 1995), and the CI repressor of the lambdoid phage 880. In the alignment with the CI repressor of the lambdoid phage 880, the similarity was over the helix-turn-helix motif (Fig. 4). The alignment of the orf 67 gp with CI-like and Cro-like proteins from two well-studied lambdoid phages, Cro-like proteins from two phages of lactic acid bacteria, and several bacterial proteins led to...
FIG. 1. Alignment of the putative lysogeny modules from *S. thermophilus* phages Sfi21 (top line) and TP-J34 (bottom line). The predicted open reading frames are indicated with their orientations and are marked by their length in amino acids. White arrows indicate ORFs for which significant matches were found in the databases. Putative functions are indicated next to the arrow. The numbers above and below the alignments give the base pair numbers of the compared sequences. The colored bars give the percentage of nucleotide sequence similarity between both sequences as defined in the color code at the top of the figure. The figure was created with the SIM Alignmen tool and Lanlview (Huang and Miller, 1991; http://www.expasy.ch). The vertical lines indicate the transition zones from high to low sequence similarity between the two phages. The transition zones from highly conserved to low or no sequence similarity are marked by letters A to F. D1 and D2 refer to spontaneous deletions found in mutant phages.
the definition of a block of conserved aa localized in the N-terminal half of the orf 67 gp (Fig. 4). Since the gene products from the divergently orientated orfs 121 and 67 revealed homology to Cl- and Cro-like regulatory proteins, respectively, the 168-bp intergenic region was examined for further "genetic switch" motifs (see Fig. 9). This region contains two divergently orientated putative promoter elements showing high homology to *E. coli* consensus promoters (Mulligan et al., 1984). A 15-bp direct repeat with internal dyad symmetry, which may function as a putative operator site (O₁₇), is overlapping the rightward promoter sequence. An identical site (O₁₈) is located in front of orf 238 specifying the putative TP-J34 anti-repressor (Fig. 5).

Orf 238 gp showed, mainly over the N-terminal half, similarity with a hypothetical protein from *Haemophilus influenzae* (score = 199, \( P = 10^{-19} \)). Over the C-terminal half the orf 238 gp demonstrated similarity with the Ant and Kla proteins from coliphage P1 (coding, respectively, for the anti-repressor and a phage protein lethal to the host cell) as well as the DNA-binding protein ROI from the lambdoid phage HK022 (scores > 150, \( P < 10^{-19} \), Fig. 6). Orf 238 gp showed similarity to the helix-turn-helix motif from the coliphage proteins, except for a 5-aa insertion (Fig. 6).

According to the significance criteria that we have used no other predicted proteins from \( \Phi \)TP-J34 showed similarities to predicted proteins from the databases. However, orf 142 gp showed a similarity to *L. lactis* chromosomal fragment (clone SS38), which allowed excretion/secretion of the *E. coli* \( \beta \)-lactamase in lactococci. This similarity was at the borderline of our significance.

**FIG. 2.** Alignment of the central part of the predicted orf 122 gp from \( \Phi \)TP-J34 with orf 3 gp from *Lactococcus lactis* bacteriophages TP901-1 (Genbank Accession No. X95213) showing the conservation of the zinc-binding motif VxxHE(L)G from metalloproteases (boxed).

**FIG. 3.** A Multiple alignment of the \( \Phi \)TP-J34 orf 121 gp with (from above to below) the c2 repressor proteins from the lambdoid phage P22 (P03035, the helix-turn-helix motifs are indicated for this repressor), the putative repressor from *L. lactis* phage Tuc209 (L26192), an anonymous protein from *Clostridium perfringens* (X37369), orf 127 gp from *S. aureus* (P07666), the PBSX repressor and a putative repressor from a defective *Bacillus subtilis* prophage (M36477, P23789), the putative repressor from the Lactococcus phage gfe (X88308), the \( \alpha \)s operon repressor from *B. subtilis* (Q07683), the yycG gene fragment of *Pseudomonas aeruginosa* (L05240), and the transcription regulator Y4D from *Rhizobium* sp. (P03035). The aa positions identical with the \( \Phi \)TP-J34 orf 121 gp are given in bold. The multiple alignment was done using the Computational Biochemistry server at ETH Zurich (http://cbgr.inf.ethz.ch). (B) Alignment of the \( \Phi \)TP-J34 orf 121 gp with orf 127 gp from *S. aureus* and localization of the deletion in the spontaneous \( \Phi \)Sto21 deletion mutant D2.
FIG 4 Multiple alignment of the φTP-J34 orf 67 gp with (from above to below) orf 75 gp from φSfi21, the CI and Cro repressors from the lambdoid bacteriophage 434 (Y00118, P03036), the anonymous proteins from Methanococcus jannaschii (U67482) and Agrobacterium radiobacter (X95394), CI protein from bacteriophage phi-80 (X19065), PvuI restriction endonuclease regulator (A41879), an anonymous protein from Clostridium perfringens (X87369), orf 79 gp from the L. lactis phage BK5 (L44593), and a repressor from Streptococcus pneumoniae (L29324). Aa positions identical to the φTP-J34 orf 121 gp are given in bold.

Comparison with predicted φSfi21 proteins

Several orfs from φTP-J34 showed significant similarities to orfs from φSfi21 located at comparable topological positions (Fig. 1). However, the degree of relatedness of the predicted proteins varied (Table 1). Between the two phages orf 87, orf 122, and orf 359 (integrate) gsp showed no aa differences, while orf 93 gps showed one aa difference. As expected for identical integrases, PCR analysis (data not shown) demonstrated that φTP-J34 integrated its genome into the mRNA gene of the chromatid DNA precisely as described previously for φSfi21 (Bruttin et al., 1997b). Sequencing of the PCR products revealed a common core sequence of 40 bp for attB, attR, attL, and attP of φTP-J34 that was molecularly identical to that of φSfi21.

Other predicted proteins, while clearly related, showed more aa variation (orf 111, 103a, 238, and 121 gps, Table 1). Two different patterns of variation were observed. On the one hand the respective orf 111/orf 110 gp comparison indicated aa sequence variability throughout the predicted proteins from both phages, while on the other hand the respective orf 103a, 238, and 121 gp comparisons showed significant clustering of aa differences in either the N-terminal parts (orf 238, orf 103a gps) or the C-terminal part (orf 121 gp) of the predicted proteins (Figs. 3B and 6; Table 1).

Alignment of the two phage nucleotide sequences

The lysogeny module in φSfi21 extends from the 3' end of orf 77 to the 3' end of orf 87b (Bruttin et al., 1997b). Along, but not beyond, this region a global alignment between the two S. thermophilus phage DNA sequences could be established (Fig. 1). The alignment of the φSfi21 and φTP-J34 nucleotide sequences revealed regions of high sequence similarity interspersed with regions of low or no sequence similarity (Fig. 1). Three regions of
FIG. 6 Multiple alignment of the 4TPJ3-ori238 gp with (from above to below) orf 287 gp from Sfi21, KilA protein from bacteriophage PI (P19653), ROI protein from bacteriophage HK022 (U4336), and an anonymous protein from Haemophilus influenzae (P44193). The aa positions identical to ATPJ3 orf238 gp are indicated in bold. The genetic information deleted from the spontaneous D1 (Buttin and Brussow 1998) is marked by a bended arrow. The alignment was done according to Corpet (1988).

0.5 to 1.3 kb in length showed high conservation between the two phages (region 1 covered the int gene and part of orf 103a, region 2 covered orf 122 and part of orf 121, region 3 covered part of orf 238 and orfs 93 and 87). Highly conserved DNA sequences were defined in this study as contiguous DNA segments of longer than 200 bp which did not show two consecutive base changes and which possessed an overall sequence diversity of less than 5%. All six transition zones from high to low or no sequence conservation (marked by letters A to F in Fig. 1) were sharply defined (Fig. 8). Four of these six sharp transition zones lay within coding sequences (5')
TP-J34 and Sfi21 Lysogeny Modules

FIG. 7. Alignment of the N-terminus of orf 142 gp with the L. lactis protein SS38. The signal sequences of the five functionally characterized, phage-encoded lipoproteins are shown. Putative or proven cleavage sites (*) are indicated just in front of the essential acylated cystein residue (+1). The four-residue consensus lipoprotein box (L, V, I|A, S, T, G)(G, A)C is shown at the bottom of the figure. Identification of a signal sequence as that of a lipoprotein requires at least one match to the first two and precise matches to the final two aa residues of the consensus (von Heijne, 1989).

end of orf 103a: Fig. 8A; 3’ end of orf 142: Fig. 8B; middle of orf 121: Fig. 8D; middle of orf 238: Fig. 8E). No consensus sequence could be defined when the DNA sequences from the different transition points were aligned. Interestingly, three transition points lay in the vicinity (< 70 bp apart) of deletion points observed in spontaneous phage mutants (Figs. 1, 8D, and 8E) or plasmids containing cloned phage DNA (Fig. 8C).

DISCUSSION

With the presented sequence data from two temperate S. thermophilus phages we wanted to test predictions of the modular theory of bacteriophage evolution. The lysogeny module was chosen because the Sfi21 lysogeny module contained genes which are similar to well-characterized genetic elements from intensively investigated phages (Bruttin et al., 1997b). All S. thermophilus phages characterized until now, whether temperate or virulent, share DNA sequence similarity (Brüssow and Bruttin, 1995). Despite this uniformity, the differences between any two phages can be substantial. The two phages of our study, Sfi21 and TP-J34, isolated in France and Germany, respectively, differ from each other in their phage DNA packaging mechanism: Sfi21 is a cos-site phage (Neve et al., 1998), while TP-J34 is a pac-site phage (Neve et al., 1998), while Sfi21 is a cos-site phage (Brüssow and Bruttin, 1995). Further differences are that lytic growth of TP-J34 is inducible by a temperature upshift (Neve et al., 1998) and that TP-J34

TABLE 1

Comparison of the Lysogeny Modules from Phage Sfi21 and TP-J34 at the Sequence Level

<table>
<thead>
<tr>
<th>ORF TP-J34</th>
<th>ORF Sfi21</th>
<th>aa changes (%)</th>
<th>Nucleotide conservation (%)</th>
<th>Putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>110</td>
<td>19 (17)</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>103b</td>
<td>n.a.</td>
<td>n.a.</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>n.a.</td>
<td>n.a.</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>n.a.</td>
<td>n.a.</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>103a</td>
<td>140b</td>
<td>15 (14.6)</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>103a (1–22)</td>
<td></td>
<td>10 (45)</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>103a (23–103)</td>
<td></td>
<td>4 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>359</td>
<td>359</td>
<td>0</td>
<td>99</td>
<td>Integrate</td>
</tr>
<tr>
<td>142</td>
<td>n.a.</td>
<td>n.a.</td>
<td>44</td>
<td>Lipoprotein?</td>
</tr>
<tr>
<td>122</td>
<td>122</td>
<td>0</td>
<td>100</td>
<td>Metalloproteinase?</td>
</tr>
<tr>
<td>121</td>
<td>127</td>
<td>57 (47)</td>
<td>57</td>
<td>Cl repressor</td>
</tr>
<tr>
<td>121 (1–117)</td>
<td>127</td>
<td>57 (47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>121 (118–246)</td>
<td></td>
<td>58 (72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>n.a.</td>
<td>n.a.</td>
<td>40</td>
<td>Cro repressor</td>
</tr>
<tr>
<td>238</td>
<td>287</td>
<td>55 (23)</td>
<td>82</td>
<td>Anti repressor</td>
</tr>
<tr>
<td>238 (1–105)</td>
<td>287</td>
<td>54 (51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>238 (105–238)</td>
<td></td>
<td>1 (0.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>93</td>
<td>1 (1.2)</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>87</td>
<td>0</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Note. The first two columns align the corresponding orfs from the two phages. In cases where a corresponding orf could be attributed, the third column provides the number of aa changes (percent) between the corresponding gene products and the fourth column provides the degree of nt conservation in percent. Cases in which no corresponding genes could be identified between the phages are labeled n.a., nonapplicable.
shows lysogenic conversion of its host, changing its phenotype from rapidly sedimenting to homogeneously growing (Neve et al., 1998).

The two phages share about 10 kb of DNA information (roughly one-fourth of their genome size) clustered in two different genome regions (Brussow and Bruttin, 1995). As both phages are temperate phages, we suspected that they shared the lysogeny module. In fact, sequence analysis revealed similarly organized lysogeny modules. This allowed us to ask whether the whole lysogeny module was the unit of genetic exchange between both phages, as postulated in the initial formulation of Botstein's theory. The answer depends on the definition of the terms module and genetic exchange. Notably, the extent of detectable sequence similarity between the two distinct temperate phages, as assessed by the global alignment, coincided with the length definition of the lysogeny module from \$Sfi21\$ obtained by comparison with closely related virulent phages (Bruttin et al., 1997a; Desiere et al., 1997). This match between two independent data sets indicates that there are functional constrains to conserve the whole lysogeny module in temperate phages. However, the multiple sharp transition zones from highly similar to moderately similar or different DNA sequences within the two compared temperate phages suggest multiple exchange events. Currently we are unable to state whether these DNA exchanges were recent events (DNA recombination in an ecological context) or past events reflecting the evolution of \$S.\ thermophilus\$ phages in general.

As demonstrated by our sequence comparison, a single gene might be a unit of genetic exchange, e.g., orf 111 or orf 142. A typical lipoprotein box was present in orf 142 gp, thus proposing it as a surface-exposed lipoprotein (Fig. 7). The proximity of orf 142 to the int gene might suggest the involvement of the orf 142 gp in the phenomenon of lysogenic conversion observed with \$\phi TP-J34\$ (Neve et al., 1998). Three phage-encoded lipoproteins have been described so far to be involved in lysogenic conversion: phage \$\lambda\$ Bor protein (Barondess and Beck-
TP-J34 AND Sfi21 LYSOGENY MODULES

FIG. 8—Continued

with, 1990) and the Cor proteins of phages φ80 and N5 (Vostrov et al., 1996). Other phage-encoded lipoproteins are Lip of phage T5 involved in lytic conversion (Decker et al., 1994) and R71 of phage λ involved in lysis (Kedzierska et al., 1996). Furthermore, our analysis reveals that an exchange reaction can occur even within a gene, as demonstrated by transition points within the repressor and anti-repressor genes. In these two cases it appears that the exchange points separate gene segments coding for distinct domains of the respective proteins. The distinct, although related, N-terminal parts of the two Cl-like repressor proteins from the two streptococcal phages are similar to the DNA-binding regions in well-defined repressors, while the nearly identical C-terminal parts of the two repressors could represent the hypothetical proteolytic cleavage or the dimerization regions. The DNA regions between the putative cl- and cro-like genes from the two phages are likely to represent a genetic switch region (Fig. 5). Interestingly, these regions differ between both phages not only at the nucleotide sequence level, but also in the topological organization of DNA repeats.
Comparison of the φP1 proteins and φHK22 ROI protein with the predicted φSfi21 orf 287 gp and φTP-J34 orf 238 gp also suggested two apparent domains in these proteins. One domain, consisting of the C-terminal half of the orf 238 gp, is shared between coliphage and streptococcal phage proteins and is nearly identical between φSfi21 and φTP-J34. This domain contains a DNA-binding motif in the coliphage proteins. In contrast, a second domain, located in the N-terminal half of the orf 238 gp, where the similarity to the anonymous H. influenzae (>Sfi21 and <TVTP-J34. This domain contains a DNA-binding motif in the coliphage proteins. Interestingly, the C-terminal part of φSfi21 orf 287 gp is dispensable for the lytic and lysogenic life cycle, as demonstrated by the phage deletion mutant D1 (Bruttin and Brussow, 1996).

The proposal that a single gene or even gene fragments coding for a single protein domain and not a set of functionally related genes are the units of modular exchange is not entirely new. Comparative sequencing (Moore et al., 1981) and heteroduplex analysis (Highton et al., 1990) in lambdoid phages had also identified transitions from homologous to heterologous DNA sequences which were located within a gene. While this might be an exception in the exchange of genetic material between lambdoid phages, such exchange reactions appear to occur with higher frequency in S. thermophilus phages. If this concept is confirmed, it might be helpful to use the term phage "module" to specify a set of functionally related phage genes and another word, like phage "recombinon," to specify the phage DNA segments undergoing exchange reactions. The latter term will of course only be useful if similar DNA exchange units are identified when more than two related phages will be compared.

The term "recombinon" does not necessarily imply that site-directed recombinases are at work. The exchange process could first be random, with an ensuing selection under functional constraints for viable constructs (Casjens et al., 1992). Screening the borders of the lysogeny module and the internal transition zones between the two temperate phages did not reveal a consensus DNA sequence that would provide a hint of a specific recombinase. However, one transition zone overlapped a probable secondary attachment site defined by an analysis of the prophage Sfi21 integration mechanism and a spontaneous phage deletion mutant (Bruttin et al., 1997c). These deletion mutants were repetitively obtained upon serial passage of the wild-type φSfi21 in the laboratory (Bruttin and Brussow, 1996; Bruttin et al., 1997b,c). Further work on the integration system of φSfi21 revealed that this GTTG sequence which flanked the spontaneous deletion is directly adjacent to the core region of the attP site (Bruttin et al., 1997c). Notably, site-specific recombinase systems have been proposed as motors for module exchanges (Casjens et al., 1992) and in the case of the repressor genes of the two streptococcal phages the phage integrase could have mediated the module exchange. The presence of identical integrases and attP sites in both phages could facilitate this type of DNA exchange reaction in cells susceptible to both phages.

The transition zone in the anti-repressor gene (Fig. 8E) was near another site of DNA deletion repetitively observed in spontaneous phage mutants. The deletion site was followed by a GGC. The GGC sequence is found at the 5' end of the common core and partly overlaps the GTTG site mentioned above (GTT/GGCAGGGGACA...; the common core is to the right of the slash). The GGC sequence was also a DNA repeat that flanked a deletion observed in cloned phage DNA (Bruttin et al., 1997c). Interestingly, one further deletion type in cloned phage DNA was flanked by an AGGG repeat, one AGGG was in the attP region (directly adjacent to the GTTG and the GGC sequences discussed above), while the other AGGG was near the 5' end of the immunity gene (orf 203) (Bruttin et al., 1997c). Notably, the AGGG sequence in orf 203 is 67 bp to the left of the transition zone between orf 203 and 122 (Fig. 8C).

In summary, a rather complex pattern of evolutionary DNA exchanges was observed when we compared the lysogeny module from two temperate S. thermophilus phages. Sequence information from further temperate S. thermophilus phages will be necessary to assess whether our observations can be generalized. At present, one should avoid stretching conclusions from two almost arbitrarily chosen phages. There might be S. thermophilus phages with still further arrangements of the lysogeny module which will result in even more complex patterns of evolutionary DNA exchanges. It will also be important to extend the sequence analysis to other phage genome regions than the lysogeny module (Desiere et al., 1998; Lucchini et al., manuscript in preparation).

MATERIAL AND METHODS

Bacteria, phages, and plasmids

The lysogenic S. thermophilus strain J34 was grown at 40°C in modified M17 broth (Krusch et al., 1987). Escherichia coli strains were grown at 37°C with constant shaking in LB broth (Sambrook et al., 1989) containing the appropriate antibiotics. For isolation of phage TP-J34, a 250-ml culture of S. thermophilus J34 grown to the early logarithmic phase was induced by UV light induction. For this, the cells were suspended in 0.1 M MgSO4 and pumped through a quartz tube (internal diameter, 1.3 mm; length, 75 cm) placed under a laboratory UV lamp. Purification and concentration of the phage particles by NaCl and polyethylene glycol 6000 precipitation and subsequent twofold CsCl banding were done essentially as described earlier (Neve et al., 1989).
DNA techniques

DNA manipulations (i.e., restriction enzyme digestion, ligations, electrophoresis) were done according to Sambrook et al. (1989). Sau3A-generated fragments of the phage TP-J34 DNA were cloned in the lactococcal cloning vectors pGKV210 and pGKV110, respectively (van der Vossen et al., 1985). Furthermore, phage DNA fragments obtained from HindIII digests were cloned in pACYC177 (New England Biolabs) or in pGEM-7Z(f+) (Promega). Transformations were routinely done in E. coli C600 or XL1-Blue (both obtained from Stratagene) according to the method of Hanahan (1985).

Nucleotide sequencing

Cloned DNA was prepared from 5- or 400-ml overnight cultures and purified using either the QIAprep-spin or the QIAGen plasmid maxi kit (both from Qiagen) and was sequenced by the Sanger method with the Sequencing Kit (USB) and the Sequenase (USB) or in pGEM7 (+) (Promega). The nucleotide sequence of the phage DNA was determined from cloned phage DNA. Sequences were obtained using the T7 gene 6 exonuclease (USB) or in pGEM-7Z(f+) (Promega). The nucleotide sequence of the phage DNA was confirmed by the method of Hanahan (1985).

ACKNOWLEDGMENTS

The authors are grateful to Jonathan Brown for his help in the initial sequencing experiments and to Angela Beck for her technical assistance. Work on S. thermophilus phage TP-J34 was supported by the European Union BIOTEC program (Contract B102-CT94-3055). The authors thank the Swiss National Science Foundation for financial support of Frank Desiere and for the Swiss National Bank for financial support of Frank Desiere in the framework of its Biotechnology Module (Grant 5002-04454/1).

REFERENCES


A Highly Conserved DNA Replication Module from *Streptococcus thermophilus* Phages is Similar in Sequence and Topology to a Module from *Lactococcus lactis* Phages

Published in *Virology*. 1997 Aug 4;234(2):372-82.

Desiere F, Lucchini S, Bruttin A, Zwahlen MC, Brüssow H
A Highly Conserved DNA Replication Module from *Streptococcus thermophilus* Phages Is Similar in Sequence and Topology to a Module from *Lactococcus lactis* Phages

Frank Desiere, Sacha Lucchini, Anne Bruttin, Marie-Camille Zwahlen, and Harald Brüssow

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

Received May 12, 1997, returned to author for revision May 14, 1997: accepted May 27, 1997

A highly conserved DNA region extending over 5 kb was observed in *Streptococcus thermophilus* bacteriophages. Comparative sequencing of one temperate and 26 virulent phages demonstrated in the most extreme case an 18% aa difference for a predicted protein, while the majority of the phages showed fewer, if any aa changes. The relative degree of aa conservation was not homogeneous over the DNA segment investigated. Sequence analysis of the conserved segment revealed genes possibly involved in DNA transactions. Three predicted proteins (orf 233, 443, and 382 gene product (gp)) showed nucleoside triphosphate binding motifs. Orf 443 gp showed in addition a DEAH box motif, characteristically found in a subgroup of helicases, and a variant zinc finger motif known from a phage T7 helicase/primase. Tree analysis classified orf 443 gp as a distant member of the helicase superfamily. Orf 382 gp showed similarity to putative plasmid DNA primases. Downstream of orf 382 a noncoding repeat region was identified that showed similarity to a putative minus origin from a cryptic *S. thermophilus* plasmid. Four predicted proteins showed not only high degrees of aa identity (34 to 63%) with proteins from *Lactococcus lactis* phages, but their genes showed a similar topological organization. We interpret this as evidence for a horizontal gene transfer event between phages of the two bacterial genera in the distant past. © 1997 Academic Press

**INTRODUCTION**

Analysis of virus evolution by comparative DNA sequencing of selected genes has become common place in contemporary virus research. These analyses have led to phylogenetic trees for specific virus genes which has allowed some conclusions to be drawn with respect to the evolution of specific virus groups as a whole. Virologists have realized that certain ideas on virus evolution established for animal viruses do not apply to bacterial viruses. During the seventies a great deal of evidence was accumulated suggesting that lambdoid phages are related in ways not easily accounted for by the standard ideas of evolution along branching trees of linear descent (reviewed in Casjens et al., 1992). On the basis of these data, Botstein (1980) has formulated a modular theory of bacteriophage evolution. The essentials of this theory were stated as follows: the product of evolution is not a given virus but a family of interchangeable genetic elements (modules) each of which carries out a particular biological function; exchange of a given module for another occurs by recombination among viruses belonging to the same interbreeding population and finally, these viruses and their hosts can differ widely in many characteristics. According to this theory evolution acts primarily at the level of the modules and not at the level of the intact virus. This theory turned out to be extremely helpful for understanding the genome organization of bacteriophages and has been accepted as a standard hypothesis for bacteriophage evolution in general. Our laboratory is interested in bacteriophages of *Streptococcus thermophilus*, a gram-positive bacterium used in milk fermentation (lactic acid bacterium (Mercenier, 1990)). Phage infection has always been a major problem in industrial fermentation, especially in the dairy industry (Peitersen, 1991). Contamination by virulent phages may result in the lysis of bacterial starter strains, causing slow fermentation or even complete starter failure with consequent loss of the product. *S. thermophilus* phage control measures have been hampered by the large variability of phages encountered in the environment (Benbadis et al., 1990; Brüssow et al, 1994a; Neve et al, 1989; Prevots et al, 1989) and the lack of knowledge of their genome organization. To remedy this situation we started the molecular characterization of our prototype phage strain, the temperate phage φ*Sfi21* (Brüssow and Bruttin, 1995; Bruttin and Brüssow, 1996). In contrast to the lysogeny module which was specific for φ*Sfi21* (Bruttin et al, 1997), the rest of the genome of this temperate phage showed extensive cross-hybridization with DNA from virulent *S. thermophilus* phages (Brüssow et al, 1994b; Brüssow and Bruttin, 1995).

In the present communication we report on the highly conserved DNA module of the temperate phage φ*Sfi21* located directly adjacent to the lysogeny module. Analysis of the DNA sequence identified several open reading frames...
frames (orf) coding for enzymes possibly involved in the initiation of DNA replication. Surprisingly, this DNA module was not only conserved between the temperate \( \phi Sfi21 \) and most virulent \( S. \) thermophilus phages, but it also showed sequence similarity and a similar gene order to genes from \( Lactococcus lactis \) phages (Boyce et al., 1995; Kim and Batt, 1991; Kim et al., 1992).

**MATERIALS AND METHODS**

**Phages, strains, and media**

Lytic phages from our collection (Brüssow et al., 1994a) were used for comparative Southern blot hybridization. Phages (\( \phi \)) \( \phi Sfi31 \), \( \phi ST44A \), \( \phi ST130 \), and \( \phi ST12 \) were isolated from a survey of a single yoghurt factory in France during the years 1993 to 1994 (H. Brüssow et al., unpublished results). \( \phi Sfi18 \) is an isolate recovered in 1987 from a different French yoghurt factory, while \( \phi ST128 \) is from a French starter culture supplier (generous gift from A. Zourari, Paris). Phages \( \phi H \), \( \phi L \), \( \phi B \), \( \phi I \), \( \phi F \), \( \phi E \), \( \phi J \), \( \phi C \), and \( \phi M \) were isolated from a survey of a single mozzarella cheese factory in Italy during the years 1993 to 1996 (Bruttin et al., 1997b). Phages \( \phi S55 \), \( \phi S19 \), and \( \phi S69 \) are Italian cheese factory isolates from a different geographical area and a different time period (before 1978), while \( \phi S94 \) and \( \phi S96 \) are Italian cheese factory isolates from a starter culture supplier after 1990 (generous gift from T. Sozzi, Milan). \( \phi L-A3 \), \( \phi L-A4 \), \( \phi L-A9 \), \( \phi L-A12 \), and \( \phi L-F1 \) are Swiss cheese factory isolates provided by the Federal Dairy Research Institute (generous gift from U. Zehntner, Bern).

The phages were propagated on their appropriate \( S. \) thermophilus hosts in lactose M17 broth as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995). *Escherichia coli* strain JM101 (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.

**DNA techniques**

Phage purification, DNA extraction, agarose gel electrophoresis, Southern blot hybridization, and labeling of plasmid DNA were done as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995). *Escherichia coli* strain JM101 (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.

**Cloning of the conserved DNA module**

\( \phi Sfi21 \) DNA was cut with restriction enzymes EcoRI, *HindIII*, *NsiI*, *XbaI*, or *Sau3A* and cloned into pUC19 vector or the E. coli/lactococcal/streptococcal shuttle vector pNZ124 (Platteau et al., 1994). Previously, we had identified a 2.2-kb EcoRI \( \phi Sfi21 \) DNA fragment (A in Fig. 1) that was conserved in most \( S. \) thermophilus phages from our collection (Brüssow et al., 1994a). By sequence analysis a unique *HindIII* site was located in this conserved fragment (Brüssow et al., 1994b). Using this conserved DNA fragment as a hybridization probe, we identified two clones, B and D, containing a 2.2- and 3.6-kb *HindIII* \( \phi Sfi21 \) DNA fragment, respectively (Fig. 1). The insert of clone B is in turn hybridized with clone C containing a 3.2-kb EcoRI \( \phi Sfi21 \) DNA fragment. Clone C overlapped a previously characterized clone pPX6 (Bruttin and Brüssow, 1996), containing the rightmost part of the lysogenic module of \( \phi Sfi21 \) (Fig. 1). The 3.6-kb *HindIII* fragment in clone D was subcloned using the internal EcoRI sites, creating the subclone E which contains a 0.8-kb insert (Fig. 1).

**Sequencing**

DNA sequencing was started with the universal forward and reverse primers of pUC19 or pNZ124 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 *T7* DNA Sequencing System of Promega (Madison, WI). The sequencing primers were end-labeled using \( \gamma \)\textsuperscript{32}PdATP according to the manufacturer's protocol. The thermal cycler (Perkin-Elmer) was programmed at 30 cycles of 95° for 30 sec, 50° for 30 sec, and 72° 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 prepare a template for comparative sequencing of five different regions of the conserved DNA module. PCR products were generated using the synthetic oligonucleotide pair designed according to the established \( \phi Sfi21 \) DNA sequence, purified phage DNA and Super Taq Polymerase (Stehelein, 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 99; EMBL
FIG. 1. Prediction of open reading frames (orfs) longer than 50 codons for the investigated DNA sequence. The length of the predicted proteins in amino acids was noted above the arrows; predicted proteins with identical length were differentiated with a or b annotations. Reading frame 1, 2, or 3 and location on the complementary strand (C) were indicated below the arrows. Orfs showing matches in the database searches were only lightly shaded. * Indicates standard ribosomal binding site. The upper line gives the position of some relevant restriction sites. The lower line gives the nucleotide position scale and the position of relevant clones mentioned in the text. The third intragenic region is shown as sequence information. Direct repeats are numbered R1 to R4 and are underlined by a broken line, an inverted repeat is underlined by an arrow (one mismatch). The region of nucleotide sequence similarity with the putative minus origin of replication from the cryptic S. thermophilus plasmid pST1 is marked by a bended arrow. The sequence is available from the GenBank Database under Accession No. AF004379.

RESULTS

Definition of the highly conserved phage DNA module

More than 70% of the phages from our collection, comprising now about 100 individual phage isolates, hybridized with a 2.2-kb EcoRI DNA fragment (Brussow et al., 1994b) from phage φSfi21 (fragment A in Fig. 1). We selected 26 lytic phages, which reacted with this probe on dot blots, for further investigation. These phages covered yogurt and cheese factory isolates from France, Italy, and Switzerland. To define the borders of this conserved phage DNA module, Southern blot hybridizations were done using cloned φSfi21 DNA fragments B to E (for an orientation of the fragments see Fig. 1) as probes against DNA from these virulent phages. The results of these hybridizations are summarized in Fig. 2. All but three of the selected 26 phages yielded a conserved 2.2-kb EcoRI hybridization signal with fragment A (Fig. 2). When fragment B was used on Southern blots, 18 phages yielded a conserved 2.2-kb HindIII hybridization signal. Nineteen phages showed a conserved 2.5-kb XbaI/EcoRI fragment when probed with fragment C, while no hybridization at all was observed when clone pPX6 was used as a probe. Therefore the transition point between the lysogeny module and the highly conserved module on φSfi21 DNA must be very near to the XbaI site. This was confirmed by sequence analysis of the conserved orf 157 which overlapped this site by 32 nucleotides (Fig. 1 and below). To explore the left border of the conserved module in the virulent phages, HindIII-digested phage DNA samples were probed with fragment C. For the 16 phages which gave the conserved 2.2-kb signal, corresponding to fragment B, a second signal of variable size was obtained corresponding to the left adjacent HindIII fragment. Based on size conservation the investigated phages yielded three subgroups comprising 6, 5, and 3 members, respectively (Fig. 2).

To explore the right border of the conserved module,
A HIGHLY CONSERVED DNA REPLICATION MODULE FROM *S. thermophilus*

### FIG 2 Partial restriction map of the DNA segment containing the highly conserved DNA replication module from temperate \(\phi Sfi21\) and 26 virulent *S. thermophilus* phages All *EcoRI* (E), HindIII (H), and *XbaI* (X) restriction sites that could be mapped are shown. The sites shown in bold for \(\phi Sfi21\) DNA are those depicted in Fig. 1 except for three narrowly spaced *EcoRI* sites that could not be resolved by Southern blot hybridization. *X*, the topological order of the sites was not established. The darkly shaded line gives the extent of cross-hybridization of the DNA from the phage strain indicated at the left side with \(\phi Sfi21\) DNA observed under high stringency conditions (90% DNA homology). If the darkly shaded line ends with an interruption (e.g., right side of \(\phi S55\)), the length of the cross-hybridizing DNA fragment was not established accurately.

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>(\phi Sfi21)</th>
<th>(\phi Sfi18)</th>
<th>(\phi L)</th>
<th>(\phi Ba55)</th>
<th>(\phi H)</th>
<th>(\phi F)</th>
<th>(\phi Sfi33)</th>
<th>(\phi ST44)</th>
<th>(\phi ST130)</th>
<th>(\phi Ba519)</th>
<th>(\phi ST3)</th>
<th>(\phi Ba569)</th>
<th>(\phi F)</th>
<th>(\phi ST128)</th>
<th>(\phi J)</th>
<th>(\phi Ba594)</th>
<th>(\phi Ba596)</th>
<th>(\phi ST12)</th>
<th>(\phi L-A3)</th>
<th>(\phi L-A4)</th>
<th>(\phi L-A9)</th>
<th>(\phi L-A12)</th>
<th>(\phi L-F1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td>X</td>
<td>H</td>
<td>E</td>
<td>H</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>H</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>H</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>H</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>XbaI</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>
<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>HindIII</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>
<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>

### Sequencing of the conserved DNA module of \(\phi Sfi21\)

DNA fragments A to E were sequenced on both strands and investigated for open reading frames (ORF) of 50 or more codons using ATG as the start. Thirteen ORFs potentially coding for proteins ranging from 51 to 443 amino acids (aa) were detected (Fig. 1). Only 3 ORFs were preceded by a standard *S. thermophilus* ribosomal binding site (GAG; Guedon et al., 1995) in appropriate spacing with respect to the start codon. Eight of the 13 predicted ORFs pointed rightward (Fig. 1) from which 5 predicted proteins showed similarity to known sequences (see below). In contrast, the coding potential of the leftward ORFs was small (range: 51 to 81 aa) and none showed similarity to predicted proteins from the databases. The genetic structure of the sequenced phage DNA was relatively dense; only two intergenic regions larger than 100 nt were observed.

**ORFs 157 and 233**

The predicted protein product from ORF 157 showed significant similarity (\(P = 10^{-5}\)) to an anonymous 169-aa protein from the lactococcal phage \(\phi BK5-T\) (Boyce et al., 1995). Alignment of both proteins revealed 34% identical aa. At the N- and the C-terminals, the identity rate was even higher (52 and 41%, respectively). The gp from the adjacent ORF 233 also showed similarity to a hypothetical protein from the same lactococcal phage (\(\phi BK5-T\) ORF 234 gp). The similarity between the two proteins was strong.
(P = \(10^{-67}\); 140 aa positions shared, 63% identity; Fig. 3). No function was attributed to these proteins, but the presence of the "Walker A" motif in both proteins (aa positions 21 to 28) indicates NTP (nucleoside triphosphate) binding activity (Walker et al., 1982). Interestingly, the streptococcal \(6\)Sfi21 orf 233 gp (gene product) showed higher similarity to the lactococcal \(6\)BK5-T orf 234 gp (140 aa shared) than the two lactococcal phages BK5-T and phi-31 to each other (78 aa shared; Fig. 3), ORF 443

Orf 443 gp showed strong similarity (39% identity; \(P = 10^{-109}\)) to a 452-aa protein from lactococcal phage 7-9 (Fig. 4). This lactococcal phage protein was reported to be extremely conserved in many lactococcal phages (Kim and Batt, 1991). Database searches identified in addition significant similarity (\(P = 10^{-12} \) to \(10^{-14}\)) to numerous helicases of various origin. A tree analysis of all proteins showing similarities with \(P \) values < 0.0003 yielded essentially the dendrogram described for the helicase superfamily by Koonin (1991) (Fig. 5). \(6\)Sfi21 orf 443 gp did not cluster with one of the well-known subfamilies (DEAD and DEAH box proteins and virally encoded helicases) even though orf 443 gp shows a DEAH box, but was found in a side branch together with the endonuclease (R) subunit of a type 1 restriction enzyme from \(E\). coli (hsdR gene; Loenen et al., 1987). Sequence alignment had previously identified HsdR as a peripheral member of the helicase superfamily (Gorbalenya and Koonin, 1991). Sequence alignments of helicases have revealed six to seven conserved motifs (Pause and Sonnenburg, 1992), most of which are located around the ATP-binding site (Subramanya et al., 1996). Two well-characterized motifs were identified in both orf 443 gp and HsdR: the DEAH box or motif II and the "Walker A" or motif I commonly found in NTP-binding proteins (Walker et al., 1982). Interestingly, motif I was found in two different regions of orf 443 gp (Fig. 4). The phase 7-9 homologue showed only a DEEx motif (Fig. 4). Multiple alignment of orf 443 gp with five related proteins led to the definition of a further four conserved aa blocks (Fig. 4). The conserved aa block A showed weak similarity with helicase motif Ia (the PTEALE motif; Pause and Sonnenberg, 1992), while the other conserved aa blocks B, C, and D showed no recognizable similarity with motifs III, IV, V, and VI from known helicases (Gorbalenya and Koonin, 1991; Schmid and Linder, 1992). Interestingly, block D was within a variant zinc-finger motif, C-X-X-C-X(10-15)-C-X-X-C, known from bacteriophage T7 gene 4 protein (Bernstein and Richardson, 1988).

ORF 124 and 271

The predicted \(6\)Sfi21 orf 124 gp showed 35% identity and 46% similarity with an anonymous 18-kDa predicted protein from lactococcal phage 7-9 (\(P = 10^{-29}\)). Interestingly the two proteins showed an identical topological position: both were directly downstream of a DEAH or DEEx protein, respectively (Kim et al., 1991). In addition, both proteins showed an unusual percentage of charged aa (e.g., 19 basic and 17 acidic aa for orf 124 gp from \(6\)Sfi21). The adjacent \(6\)Sfi21 orf 271 gp showed no similarity to any sequence from the databases.

ORF 382

The gene product of orf 382 showed significant similarity to putative DNA primases from the cryptic plasmid pWS5 of \(Lactobacillus delbrueckii\) subsp. lactis (\(P = 10^{-29}\)) and from plasmid pRN1 of \(Sulfolobus islandicus\).
A HIGHLY CONSERVED DNA REPLICATION MODULE FROM S. thermophilus

377

FIG. 4. Sequence alignment of the deduced protein product of Sfi21 orf 443 with the lactococcal phages \( \phi 7-9 \) orf 452 gene product. Identical amino acids are indicated by a colon; similar amino acids are indicated with a dot. The helicase motif I (modified Walker A motif; Pause and Soncnberg, 1992) and II (DEAH box) are indicated. The position of a further four conserved aa blocks A, B, C, and D, deduced from a comparison with five related proteins (marked in Fig. 5 by an asterisk), are indicated (asterisk, conserved in all 6 proteins; dot, conserved in 3–4 proteins). The position of a variant zinc finger motif is given.

\( \text{P} = 10^{-11} \) (Keeling et al., 1996), and to an anonymous ATP-binding protein from chlorella virus 1 (P = 10^{-10}) (Kutish et al., 1996). Clustal W multiple sequence alignment (Fig. 6) resulted in 19 aa common to all four proteins, and 119, 75, and 51 identical aa between the orf 382 gene product and the proteins coded by the Lactobacillus and Sulfolobus plasmids and Chlorella virus, respectively (31, 20, and 13% identity). One region corresponded to a type I nucleotide binding site (Fig. 6). A weak homology (43 identical aa over 203 aa length; \( \text{P} = 0.004 \)) was detected with the multifunctional DNA primase of phage P4 (Strack et al., 1992; Ziegelin et al., 1993). Notably, downstream of Sfi21 orf 338 a 218-nt noncoding DNA region showing conspicuous DNA repeats, four direct and one inverted, was observed (Fig. 1). Using the FastA/Bestfit program we found an 80% sequence identity over 220 nt (\( \text{P} = 10^{-22} \)) between this AT-rich intergenic region and a region following the repA gene of the cryptic S. thermophilus plasmid pST1 (Lanzen et al., 1992). When alternative initiation codons were allowed, orf 382 experienced an N-terminal extension resulting in an orf 500 which filled most of the intergenic region between orf 271 and orf 382 (\( \text{P} = 10^{-40} \) with the predicted pWS58 protein).

Sequence variation in the conserved module among the 26 virulent phages

Five distinct regions (see Fig. 7 for an orientation) of the highly conserved module (orf 157, two segments from orf 443, orf 124, orf 271, and orf 382) were partially sequenced in the 26 virulent phages and compared with the corresponding Sfi21 sequence. A maximum of 40 nt changes was observed over 252 nt (corresponding to 15 aa changes over 84 aa) (Table 1). When median values were calculated for each region excluding orf 382 the
highest values obtained were of 2 nt and 1 aa changes. Orf 382 showed a median value of 18 nt changes most of which were silent. In contrast when the most related lactococcal phage proteins were included in the comparison the number of aa differences increased dramatically with a minimum of 47 changes over a stretch of 83 amino acids (Table 1).

When compared to the φSI21 DNA sequence, the virulent S. thermophilus phages showed no constant rate of sequence difference over the conserved module. Of the 26 virulent phages, φI was the most divergent phage with respect to orf 157 and orf 382, while it showed no or only two nt changes for the 3′-end of orf 443 and orf 124, respectively (Table 1). Conversely, φL-A9 was the most divergent phage for orf 443 and orf 124, whereas it showed no nt changes in orf 157.

One relatively divergent phage, φS94, was sequenced across the whole conserved module and compared with φSI21. In a similarity plot, DNA sequence differences were unevenly distributed: relative minima of aa conservation were observed over the left 2-kb (orf 233 and orf 443) and the right 1-kb DNA segment (orf 382) of the conserved module, while a maximum was seen in the central 2-kb DNA segment (comprising orf 124 and orf 271 and including the two intergenic regions; Fig. 7). Comparable differences were seen at the aa level (φS94 orf 157, 2.7%; orf 233, 3.9%; orf 443, 5.9%; orf 124, 0.8%; orf 271, 1.8%; and orf 382, 3% aa difference with respect to the φSI21 homologues). There were still other parameters that changed over the conserved DNA module; e.g., the ratio of changes in the third position of the triplet to the first and second positions. This ratio was, for example, higher in the orf 271 comparison than in the orf 382 comparison (7.9 vs 1.8), possibly indicating a decoupling of the evolutionary histories of these two orfs.

**DISCUSSION**

Database searches with the conserved DNA segment identified genes coding for enzymes possibly involved in the initiation of DNA synthesis. Three orfs coded for proteins that showed a "Walker A" motif typically found in nucleoside triphosphate binding proteins (orf 233, 443, and 382). One of them, orf 443 gp is a peripheral member of the helicase superfamily, The HsdR protein, another peripheral member of the helicase superfamily (Gorbalenya and Koorn, 1991) showed as orf 443 gp motifs for NTP binding, DNA binding, and the DEAH box (Loenen et al., 1987), Bacteriophage T7 gene 4 protein, which combines DNA binding with primase and helicase activity (Bernstein and Richardson, 1988), shows a similar motif composition (variant zinc finger, two NTP binding
sites, a DEA box), but no sequence similarity with orf 413 gp. Orf 382 gp from φS321 showed significant similarity to DNA primases from two plasmids and weak similarity to a multifunctional protein from phage P4, which combined DNA binding, helicase, and primase activity (Ziegelin et al., 1993). In support of this primase attribution, a transfer of DNA segments between phages of different hosts is predicted by the modular theory. However, a gene transfer between phages which do not constitute an interspecific complex. This argument would implicate the intervening orfs 124 and 271 in replicative functions. In fact, antisense messenger RNA to the lactococcal phage homologue of orf 124 inhibited the replication of lactococcal phages carrying this sequence (Kim et al., 1992).

The high similarity of predicted streptococcal phage proteins to proteins from lactococcal phages and the comparable topological organization of their genes was a startling finding. A transfer of DNA segments between different phages from the same streptococcal host is predicted by the modular theory. However, a gene transfer between phages which do not constitute an interbreeding phage population and which attack hosts belonging to different bacterial genera (Wood and Holzapfel, 1992) was unexpected. Over a stretch of about 14 kb intervening orfs 124 and 271 in replicative functions. In fact, antisense messenger RNA to the lactococcal phage homologue of orf 124 inhibited the replication of lactococcal phages carrying this sequence (Kim et al., 1992).

The high similarity of predicted streptococcal phage proteins to proteins from lactococcal phages and the comparable topological organization of their genes was a startling finding. A transfer of DNA segments between different phages from the same streptococcal host is predicted by the modular theory. However, a gene transfer between phages which do not constitute an interbreeding phage population and which attack hosts belonging to different bacterial genera (Wood and Holzapfel, 1992) was unexpected. Over a stretch of about 14 kb

![FIG. 6. Alignment of the deduced protein product of φS321 orf 382 and putative primases from the cryptic plasmid pWS58 of Lactobacillus delbrueckii subsp. lactis (Accession No. Z50864) and from plasmid pRN1 of Sulfitobacter islandicus (Keeling et al., 1996) and orf A456L from Chlorella virus 1 (Kutish et al., 1996). Amino acid positions conserved in all four sequences were marked by an asterisk; those only conserved in three are marked with an open circle. The variant Walker A motif found in adenylate kinase (Walker et al., 1982) is boxed.](image-url)
thermophilus and L. lactis (Heller et al., 1995). In addition, comparative sequencing of insertion sequences (IS) in S. thermophilus and L. lactis (Guédon et al., 1995; Bourgoin et al., 1996) revealed quasi-identity of their nt sequences (>99%, while their 23S rRNA shares only 89.2% identity), suggesting that intergeneric transfers actually occurred during cocultures used in the manufacture of cheese. However, none of the analyzed genes from lactococcal and streptococcal phages showed such a high sequence conservation indicative of a recent horizontal gene transfer. The closest relatives detected until now (φSFi21 orf 233 and φBK-5T orf 234) still differed in 77 of 233 aa. This divergence pushes the hypothetical gene transfer event to a more distant past.

Alternatively, one might speculate that Streptococcus and Lactococcus phages derived their conserved module from a common precursor module which evolved differently in the two bacterial genera. However, no conceptual framework exists for coevolution of phages or phage modules with their bacterial hosts (Botstein, 1980; Bernstein and Bernstein, 1989; Casjens et al., 1992). If this would be the case one would expect a close relationship between S. thermophilus and S. pneumoniae phages (Martin et al., 1996), which has not been ob-

**TABLE 1**

<table>
<thead>
<tr>
<th>Gene</th>
<th>n°</th>
<th>nt (aa)*</th>
<th>nt change£</th>
<th>aa change£</th>
<th>Nearest homologue*</th>
<th>Most divergent phages†</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf 157</td>
<td>21</td>
<td>252 (84)</td>
<td>0 0 0</td>
<td>0 15 0</td>
<td>64</td>
<td>L, S94, M, E, ST130, SF153 (40-3)</td>
</tr>
<tr>
<td>orf 443.5*</td>
<td>23</td>
<td>243 (81)</td>
<td>0 34 0</td>
<td>0 9 0</td>
<td>53</td>
<td>L/A9, L/A3, L/A4, M, S94, S5 (34-20)</td>
</tr>
<tr>
<td>orf 443.3*</td>
<td>26</td>
<td>249 (83)</td>
<td>0 32 0</td>
<td>0 6 0</td>
<td>47</td>
<td>S5, L/A9, L/A3, L/A4, H (32-25)</td>
</tr>
<tr>
<td>orf 124</td>
<td>24</td>
<td>258 (86)</td>
<td>0 11 2</td>
<td>0 2 1</td>
<td>56</td>
<td>L/A9, LF1, L/S69, B, SF18 (11-4)</td>
</tr>
<tr>
<td>orf 271</td>
<td>26</td>
<td>267 (89)</td>
<td>0 5 2</td>
<td>0 3 1</td>
<td>†</td>
<td>S94, S96, M, L/A9 (5-2)</td>
</tr>
<tr>
<td>orf 382</td>
<td>24</td>
<td>282 (94)</td>
<td>0 29 18</td>
<td>0 3 0</td>
<td>54*</td>
<td>J, S96, E, ST128, M, S94 (29-23)</td>
</tr>
</tbody>
</table>

* Number of investigated phages.
* Number of nucleotides (amino acids) compared.
£ Number of nucleotide changes range: median in comparison with φSFi21.
£ Number of amino acid changes range: median in comparison with φSFi21.
£ Number of amino acid changes between φSFi21 and the nearest lactococcal phage.
† The most divergent phage strains in order of decreasing diversity (range of nt changes in this subgroup).
* 5' and 3' part of orf 443, respectively.
* From a lactobacillus plasmid, no lactococcal phage available for comparison.
A HIGHLY CONSERVED DNA REPLICATION MODULE FROM \textit{S. thermophilus}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{alignment.png}
\caption{Alignment of the predicted ORFs in the streptococcal phage \phiSfi21 genome (middle panel) with that of the lactococcal phages \phiBK5-T (lower panel), \phiT-9 and \phiLC-3, a Lactobacillus plasmid and \textit{E. coli} phages (all upper panel). The ORFs are named according to their length in amino acids. The relatedness between the predicted proteins is indicated as percentage of aa identity and similarity. Related proteins are connected by grey shading.}
\end{figure}

served. In addition, the observation of \phiSfi21 genes showing significant similarities to genes from \textit{Leuconostoc} oenos, \textit{Lactobacillus delbrueckii}, \textit{Lactobacillus casei}, and \textit{Mycobacterium} spec. phages (Desiere et al., manuscript in preparation) excludes evolution of the \phiSfi21 genome along lines of linear descent. Furthermore, the streptococcal phage genome seems to be remodelled by recombination events as seen by interspersed coliphage-like DNA elements between the lactococcal-like DNA (Fig. 8). Sequence analysis (variation in the degree of sequence conservation, ratio of nt changes at different codon positions) seems to indicate recombination events even within the conserved DNA module. The \phiSfi21 genome could thus reflect a number of horizontal gene transfer events mainly, but not exclusively, with phages from lactic acid bacteria, in a distant past, followed by some evolution and recombination within their current host in the more recent past. It is startling that most \textit{S. thermophilus} phages have faithfully conserved the gene constellation of the investigated module, while only part of the lactococcal phages adopted this gene constellation and those that acquired it showed much less conservation of the genes (e.g., the orf 233 homologues from lactococcal phages BK-5T and phi-31 showed 66% aa diversity). This difference could reflect the evolutionary forces leading to many distinct DNA homology groups in lactococcal phages (Jarvis, 1984), while no such splitting has been seen with \textit{S. thermophilus} phages which all belong to one DNA homology group (Brüssow et al., 1994a).

Analysis of further conserved DNA segments in other genome regions of \textit{S. thermophilus} phages (Brüssow and Bruttin, 1995) will be necessary to further constrain hypotheses on the "evolution" of this phage group.

ACKNOWLEDGMENTS

We thank Thomas Janzen (Chr. Hansen Laboratory) for contributions in the initial phase of the work and S. Foley for helpful discussions. We thank the Swiss National Science Foundation for financial support of Frank Desiere and Paolo Lucchini in the framework of its Biotechnology Module (Grant 5002-044545/1).

REFERENCES


Evolution of *Streptococcus thermophilus* Bacteriophage Genomes by Modular Exchanges Followed by Point Mutations and Small Deletions and Insertions


Desiere F, Lucchini S, Brüssow H
Evolution of *Streptococcus thermophilus* Bacteriophage Genomes by Modular Exchanges Followed by Point Mutations and Small Deletions and Insertions

Frank Desiere, Sacha Lucchini, and Harald Brüssow

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

Received October 16, 1997; returned to author for revision October 23, 1997; accepted November 19, 1997

Comparative sequence analysis of 40% of the genomes from two prototype *Streptococcus thermophilus* bacteriophages (lytic group I phage ϕSfi19 and the cos site containing temperate phage ϕSfi21) suggested two processes in the evolution of their genomes. In a first evolutionarily distant phase the basic genome structure was apparently constituted by modular exchanges. Over the 17-kb-long DNA segment analyzed in the present report, we observed clusters of genes with similarity to genes from *Leuconostoc oenos* phage L10, *Lactococcus lactis* phage BK5-T, and *Streptococcus pneumoniae* phage Dp-1. A chimeric protein was predicted for orf 1291 which showed similarity to both phage BK5-T and phage Dp-1 proteins. The very large orf 1626 gene product showed similarity to two adjacent genes from the *Lactobacillus delbrueckii* phage LL-H and further phage proteins (*Lactococcus lactis*, *Bacillus subtilis*). The similarities were localized to distinct parts of this apparently multifunctional protein. The putative ϕSfi19 lysin showed similarity to both lysins of phages and cellular enzymes. In a second, evolutionarily more recent, phase the *S. thermophilus* phage genomes apparently diversified by point mutations and small deletions/insertions. Over the investigated 17-kb DNA region ϕSfi19 differed from ϕSfi21 by 10% base pair changes, the majority of which were point mutations (mainly at the third codon position), while a third of the base pair differences were contributed by small deletions/insertions. The base pair changes were unevenly distributed: Over the *Leuconostoc* phage-related DNA the change rate was high, while over the *Lactobacillus* and *S. pneumoniae* phage-related DNA the change rate was low. We speculate that the degree of base pair change could provide relative time scales for the modular exchange reactions observed in *S. thermophilus* phages.

INTRODUCTION

Virologists realized nearly 20 years ago that certain ideas about virus evolution established for animal viruses do not apply to bacterial viruses (Casjens et al., 1992). For example, no lines of linear descent based on gradual accumulation of point mutations have been established for bacterial viruses. Experimental data from lambdoid phages led to the formulation of the modular theory of phage evolution by Botstein (1980). According to that theory the product of phage evolution is not a given virus, but a family of interchangeable genetic elements (modules) each of which carries out a particular function; exchange of a given module for another occurs by recombination among viruses belonging to the same interbreeding population; and finally, these viruses can differ widely in many characteristics. This theory turned out to be extremely helpful for understanding the genome organization of other bacteriophages, e.g., T4 phages (Monod et al., 1997) and has subsequently been accepted as a standard hypothesis for bacteriophage evolution in general.

Our laboratory is interested in phages attacking *Streptococcus thermophilus*, a Gram-positive lactic acid bacterium used as a starter in industrial milk fermentation (Mercenier, 1990). Several laboratories have classified *S. thermophilus* phages systematically collected from milk fermentation failures. This led to the characterization of two to four subgroups of phages (Brüssow et al., 1994a; Le Marrec et al., 1997; Prévôts et al., 1989), but all authors found DNA cross-hybridization between phages of all subgroups, *S. thermophilus* phages represent thus a relatively homogeneous phage group. An ecological survey of *S. thermophilus* phages in their natural environment confirmed the impression of a viral quasispecies (Bruttin et al., 1997a). Systematic dot blot hybridization experiments with defined restriction fragments from a type phage strain (ϕSfi21) against more than 30 distinct phage isolates suggested a modular organization of their genome structure (Brüssow and Bruttin, 1995). This observation was confirmed by comparative sequencing in the lysogeny module from two temperate *S. thermophilus* phages (Bruttin et al., 1997b; Newe et al., 1998). Areas of high sequence conservation were interspersed with regions of low or no sequence similarity, suggesting the involvement of recombination processes. In contrast with the original modular theory, four of the six transition zones from high to low sequence conservation were found within genes. The transition points appeared to separate gene segments coding for distinct functional
domains of proteins. Similar modifications of the original theory have been proposed by more recent reinvestigations of lambdoid phages (Highton et al., 1990). A different picture emerged from comparative sequencing of more than 20 distinct S. thermophilus phage isolates over the adjacent DNA replication module. Most phages differed from each other only by point mutations (Brüssow et al., 1994b; Desiere et al., 1997). Why should recombination processes contribute to the diversification of phages in one module, but not in another? Integrase-mediated recombination processes were suspected in the lysogeny module (Bruttin et al., 1997c; Neve et al., 1998); Recombination processes catalyzed by the integrase might be less likely outside of the lysogeny module. To further our understanding of the genome organization of S. thermophilus phages and to better define the mechanisms that lead to genetic diversity in this phage group we continued with the comparative sequencing approach over a previously uncharacterized 17-kb-long genome segment representing 40% of the genome size of S. thermophilus phages. For this comparison we chose as our prototype lytic group I phage $\phi$Sfi19 and the only temperate phage from our collection, $\phi$Sfi21. The sequence analysis confirmed the overall modular organization of the S. thermophilus phage genome, but the two phages differed from each other only by point mutations and small insertion/deletions. We propose a two-step evolution model to account for the diversity of S. thermophilus phages.

**RESULTS**

$\phi$Sfi19 and $\phi$Sfi21 are distinct phage strains

Bacteriophages $\phi$Sfi21 and $\phi$Sfi19 demonstrated a clearly distinct restriction pattern when digested with four restriction enzymes (EcoRV, HindIII, EcoRI, and PvuII, Brüssow and Bruttin, 1995; and data not shown). Comparison of heated and unheated restricted phage DNA suggested furthermore that $\phi$Sfi19 is, like $\phi$Sfi21, a cos site containing phage (data not shown).

Host range analysis of the two phages demonstrated different biological properties. The two phages showed a distinct, although overlapping host range when tested on 120 S. thermodilius strains from our collection (Table 1). $\phi$Sfi19 has a clearly broader host range than $\phi$Sfi21. However, infectivity of $\phi$Sfi19 could be neutralized with an antiserum to $\phi$Sfi21 and vice versa (Brüssow et al., 1994a), indicating a sharing of proteins essential for the infection process. Furthermore, $\phi$Sfi19 was unable to lysoygenize any of the hosts it could infect, while $\phi$Sfi21 could lysogenize all its host cells. Finally, $\phi$Sfi19 could grow lytically on cells harboring the $\phi$Sfi21 prophage, i.e., $\phi$Sfi19 is heteroimmune to $\phi$Sfi21 (Bruttin et al., 1997b).

<table>
<thead>
<tr>
<th>Strain</th>
<th>$\phi$Sfi21</th>
<th>$\phi$Sfi19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sfi1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sfi2</td>
<td>-</td>
<td>(+)</td>
</tr>
<tr>
<td>Sfi3</td>
<td>-</td>
<td>(+)</td>
</tr>
<tr>
<td>Sfi4</td>
<td>-</td>
<td>(+)</td>
</tr>
<tr>
<td>Sfi7</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sfi10</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sfi15</td>
<td>-</td>
<td>(+)</td>
</tr>
<tr>
<td>Sfi18</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sfi19</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sfi21</td>
<td>(+)</td>
<td>-</td>
</tr>
<tr>
<td>Sfi22</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sfi32</td>
<td>-</td>
<td>(+)</td>
</tr>
<tr>
<td>Sfi33</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sfi1c16</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>S3</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>W3</td>
<td>+</td>
<td>(+)</td>
</tr>
<tr>
<td>YS3</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: +, Replication; −, absence of replication (<10 PFU/mL) of the indicated phage on the given S. thermophilus strain. To control for effects of restriction modification systems, the phages were propagated on strain Sfi1 (+) Replication of the phage on the given cell after one subpassage on that cell.

**DNA sequence from $\phi$Sfi19**

A 17-kb-long DNA segment covering PvuII/XbaI (P/X) restriction fragments 2 (partly), 5, 4, 7, 8, and 3 (partly) from $\phi$Sfi21 (Fig. 1) was targeted for parallel sequencing of $\phi$Sfi21 and $\phi$Sfi19. Previous hybridization experiments between the two phages had demonstrated a transition from cross-hybridization (fragments 4, 7, 8, and 3) to nonhybridization (fragments 2 and 5) (Brüssow and Bruttin, 1995). When only ATG start codons were accepted, the 17-kb DNA segment from $\phi$Sfi19 contained 30 open reading frames (orf) longer than 60 aa. Two very long phage genes were observed (orf 1626 and orf 1291) together corresponding to roughly one-fourth of the total coding capacity of the S. thermophilus phages (≈40 kb, Brüssow et al., 1994a). With one exception, all orfs larger than 100 codons were located on the same strand (rightward orientation). None of the predicted proteins for the leftward-oriented orfs showed a match to proteins from the database. In addition, all leftward-oriented orfs were situated opposite to larger rightward orfs whose predicted proteins showed matches to entries from the database. These considerations reduced the likely number of genes in this genome area to 14 (Fig. 1). Orf 123 has an alternative start codon (GTG), but was included due to its good database match to gene I from Leuconostoc oenos $\Phi$ L10 (see below). All were rightward oriented and all but two showed database matches (see below for the individual predicted proteins). Ten orfs were preceded by a standard S. thermophilus ribosomal binding
FIG. 1 Prediction of open reading frames (orfs) in the 17-kb fragment of 4>Sfi19 and comparison with 4>Sfi21. The orfs were marked below the arrows by their length in aa; the reading frame is indicated above the arrow. An asterisk indicates standard ribosomal binding sites. Putative terminator structures are marked by a hairpin loop above the arrows. T indicates the termination of the nucleotide sequence in the database. PvuII and XbaI restriction sites and the restriction fragments corresponding to the published map of 4>Sfi21 (Brutin and Brussow, 1996) were noted together with the nucleotide scale at the bottom of the figure. Similarities to proteins from the database are indicated by shading. When the shading did not cover the whole 4>Sfi19 gene, only part of the gene showed similarity to the database entry. The boxes at the top of the figure indicate topologically similarly organized gene clusters in the indicated phages.

site (GAG; Guédon et al., 1995) in appropriate spacing with respect to the start codon. Orfs 203 and 289 were followed by an inverted repeat suggestive of a transcriptional terminator (ΔG = -86 and -70 kJ mol\(^{-1}\) at 37°C, respectively). Only three small intergenic regions were seen, the largest of which was 220 bp.

Comparison with 4>Sfi21

In 4>Sfi21 topological corresponding positions were occupied by orfs of nearly identical size (Fig. 1, Table 2). The similarity between the two phages extended to the nucleotide sequence level. No modular exchanges were observed. When aligned by the SIM program only 10 nonalignments were found, all of which represented deletions/insertions of base pairs in multiples of three (Table 2). Three orfs from 4>Sfi21 (orfs 1560, 1276, and 117b) were significantly smaller than their complements in 4>Sfi19. The differences were accounted by deletions or insertions (Figs. 2D, 3C, 4A, and 5).

A total of 1742 bp differences were detected over 17 kb aligned DNA (10% difference); 310 bp differences were accounted by deletion/insertions (Table 2). The majority of the changes were point mutations, mainly at the third base position. Three or more successive base pair changes (gaps) were rare: A total of 60 such regions were found, the longest being 11 bp long. The distribution of base pair differences was uneven. The DNA segment covering orfs 104 to 1626 in 4>Sfi19 showed a significantly higher number of base pair changes when compared to the corresponding DNA segment in 4>Sfi21 than the segment covering orf 515 to 289 (16 vs 2% bp changes). In the latter segment the base pair changes were clustered.

Similarity searches

The protein products predicted for the 14 4>Sfi19 orfs were compared to entries from the database. Scores higher than 100 and P values <10\(^{-6}\) were considered significant.

Five of the seven orfs located in restriction fragment PX2 (Fig. 1) predicted proteins that showed highly significant similarities with proteins from L. oenos bacteriophage L10 (Sutherland et al., 1994) (Table 3). The gene cluster from 4>L10 showed a topologically identical orientation to that of 4>Sfi19; however, orfs 116 and 117 showed no sequence similarity to the topologically corresponding 4>L10 genes H and F, respectively (Fig. 1).

Orf 1626 gp showed high similarities with predicted proteins from phages and phage-like elements from Gram-positive bacteria; however, weak similarities to nonphage proteins were also detected (Table 4). Several observations are notable: over the C-terminal part, matches with a number of different phage proteins led to the definition of a consensus sequence (Fig. 2B). Apparently this region defines a motif common to several phage systems. In contrast, the central part of the orf 1626 gp showed no database matches. Over this region even the corresponding orf 1560 gp from 4>Sfi21 differed substantially from the 4>Sfi19 protein (Figs. 2C and 2D). Furthermore, two adjacent genes from Lactobacillus orf L-H (Mikkonen and Alatossava, 1994) showed similarities with two different regions of the same orf 1626 gp (Fig. 1, Table 4), suggesting a combination of different functions in this very large streptococcal protein. Finally, the N-terminal third of the orf 1626 gp showed a hydrophobic region for which a coiled coil was predicted (Fig.
### TABLE 2

Comparison of the Indicated 4>Sfi19 Gene and Gene Product with the Corresponding Elements from 4>Sfi21 and Analysis of the Differences

<table>
<thead>
<tr>
<th>Orf</th>
<th>Base pair changes (%)</th>
<th>Change in codon position 1 + 2 (%)</th>
<th>Amino acid differences (%) in gp</th>
<th>Distribution of base pair changes</th>
<th>Gaps</th>
<th>Deletions/insertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>21</td>
<td>25</td>
<td>20</td>
<td>H</td>
<td>1</td>
<td>D:1</td>
</tr>
<tr>
<td>116</td>
<td>20</td>
<td>43</td>
<td>22</td>
<td>H</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>153</td>
<td>19</td>
<td>39</td>
<td>16</td>
<td>H</td>
<td>3</td>
<td>—</td>
</tr>
<tr>
<td>123</td>
<td>14</td>
<td>38</td>
<td>11</td>
<td>H</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>203</td>
<td>21</td>
<td>32</td>
<td>20</td>
<td>H</td>
<td>4</td>
<td>D:6</td>
</tr>
<tr>
<td>117</td>
<td>17</td>
<td>25</td>
<td>15</td>
<td>H</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>1626</td>
<td>16</td>
<td>37</td>
<td>19</td>
<td>H</td>
<td>4</td>
<td>3, 9, 21, 33, 141</td>
</tr>
<tr>
<td>515</td>
<td>5</td>
<td>36</td>
<td>4</td>
<td>C: 5'</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>1291</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>C: middle</td>
<td>2</td>
<td>D:45</td>
</tr>
<tr>
<td>670</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>H</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>131</td>
<td>12</td>
<td>27</td>
<td>21</td>
<td>5': H; middle</td>
<td>2</td>
<td>D: 42</td>
</tr>
<tr>
<td>141</td>
<td>0</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>87</td>
<td>0</td>
<td>—</td>
<td>0</td>
<td>—</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>289</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>C: 3'</td>
<td>0</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. Orf: the orfs were listed according to their order on the 4>Sfi19 genome (see Fig. 1). Base pair changes: the percentage of base pair changes between the corresponding orfs of the two phages. Change in the percentage of base pair changes found in the first and second codon position over the total base pair changes. Amino acid differences: the percentage of amino acid changes between the two proteins. Distribution: the distribution of the base pair changes in the compared orfs was classified as homogeneous (H) or clustered (C). The location of the clustered base pair changes in the 5' or 3' end or middle of the orf was given. Gaps: the number of ≥3 successive base pair changes (gaps) in the alignment of the two genes. Deletion: deletions (D) and insertions (I) observed in the alignment of the two corresponding orfs. The numbers give the length of the D/I in base pairs deduced from the SIM alignment. If more than one number is given, multiple D/I were observed. The distinction of D/I is arbitrarily based on the phage Sfi21 DNA sequence.

2A). This region showed similarity to the S-2 (hinge) region of myosin (Table 4) and ends with a leucine zipper (Fig. 2A). Over the same region the streptococcal protein showed also similarity with orf 31 from Lactococcus lactis phage BL67. This protein has been localized to the end of the phage tail by immunoelectron microscopy (Schouler et al., 1994). Since this protein showed an EF hand, a calcium binding domain, and since calcium is essential to many phages for adsorption, orf 31 is a reasonable candidate for the lactococcal phage anti-receptor protein. It should be noted that similarity between the two proteins ended just in front of the EF hand and that the leucine zipper is found at the exact place of the EF hand.

The next two streptococcal phage genes showed high similarity with two adjacent genes from the L. lactis phage BKS-T: orf 515 gp and orf 1291 gp showed similarity to orf 410 gp (score 231; P = 10⁻¹⁶) and very high similarity to orf 1904 from φBK5-T, respectively (Table 5). In the orf 1291 gp an approximately 200-aa-long region was apparently conserved in many lactococcal phages (Figs. 3B and 4B). Paradoxically, the DNA region coding for this conserved motif could be lost from the streptococcal phage without a measurable effect on the biological properties of the deletion mutant. This point was demonstrated by the spontaneous deletion mutant D3 from φSfi21 which occurred during serial passage of the wildtype phage (Bruttin and Brüssow, 1996). The deletion was flanked on both sides by a nearly perfect 53-nt repeat (Fig. 3C). The deletion mutant had lost the first repeat and the 658 bp located between the first and the second repeat. The deletion mutant had the same host range as the parental φSfi21, could establish lysogeny, and could grow lytically (Bruttin and Brüssow, 1996).

The similarity with the very large orf 1904 gp from φBK5-T extended with three interruptions up to aa position 1095 of orf 1291 gp (Fig. 3B), while from aa position 1095 to the C-terminus a high similarity with orf 1 from the S. pneumoniae phage Dp-1 was observed (Figs. 3B and 4C) (Sheehan et al., 1997).

Orf 1291 gp thus combined structural motifs from two proteins. With the lactococcal phage protein it shared collagen-like tripeptide GXY repeats (Figs. 3B and 4A). Contrary to the situation in the φBK5-T protein the four collagen motifs in the orf 1276 gp were distinct from each other, but a consensus sequence could be derived (Fig. 4A). A nearly perfect 17-aa repeat overlapping the first and third repeats was observed, bracketing the deletion in the D3 mutant. With the pneumococcal phage protein it shared a protein region that was predicted to build a coiled coil in both proteins (Fig. 3A). No coiled coil structure was predicted for the φBK5-T protein.

Over six adjacent orfs the streptococcal phages showed significant similarities with five genes from S.
EVOLUTION OF S. thermophilus GENOMES

**Figure 2.** Analysis of the orf1626 gp from φSfi19. (A) Probability of a coiled coil region in orf1626 gp predicted according to the Multicoll program (Wolf et al., 1997). The position and the sequence of a putative leucine zipper is indicated. (B) Multiple sequence alignment of the indicated proteins showing similarity to the C-terminal part of the orf1626 gp. Amino acids identical with the φSfi21 protein are in bold. In the consensus sequence aa positions are in uppercase when conserved in all proteins and in lowercase when conserved in all but one protein. LL-H, Lactobacillus phage LL-H orf360 gp (Accession No. Q38352), phigle, Lactobacillus phage ϕ g l o orf1608 gp (Y003937); XQBO_BACSU, Bacillus subtilis phage-like element skin, YQBO protein (P45931); XKDO_BACSU, B. subtilis phage like element PBSX, XKDO protein (P54334); M.tuberc, Mycobacterium tuberculosis hypothetical 328-kDa protein (Z94121). (C) Localization of the differences between the φSfi19 orf1629 and φSfi21 orf1560gps. Filled circle, point mutation; open boxes, gaps in alignment, open triangle deletion/insertions; the number above the triangle indicates the number of aa deleted/inserted. (D) Comparison of the φSfi19 orf1629 and φSfi21 orf1560gps over the two deletions/insertions (clashes) shown in C. A conspicuous 12-aa repeat was boxed. Identical aa are marked by asterisks.
like inhibitory substance from *S. zooepepideicu* (Simmonds et al., 1997) were observed (data not shown). Areas of similarity with the φDp-1 lysis were concentrated over the N-terminal two-thirds, while the similarity with the cellular protein was concentrated over the C-terminal third.

Structural proteins

Orf 1626 and 1291 gps showed similarity to minor structural proteins from *Lactococcus* and *Lactobacillus* phages (Tables 4 and 5). In both phage groups the structural genes are commonly clustered (Kakikawa et al., 1996; Lubbers et al., 1995; Mikkonen and Alatorssava, 1994; van Sinderen et al., 1996) and in lactococcal phages the structural genes were found upstream of the lysis module (Lubbers et al., 1995; van Sinderen et al., 1996). Since the genome organization of lactococcal phages is very similar to that of the streptococcal φSfi21 (Brutin et al., 1997b; Desiere et al., 1997) one might suspect a structural gene cluster over orf 104 to 610. Extracellular φSfi19 particles were concentrated by PEG precipitation and purified by two rounds of CsCl gradient centrifugation. SDS-polyacrylamide gel electrophoresis yielded molecular weights of 155, 105, 82, 64, 55, and 44 kDa for possible minor phage proteins and 33 and 28 kDa for the two major proteins of φSfi19 (Fig. 6). Since two very large structural proteins were detected one might suspect in orf 1626 and 1291 gp minor structural proteins of the streptococcal phage. It should be noted that no good correlation was found between observed and predicted φSfi19 proteins, as also reported previously for lactococcal phages (van Sinderen et al., 1996). According to N-terminal sequencing, orf 203 codes for a
FIG 4. Amino acid alignments of the orf 1276 gp from \( \text{Sfi} 21 \) (A) Alignment of the collagen-like tripeptide repeats from orf1276 gp. The numbers at the left side give the aa position for the first aa shown in each line of the alignment. Conserved glycine (G) residues are in bold. Boxes indicate two nearly perfectly conserved amino acid stretches. A consensus sequence is given below the alignment which gives positions in uppercase where conserved in all repeats and in lowercase when conserved in all but one protein. The position and sequence of an apparent insertion in \( \text{Sfi} 19 \) is given in the lowest line (B) Pairwise sequence alignment of orf 1276 gp from \( \text{Sfi} 21 \) and orf 1904 gp (Accession No. Q83410). The numbers at the left side give the aa position for the first aa shown in each line of the alignment. Identical aa were noted between the two sequences; they are in bold when they were identical in at least two further lactococcal phages (see Fig 3B). (C) Pairwise sequence alignment of orf 1276 gp with \( \text{Streptococcus pneumoniae} \) phase Dp 1 orf 1 gp (Accession No. 283948). The numbers at the left side give the aa position for the first aa shown in each line of the alignment. The consensus sequence is given in between (+) conserved aa positions.

major phage protein (Le Marrec et al., 1997, GenBank Accession No. AF001793).

**DISCUSSION**

Bioinformatic analysis of a DNA segment covering 40% of the \( S. \ thermophilus \) phage genome indicated that the \( S. \ thermophilus \) phage genome is probably the result of recombination processes between diverse genetic elements from phages infecting Gram-positive bacteria. Our observations concur with the modular theory of phage evolution since exchange reactions between phages differing in many characteristics have been predicted. In the original formulation of the theory the unit of exchange was considered to be a set of related phage genes (Botstein, 1980), while in later modifications of the theory the exchange unit was only defined as a functional segment of a phage genome. A module could therefore be as small as a single gene or even a gene segment coding for a single domain of a protein (Highton et al., 1990; Moore et al., 1981). Several observations made on the \( S. \ thermophilus \) phage genome concur with these later modifications. The similarity of the streptococcal phage DNA with \( \text{Leuconostoc} \) \( \Phi \)10 DNA was
twice interrupted by an unattributed gene and the similarity with \textit{S. pneumoniae} phage DNA was interrupted by one unattributed gene. These observations suggest individual genes as units of exchange. The bioinformatic analysis of three further genes suggested that DNA segments smaller than a whole gene were probably units of modular exchange in the evolution of \textit{S. thermophilus} phage genomes. Orf 1291 gp (or alternatively orf 1904 gp from phage BK5-T) appears as a chimeric protein since distinct protein parts (possibly corresponding to distinct protein domains) could be traced to two different phases. A comparable situation was found in orf 289 gp, whose N-terminal part resembled a phage lysis protein, while the C-terminal part resembled a cellular enzyme. A very similar situation was recently reported for the pneumococcal \phi Dp-1 lysis and this was quoted as an example for the modular evolution of proteins (Sheehan et al. 1997).

The genetic structure of orf 1626 gp is even more complicated since three distinct regions of the protein showed similarities with proteins from phages infecting different bacterial genera. In addition, orf 1626 gp showed similarity to two proteins encoded by two adjacent genes from \textit{Lactobacillales} \phi LL-H. Apparently, at least three distinct domains were combined in this probably multifunctional protein.

In contrast, the comparison of the two streptococcal phage genomes did not reveal modular exchanges. Over the 17-kb DNA fragment investigated both phages differed from each other only by point mutations and small insertion/deletions. Insertion/deletions seem to contribute to the diversification of \textit{S. thermophilus} phages. Three distinct spontaneous deletion mutants of \phi Sfi19 were isolated in the laboratory (Bruttin and Brussow, 1996). One deletion (D2) is most likely a phage integrase-mediated event (Bruttin et al., 1997c). In contrast, the deletion described in this report (D3) was flanked with a nearly perfect 53-bp repeat. Slippage of the DNA polymerase (Singer and Westlye, 1988) is here the most likely interpretation. Interestingly, the smaller deletions observed in two genes of the \phi Sfi19/\phi Sfi21 comparison were also in regions showing DNA repeats.

The DNA sequences from the two phages differed over the investigated genome segment by 10%. We do not know the mutation rate of streptococcal phages. Comparative sequencing in the DNA replication module of multiple isolates of \textit{S. thermophilus} \phi B revealed no point mutations over a 3-year observation period (Bruttin et al., 1997a). \phi B was isolated from a dairy environment in which the phage persisted with high titers and underwent many replication cycles. The average of 19% bp difference observed over orf 104 to 1626 suggests there was a substantial time period since the corresponding DNA segments from \phi Sfi19 and \phi Sfi21 have separated. The fact that such divergent DNA segments could be

| TABLE 3 |

<table>
<thead>
<tr>
<th>gp from orf</th>
<th>Similarity</th>
<th>Identical/Aligned aa</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>\textit{L. delbrueckii} \phi L10 orf K (tr.)</td>
<td>38/84</td>
<td>−20</td>
</tr>
<tr>
<td>116</td>
<td>None</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>153</td>
<td>\textit{L. casei} \phi L10 orf A (164)</td>
<td>43/113</td>
<td>−25</td>
</tr>
<tr>
<td>123</td>
<td>\textit{L. delbrueckii} \phi L10 orf I (127)</td>
<td>31/100</td>
<td>−10</td>
</tr>
<tr>
<td>203</td>
<td>\textit{L. delbrueckii} \phi L10 orf E (206)</td>
<td>81/200</td>
<td>−18</td>
</tr>
<tr>
<td>117</td>
<td>\textit{S. thermophilus} \phi D201 orf Y (tr.)</td>
<td>41/43</td>
<td>−23</td>
</tr>
<tr>
<td>1626</td>
<td>\textit{L. delbrueckii} \phi L10 orf C (tr.)</td>
<td>23/64</td>
<td>−7</td>
</tr>
</tbody>
</table>

Note. gp: identification of the predicted gp from \phi Sfi19. The orfs were listed according to their order on the \phi Sfi19 genome (see Fig. 1). Similarity: identification of the matched proteins by the bacterial species/phage/orf (the lengths in aa of the matched protein, tr. truncated). Additional matches for the orf 1626 gp are given in Table 4: Identical: the number of identical aa for the two compared proteins over the indicated length of computer aligned aa. P value: Probability derived from BLASTP score for obtaining a match by chance. P values are expressed as logarithms: “−20” means \( P = 10^{-20} \). References: \textit{L. delbrueckii} \phi L10 (Sutherland et al., 1994); \textit{S. thermophilus} \phi D201 (Le Marrec et al., 1997).
isolated from two phages of the same factory indicates that homologous recombination has not constantly reshuffled the genomes from both phages. It should be noted that $\phi$Sfi19 was isolated from an fermentation failure in a factory that used a lysogenic starter which contained $\phi$Sfi21 as a prophage. In addition, we deduce that virulent *S. thermophilus* $\phi$Sfi19 was not directly derived from the resident prophage in the lysogenic cell by deletion processes, as could have been suspected from the analysis of deletion mutants from the temperate $\phi$Sfi21 (Bruttin and Brussow, 1996).

Notably, an uneven degree of base pair change was found over the investigated DNA segment: over the *Leuconostoc/Lactobacillus* phage-related DNA segment the two phages showed a relatively high number of base pair differences, while over the *Lactococcus/Streptococcus* phage-related DNA (with the exception of orf 131) only a low number of base pair differences was seen. This observation is not a peculiarity of the $\phi$Sfi19/$\phi$Sfi21 comparison. Dot blot hybridizations with more than 30 phages showed cross-hybridization over the latter, while cross-hybridization with DNA over the former region was much less frequent (Brüssow and Bruttin, 1995). The degree of base pair change between related phages may provide information on the exchange reactions in phages with respect to both the length of the exchanged DNA segments and the temporal order of the exchange reactions. Contiguous DNA segments showing a similar base pair change rate are likely to belong to the same exchange reaction. In addition, DNA segments with a higher base pair change rate are likely to have been acquired in a more distant past than segments with a lower base pair change rate. This hypothesis hinges on the postulation of a temporal accumulation of point mutations. In the case of the streptococcal phages one would thus predict that the *Leuconostoc* phage-related

<table>
<thead>
<tr>
<th>Region: aa position</th>
<th>Similarity</th>
<th>Identical/aligned aa</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-60</td>
<td>L. oenos $\phi$L10 orf C (tr)</td>
<td>23/64</td>
<td>-7</td>
</tr>
<tr>
<td>10-200</td>
<td>Streptococcus pyogenes FeReceptor (405)</td>
<td>40/193</td>
<td>-5</td>
</tr>
<tr>
<td>70-520</td>
<td>Chicken myosin (1940)</td>
<td>83/430</td>
<td>-5</td>
</tr>
<tr>
<td>70-520</td>
<td>Lactococcus lactis $\phi$L167 orf 31 (620) minor tail protein</td>
<td>86/403</td>
<td>-6</td>
</tr>
<tr>
<td>130-1550</td>
<td>Lactobacillus $\phi$Te orf 1608 minor structural protein</td>
<td>355/1512</td>
<td>-33</td>
</tr>
<tr>
<td>920-1030</td>
<td>Lactobacillus delbrueckii $\phi$LL-H orf 150 (150)</td>
<td>43/120</td>
<td>-10</td>
</tr>
<tr>
<td>1360-1580</td>
<td>L. delbrueckii $\phi$LL-H orf 360 (360)</td>
<td>114/235</td>
<td>-49</td>
</tr>
<tr>
<td>1420-1540</td>
<td>Bacillus subtilis $\phi$-like element skin: papa gene (1581)</td>
<td>65/129</td>
<td>-22</td>
</tr>
<tr>
<td>1350-1550</td>
<td>B. subtilis $\phi$-like element PBSX, xdo gene (1332)</td>
<td>72/176</td>
<td>-21</td>
</tr>
<tr>
<td>1350-1510</td>
<td>Mycobacterium tuberculosis: gene CT15F10.16 (302)</td>
<td>40/108</td>
<td>-6</td>
</tr>
</tbody>
</table>

Note: Region: Identification of the aa position from the orf 1626 gp which showed a match to the indicated entry of the database. Similarity: The matched protein was identified by the bacterial species/phage/orf (if applicable)/orf (the lengths in aa of the matched protein: tr, truncated). Identical: the number of identical aa for the two compared proteins over the indicated length of computer-aligned aa. $P$ value: Probability derived from BLASTP score for obtaining a match by chance. $P$ values are expressed as logarithms: "$-7" means $P = 10^{-7}$. References: $\phi$L167 (Schouler et al., 1994), $\phi$LL-H (Vasala et al., 1995), $\phi$Te (Kakikawa et al., 1996, Kodaira et al., 1997), the bacteriocin-like PBSX element (Krogh et al., 1996), and the sporulation-related skin element (Medigue et al., 1995). Access Nos: S. pyogenes Q54862, myosin P02565, Mycobacterium 005449.
Table 6

Similarity of the Indicated Gene Products from φSfi19 to Proteins from the Database

<table>
<thead>
<tr>
<th>gp from orf</th>
<th>S. pneumoniae φOp-1 orf 1 (truncated)</th>
<th>S. pneumoniae φOp-1 orf 2 (532)</th>
<th>S. pneumoniae φOp-1, holin (134)</th>
<th>B. subtilis φ-like element skin, yqfh gene (140)</th>
<th>S. pneumoniae φEJ-1, holin (85)</th>
<th>S. pneumoniae φOp-1, pal gene (687) lysin</th>
<th>L. lactis φBK5-T, orf 259 (253)</th>
<th>Streptococcus zoopneumoniae, zooA gene (255), bacteriocin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1291</td>
<td>93/102</td>
<td>58/153</td>
<td>23/87</td>
<td>39/121</td>
<td>31/63</td>
<td>79/248</td>
<td>40/125</td>
<td>58/204</td>
</tr>
<tr>
<td>670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>141</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>289</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: gp, identification of the predicted gp from φSfi19. The orfs were listed according to their order on the φSfi19 genome (see Fig. 1). Similarity: identification of the matched proteins by the bacterial species/phae (if applicable)/orf (the lengths in aa of the matched protein). Identical: the number of identical aa for the two compared proteins over the indicated length of computer-aligned aa. P value: Probability derived from BLASTP score for obtaining a match by chance. P values are expressed as logarithms. "−36" means P = 10^−36. References: φOp-1 (Sheehan et al., 1997), φEJ-1 (Martin et al., 1996), φEJ-1 (Lopez et al., 1992), zooA gene (Simmonds et al., 1997).

DNA was acquired before the Lactococcus/Streptococcus phage-related DNA. The higher degree of bp changes in orf 131/117b than in the adjacent genes makes it unlikely that the Lactococcus/Streptococcus phage-related DNA was the result of a single exchange reaction. We suspect that the DNA related to orfs 410 and 1904 from the lactococcal φBK5-T, the DNA related to orfs 1 and 2 from the pneumococcal φOp-1, the unattributed orf 131, and the DNA covering the lysis cassette consisting of the two holins and the lysin represent at least four distinct exchange reactions.

In summary, the comparative sequence analysis of two S. thermophilus phages suggests two processes in the creation of S. thermophilus phage genomes: in a first evolutionarily distant phase the constitution of the basic genome structure by recombination processes and in a second more recent phase the diversification of the basic genome structure by the accumulation of point mutations and small deletions and insertions.

Materials and Methods

Phages, strains and media

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

DNA techniques

Phage purification and DNA extraction were done as described previously (Bruttin and Brüssow, 1996; Brüssow and Bruttin, 1995; Bruttin et al., 1997a,b). Plasmid DNA was isolated using Qiagen midiplasmid 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 five-step preformed CsCl gradient (n₀ = 1.4, 1.372, 1.3698, 1.3682, 1.367). The phage bands were recovered with a Pasteur pipette, diluted in phage buffer (Brüssow and Bruttin, 1995), and concentrated by high-speed centrifugation (1 h, 40,000 rpm, SW55.5 rotor Beckman). The purified phage particles were then denatured for 2 min at 100°C using SDS gel loading buffer with β-mercaptoethanol. SDS-PAGE was done on 8% gels for 2 mm at 100°C using SDS gel loading buffer with β-mercaptoethanol.

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

Sequence analysis

The Genetics Computer Group sequence analysis package (University of Wisconsin) was used to assemble and analyze the sequences. Nucleotide (nt) and predicted amino acid (aa) sequences were compared to those in the databases (GenBank, Release 102; EMBL (abridged), Release 51; PIR-Protein, Release 53; SWISS-PROT, Release 34; PROSITE, Release 13.0) using 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 Multalin program (Corpet, 1988), and the SIM alignment tool (Huang and Miller, 1991). The φSfi21 sequence was deposited in the GenBank database under Accession No. AF032121 and the φSfi19 sequence under Accession No. AF032122.

ACKNOWLEDGMENTS

We thank Anne Bruttin for help in protein analysis, Sophie Foley for critical reading of the manuscript, and the Swiss National Science Foundation for financial support of Frank Desiere and Sasha Lucchini in the framework of its Biotechnology Module (Grant 5002-04545/1).

REFERENCES


cilus temperate phage phi g1e: The whole genome sequence and the putative promoter/repressor system. Gene 187, 45-53.

Krogh, S., O’Riilly, M., Noilan, N., and Davine, K. M. (1996). The phage-like element PBSX and part of the skin element, which are resident at different locations on the Bacillus subtilis chromosome, are highly homologous. Microbiology 142, 2031-2040.


9 Comparative Sequence Analysis of the DNA Packaging and Morphogenesis Genes in the Temperate cos-site Streptococcus thermophilus Bacteriophage Sfi21

Published in Virology. 1999 August 1;260, 244-253.

Desiere F, Lucchini S, Brüssow H
Comparative Sequence Analysis of the DNA Packaging, Head, and Tail Morphogenesis Modules in the Temperate cos-Site Streptococcus thermophilus Bacteriophage Sfi21

Frank Desiere, Sacha Lucchini, and Harald Brüssow

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

Received December 22, 1998; returned to author for revision May 13, 1999; accepted June 1, 1999

The temperate Streptococcus thermophilus bacteriophage Sfi21 possesses 15-nucleotide-long cohesive ends with a 3' overhang that reconstitutes a cos site with twofold hyphenated rotational symmetry. Over the DNA packaging, head and tail morphogenesis modules, the Sfi21 sequence predicts a gene map that is strikingly similar to that of lambdoid coliphages in the absence of any sequence similarity. A nearly one to one gene correlation was found with the phage lambda genes Not to H2 except for gene B-to-E complex, where the Sfi21 map resembled that of coliphage HK97. The similarity between Sfi21 and HK97 was striking: both major head proteins showed an N-terminal coiled-coil structure, the mature major head proteins started at amino acid positions 105 and 104, respectively, and both major head genes were preceded by genes encoding a possible protease and portal protein. The purported Sfi21 protease is the first viral member of the ClpP protease family. The prediction of Sfi21 gene functions by reference to the gene map of intensively investigated coliphages was experimentally confirmed for the major head and tail gene. Phage Sfi21 shows nucleotide sequence similarity with Lactococcus phage BK5-T and a lactococcal prohuge and amino acid sequence similarity with the Lactobacillus phage A2 and the Staphylococcus phage PVL. PVL is a missing link that connects the portal proteins from Sfi21 and HK97 with respect to sequence similarity. These observations and database searches, which demonstrate sequence similarity between proteins of phage from gram-positive bacteria, proteobacteria, and Archaea, constrain models of phage evolution.

INTRODUCTION

Streptococcus thermophilus is a gram-positive thermophilic lactic acid bacterium used in industrial milk fermentation (Mercenier, 1990). Phage attack has always been a major problem in the dairy industry and is associated with substantial economical loss. This is also the case for S. thermophilus (Brüssow et al., 1998). The genetic construction of phage-resistant bacterial starters is thus an important biotechnological goal. In contrast to the related genus Lactococcus, lactic streptococci possess few plasmids encoding natural defence systems of the bacterial host against phage infection (Larbi et al., 1992). We therefore turned our attention to the bacteriophage genome to identify and exploit phage-encoded control systems that interfere with phage replication (Foley et al., 1998). Because relatively few data were available on the biology of these phages (reviewed in Brüssow et al., 1998; Brüssow, 1999) and because tools for genetic analysis in lactic streptococci are not very developed, several laboratories decided to sequence, as a first step, a complete phage genome to obtain an overview of their genome organization (Stanley et al., 1997). However, in many bacterial genome sequencing projects, only about half of all identified open reading frames (ORFs) find a match in the database. This situation is even worse in bacteriophage genome sequencing projects: rarely more than a fourth of the phage ORFs match entries in the database. A possible remedy to this situation could be the sequencing of genetically closely related phages that differ in phenotypes. We applied this approach to S. thermophilus bacteriophages that differed in DNA packaging (cos or pac type phages), lifestyle (lytic or temperate phages), host range, or sensitivity to prophage-mediated superinfection exclusion (Desiere et al., 1998; Lucchini et al., 1998; Lucchini et al., 1999a, b). This analysis allowed the establishment of associations between phenotype and genotype and thus the attribution of possible functions to a few ORFs.

When comparing the genome of a pac-type S. thermophilus bacteriophage Sfi11 with Siphoviridae from gram-positive bacteria, we observed not only a similar gene order but also sequence relationships with a Lactococcus and a Bacillus phage (Lucchini et al., 1999). In addition, similarity with the structural gene map of phage λ was noted for the Lactococcus phage sk1 (Chandy et al., 1997), the Streptococcus phage Sfi11 (Lucchini et al., 1998), and the Bacillus phage SPP1 (Becker et al., 1997). Because phage 2 is the best investigated biological system, the similarity might be exploited to predict gene functions in gram-positive bacteria to lead future biological experimentation. However, for the prediction of gene functions, the observed correlations between the gene...
maps of the phages investigated were not sufficiently close. We suspected that the smaller genome size of the lactococcal cos-site phage sk1 (28 kb) (Chandry et al., 1997) or the different DNA packaging mechanism of the pac-site streptococcal phage Sf11 and B. subtilis phage SPP1 prevented a closer match with the gene map of the cos-site coliphage. Therefore, we investigated the genome organization of the 40-kb cos-site S. thermophilus phage Sf121. Over a 15-kb structural gene cluster, we observed a nearly 1:1 gene/ORF correlation between the two phages.

The theoretical basis for this correlation is not clear because the comparable genome organization was found in the absence of any sequence relationship. Taxonomically oriented virologists favour the comparison of genomic maps over protein sequence alignments for the understanding of phage evolution because sequence alignments have, until now, essentially confirmed relationships between phages that were known to be related (Ackermann, 1999). However, the total number of phage sequences in the database is small compared with other organisms and the lack of protein sequence alignments between more distantly related phages might reflect an observation bias. Currently, the situation is about to change with the introduction of many new phage sequences into the database. We demonstrate here similarities not only in the gene organization but partially also in protein sequences of phages whose hosts include, besides streptococci and its close relatives, Escherichia coli, Rhodobacter (a proteobactenum of the gamma subdivision), and an archaeon.

RESULTS

The cos-site of Sf121

During direct sequencing of the heat-treated unligated phage DNA by primer walking, we observed an abrupt end of the sequencing reaction upstream of ORF 152 (Fig. 1). When re-annealed, ligated Sf121 DNA was used as the template, direct DNA sequencing with a primer located in ORF 152 led into ORF 175 sequences, and vice versa. When the heat-treated Sf121 DNA was first digested with T4 DNA polymerase, which carries a 3'-to-5' exonuclease activity, then ligated and subsequently sequenced, we obtained an identical nucleotide sequence except for the absence of a 15-bp stretch between ORFs 175 and 152. This analysis suggested the presence of a 15-nucleotide-long, 3'-extended single-stranded cDNA at the two ends of the phage DNA. This 15-bp DNA stretch showed a twofold hyphenated rotational symmetry (Fig. 2). We propose that this sequence constitutes the cosN site of Sf121. The cosN site was located in an 186-bp noncoding AT-rich DNA region. This region showed a weak match with pST1 plasmid from S. thermophilus (P = .015), (Janzen et al., 1992).

Bioinformatic analysis of the cos-site associated genes of Sf121

Over the genome region depicted in Fig. 1, the predicted phage Sf121 proteins showed close sequence relationships with the corresponding proteins from the cos-site S. thermophilus phage DT1 (Tremblay and Moineau, 1999). The only major difference was the separation of Sf121 ORF 623 into two ORFs in DT1 (Table 1). Except for phage DT1, no database links could be established for ORF 132 gp from phage Sf121. ORF 175 gp showed in addition strong similarity to a protein from Staphylococcus aureus phage PVL (Table 1) and weak similarity to cos-site-associated genes from Lactobacillus casei phage A2 and Bacillus subtilis phage phi-105. A multiple alignment of the proteins revealed 32 amino acid (aa) positions shared in at least three of four phage proteins (data not shown). ORF 152 showed significant sequence similarity to a probably defective Lactococcus lactis prophage gene that abuts on a large genome inversion in the host bacterium (Daveran-Mingot et al., 1998) and probably prophage-encoded Mycobacterium tuberculosus proteins. In addition, weak similarities with proteins encoded by cos-site-associated genes from phage A2 and actinophage RP3 were detected (Table 1). Multiple alignments of the above-mentioned five proteins identified 39 aa positions that were shared in at least three proteins. ORF 623 gp showed very high similarity with proteins from the same L. lactis prophage, phages A2 and PVL (Table 1). All three phage proteins were encoded by the second gene after the respective cos-sites (Fig. 3). In addition, ORF 623 gp shared significant sequence similarity with putative terminases from the L. lactis phages c2 and bIL67 and a prophage-encoded protein from Rhodobacter capsulatus (Table 1, Fig. 3). A multiple alignment of the Streptococcus, Lactococcus, Lactobacillus, Staphylococcus, and Rhodobacter phage and prophage proteins showed 41 invariant aa positions. The presence of a typical Walker box for an ATP-binding protein further supports the suggestion that ORF 623 gp encodes a large subunit terminase.

The putative head morphogenesis genome region of Sf121

Except for phage DT1, no bioinformatic links could be established for ORF 59 gp. ORF 384 gp from Sf121 showed significant sequence similarity with a protein from phage PVL encoded at a corresponding position (Table 1, Fig. 3). Notably, the PVL protein showed significant similarity to the portal protein from coliphage HK97 (Kaneko et al., 1998), whereas the HK97 protein showed sequence similarity (P = 10⁻³) with ORF 426 gp from the Rhodobacter phage (Fig. 3). ORF 221 gp from Sf121 showed strong sequence similarity to different members of the large ClpP protease family (10 members showed P = 10⁻³ to 10⁻⁶). The attribution of ORF 221 gp to this large family of serine proteases is further supported by a multiple alignment. When compared with eight members of the ClpP protease
FIG. 1. Comparison of the DNA packaging and head and tail assembly genes in the cos-site-containing streptococcal phage Sfi21 (green box) and the cos-site E. coli phage Λ and the Methanobacterium phage psiM2. In addition, a partial gene map of the lambdoid coliphage HK97 is provided. The numbers below the maps give the lengths of the deduced proteins in aa. The phage Λ genes are identified by their letter code above the arrows and by a short description of the gene functions below the arrows. Corresponding genes in the four phages are indicated with the same color code (AC gene gives a bioinformatic protease link, but protease activity has not been identified).
Fig. 2. The nucleotide sequence of the cos-region from S. thermophilus phage Sfi21. The 16-nucleotide 3' overhang from the cohesive ends is marked by a large box; the region of twofold hyphenated rotational symmetry is shaded.

family of bacterial, plant, and animal origin, the Sfi21 protein shared 28 positions with at least six of them (data not shown). Especially well conserved were the diagnostic catalytic serine and histidine residues (Maurizi et al., 1990), which are flanked by conserved glycine residues (Fig. 5).

The Sfi21 protein is, to our knowledge, the first case of a virally encoded ClpP-like protein. The Sfi21 ClpP-like protein is unlikely to result from a horizontal gene transfer from the genome of its bacterial host: two different phylogenetic tree analyses demonstrated that the Sfi21 protein is not more closely related to bacterial (including a close sister species of S. thermophilus, namely S. salivarius) than to algal ClpP proteins (Fig. 4). Because a closely related protein was detected in the lactococcal phage BK5-T and was encoded at a corresponding map position (A. Hillier, personal communication), we might deal here with a new class of bacteriophage-encoded ClpP-like proteins. Interestingly, corresponding map positions were occupied by genes encoding likely proteases in coliphages HK97 (Hendrix and Duda, 1998) and N15 (R. W. Hendrix, unpublished observation).

The ORF 303 gp from the Rhodobacter prophage showed sequence relationship with the putative protease from N15 (P = 10^-11). However, none of these possible phage proteases belonged to the ClpP protein family.

ORF 397 gp shares very high sequence similarity with the major head protein of L. lactis phage BK5-T (A. Hillier, personal communication) and significant similarity with the major head protein of phage PVL (Table 1).

N-terminal sequencing of the larger of the two major structural proteins of phage PVL showed sequence relationship with the putative protease from N15 (P = 10^-1). However, none of these possible phage proteases belonged to the ClpP protein family.

Table 1: Open Reading Frames of the DNA Packaging, Head, and Tail Morphogenesis Module of Sfi21 and Similarities with Databases

<table>
<thead>
<tr>
<th>ORF</th>
<th>Put. function</th>
<th>Similarity to</th>
<th>Gene</th>
<th>P</th>
<th>id/sim/length (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>132</td>
<td>S. thermophilus phage DT1</td>
<td>10^-31</td>
<td>ORF 63</td>
<td>86/95/132</td>
<td>Tremblay and Moineau (1999)</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>S. thermophilus phage DT1</td>
<td>10^-26</td>
<td>ORF 2</td>
<td>84/93/168</td>
<td>Tremblay and Moineau (1999)</td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>S. thermophilus phage DT1</td>
<td>10^-14</td>
<td>ORF 153</td>
<td>76/87/162</td>
<td>Tremblay and Moineau (1998)</td>
<td></td>
</tr>
<tr>
<td>623</td>
<td>Terminase large subunit</td>
<td>10^-72</td>
<td>ORF 4</td>
<td>21/40/138</td>
<td>Garcia et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>S. thermophilus phage DT1</td>
<td>10^-14</td>
<td>ORF 59</td>
<td>30/53/542</td>
<td>Garcia et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>Portal protein</td>
<td>10^-14</td>
<td>ORF 105</td>
<td>24/64/339</td>
<td>Kaneko et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>ClpP-like protein</td>
<td>10^-14</td>
<td>ORF 221</td>
<td>90/96/221</td>
<td>Tremblay and Moineau (1998)</td>
<td></td>
</tr>
<tr>
<td>397</td>
<td>Major head protein</td>
<td>10^-14</td>
<td>ORF 234</td>
<td>86/90/290</td>
<td>Tremblay and Moineau (1989)</td>
<td></td>
</tr>
</tbody>
</table>

Note. For the database matches of the ORFs downstream of the major head gene, see Desiere et al., (1998). ORF: identification of the phage Sfi21 ORF by codon length; Put. Function: putative function deduced from bioinformatic analysis or N-terminal sequencing (ORF 397 gp); Gene: gene identification as annotated in the entry; id/sim/length: percent identity/similarity over the indicated length.
Prophage

BKY-T

Streptococcus

Sfl21

Staphylococcus

PVL

Lactobacillus

A2

Rhodobacter

Prophage

HK97

E. coli

FIG. 3. The map of the cos-site-associated genes in the temperate S. thermophilus phage Sfl21 was compared with the corresponding gene maps from Staphylococcus aureus phage PVL, Lactobacillus casei phage A2, incompletely sequenced Lactococcus lactis phage BK5-T and a L. lactis prophage, a Rhodobacter capsulatus prophage, and E. coli phage HK97. The predicted ORFs are indicated as in their published reports or with their codon length. Corresponding genes showed the same shading. The position of the cos-site is indicated where known. Genes whose predicted proteins show significant aa sequence similarity are connected.

teins from Sfl21 (Desiere et al., 1998) confirmed that the ORF 397 gp is the likely major head protein. Tail-less phase particles contain only the larger 32-kDa protein (data not shown). The N-terminal peptide sequence (LLDSKT) from the larger phase structural protein resolved a discrepancy between the predicted (44 kDa) and the observed (32 kDa) molecular mass of the putative major head protein. The N-terminal sequence corresponds to aa positions 105–110 of the predicted ORF 397 gp, suggesting possible proteolytic processing of this protein during maturation. Processing of the ORF 397 gp at aa position 105 predicts a 32-kDa protein, in agreement with the experimentally determined value.

The smaller phage structural protein, which is lost in tail-less phase particles, yielded the peptide sequence AIVGLK, which matched aa positions 2–7 from ORF 202 gp (Fig. 1). The first methionine was lacking in accordance with the rule that the N-terminal methionine is generally processed when the second aa residue is an alanine (Ben-Bassat et al., 1987).

Graded conservation of the morphogenesis module in tailed phages

When analyzing the morphogenesis genes from a pac-site S. thermophilus phage, we previously observed a hierarchy of relationships between phages that correlated with the evolutionary distance of the host bacteria to lactic streptococci (Lucchini et al., 1998). This was also seen for the cos-site S. thermophilus phage. The closest relationships of phage Sfl21 were to the lactococcal phages that could still be aligned at the nucleotide level (P < 10⁻⁵): the major head gene from BK5-T (A. Hillier, personal communication) showed 66% bp identity, and the terminase gene from the lactococcal prophage showed 64% bp identity with the corresponding Sfl21 gene. At the next level were similarities between Sfl21 and the Staphylococcus phage PVL and the Lactobacillus phage A2. The genetic organization was similar, and several phage genes showed sequence similarity at the deduced aa level but no longer at the nucleotide level. In addition, sequence-related genes were separated by sequence-unrelated genes: PVL ORF 63, 2, 4, and 7 gps were related to Sfl21 proteins, whereas PVL ORF 1, 3, 5, and 6 gps were unrelated (Fig. 3).

At the next level, Sfl21 showed similarity in the genetic organization with phage A (Fig. 1), whereas no sequence similarity at the protein level could be detected. In fact, only two differences were seen in the genetic organization of the left 15 kb from Sfl21 and A. Sfl21 ORF 221 occupied the position of the phage A gene C-Nu3-D complex and Sfl21 ORF 117 occupied the position of the gene G-T complex in A. Interestingly, over the first variant region, Sfl21 showed exactly the gene constellation described for the lambdoid coliphage HK97 (Hendrix and
FIG. 4. Phylogenetic tree analysis of the ORF 221 gp from S. thermophilus phage Sfi21 in comparison with putative CiP proteases from bacteria and plants. The tree was derived by the DARWIN function phylotree, with the distances given in PAM units. The branches are labeled with the SwissProt mnemonic of each protein or the species name. Streptococcus salivarius, the closest species to S. thermophilus, is in bold.

Duda, 1998) (Fig. 1). The similarity between Sfi21 and HK97 was even more striking: first, the major structural proteins from both phages are likely to be proteolytically cleaved at a comparable position (HK97 aa position 104; Duda et al., 1995; Sfi21 aa position 105). Second, a strong coiled-coil structure could be predicted for the processed N-terminal region of both proteins (Conway et al., 1995) (Fig. 6). Interestingly, the major head protein from S. aureus PVL showed similar properties to the HK97 and Sfi21 proteins: the mature protein starts at aa position 117 (Kaneko et al., 1998), and a coiled-coil is predicted over the N-terminal 70 aa (data not shown). Third, the HK97 and Sfi21 major head genes were preceded by a putative protease gene. Phage HK97 has been noted for its absence of a scaffolding protein. Hendrix and Duda (1998) proposed that the proteolytically processed N-terminal part of the major head protein fulfills a scaffolding function. From the similar genomic organization in phage Sfi21, we suspect that this constellation might be more widely distributed than initially thought.

It is fascinating to note that the gene map of the Methanobacterium thermoautotrophicum phage psiM2, an archaeavirus (Pfister et al., 1998), also resembles those of phages Sfi21 and A (Fig. 1). Further support for an ancestral genetic relationship between these three phage systems comes from significant sequence relationships. For example, the portal and major head protein from psiM2 showed similarity with the corresponding

FIG. 5. Multiple alignment of the ORF 221 gp from S. thermophilus phage Sfi21 with putative CiP P proteins from Borrelia burgdorferi, Paracoccus denitrificans, Synechocystis sp., Bacillus subtilis, Streptococcus salivarius, Caenorhabditis elegans, Mycobacterium tuberculosis, and Chlorella vulgaris over the characteristic catalytic CiP motif. The catalytic aa residues serine and histidine are marked by arrows. Amino acid positions identical in at least five proteins are shaded, aa positions identical in at least eight of the nine aligned proteins are underlined in black.
FIG. 6. Probability of a coiled-coil region in the major head proteins from E. coli phage HK97 (gp5) and S. thermophilus phage Sfi21 (ORF 353 gp) predicted according to the Multicoil program. The location of the N-terminus of the mature protein from the extracellular phage particle is indicated by an arrow and the aa position.

proteins from the gram-negative Haemophilus influenzae phage HP-1 \( (P = 10^{-6}; \) Esposito et al., 1996) and the gram-positive B. subtilis prophage PBSX \( (P = 10^{-4}; \) Takemaru et al., 1995; Pfister et al., 1998; and our own database searches). In fact, alignments of the three portal and head proteins demonstrated 21 and 32, respectively, identical aa positions (data not shown).

**DISCUSSION**

Over the DNA packaging, head and tail morphogenesis modules, a nearly one to one correlation was observed between genes from a cos-site S. thermophilus phage and lambdoid coliphages. The probability that this correlation occurred by chance is very low. The basic observation of the conservation of the morphogenesis genes in some tailed phages is not new; it was noted more than 25 years ago \( (\) Dove, 1971) that the prophage maps of P2, P22, and \( \lambda \) were partially congruent. The observation was confirmed by sequence comparison between P22 and \( \lambda \) \( (\) Epler et al., 1991). Evolutionary mechanisms that could favour the conservation of the structural gene order in lambdoid phages have been proposed \( (\) Casjens and Hendrix, 1974). This congruence in the gene maps was recently extended to bacteriophage-encoding gram-positive bacteria \( (\) Becker et al., 1997; Chandry et al., 1997; Lucchini et al., 1998). The correlation between the Sfi21 and phage \( \lambda / HK97 \) gene maps reported in the present communication is so close that one is tempted to exploit the similarity to predict gene functions for Sfi21 ORFs. The alignment with the phage map predicts that Sfi21 ORFs 152 and 623 encode the small and large subunits of terminase, respectively, whereas ORF 59 would be implicated in head-to-tail joining. The alignment with the HK97 map predicts that the three ORFs 384, 221, and 397 encode the portal protein, a prohead protease, and the major head protein, respectively. For ORFs 106, 116, 141, 123, and 202, one might predict analogous functions to the \( F, F, Z, U \), and \( V \) genes from \( \lambda \), which are implicated in head-to-tail joining processes. According to that prediction, ORF 202 should encode the major tail protein. The attribution of ORF 1560 to the \( H \) gene from \( \lambda \), implicated in DNA injection and tail length determination, is suggested by the map position and the sheer length of the gene. In contrast, the assignment of ORF 117 to a gene is problematic because two genes, \( G \) and \( T \), are found at the corresponding position in \( \lambda \).

Two of the above predictions were experimentally confirmed. ORFs 397 and 202 gps are the major head and tail proteins, respectively, as demonstrated by a combination of N-terminal sequencing and electron microscopy. The validity of the other assignments is supported by indirect arguments: a minor structural protein of 155 kDa \( (\) Desiere et al., 1998) is likely to be encoded by ORF 1560; ORF 623 gp shows biomorphic links to terminases and an ATP-binding motif, whereas ORF 384 gp showed biomorphic links to portal proteins. ORF 221 gp contained a motif found at the catalytic site of a specific class of serine protease. A further argument in favor of the predictive power of the correlation hypothesis is the observation that many biomorphic links to phage genes were located at corresponding positions of the phage genome maps. We believe, therefore, that comparative sequence "gazing" has, in phage genomics, a much better chance of producing new biological understanding than simple BLAST searches for individual phage genes.

What might be the theoretical basis for this correlation of the gene maps of tailed phages infecting bacterial hosts as distinct as gram-positive and -negative organisms? Taxonomically, both phages belong to the same order Caudovirales, family Siphoviridae, genus "\( \lambda \)-like viruses" \( (\) Manioff and Ackermann, 1998). Over the structural genes, which are likely to be indigenous to phages and where relationships must be sought in the first place, phages Sfi21 and \( \lambda \) might thus share a common origin. Why, then, do we not detect sequence relationships between these two phages? Ackermann \( (1999) \) compiled a list of arguments to deal with this dilemma. One is that tailed phages are so much older than plant or
animal viruses that relationships were erased and no longer detectable by aa alignments (Doolittle, 1981). Indirect evidence in favor of this argument is the observation that we actually get aa sequence alignments when we compare "A-like viruses" over smaller evolutionary distances, such as from phages infecting low GC-content gram-positive bacteria (Lucchini et al., 1998; and current report). Another argument of Ackermann (1999) is that relationships (e.g., between capsid proteins) may be so weak that they are detectable only through examination of missing links; for example, proteins A, B, and C may be related but may appear unrelated if only A and C are compared. If we equate the putative portal proteins from phages Sfi21, PVL, and HK97 with A, B, and C, we have exactly what is predicted (and we could still propose a D for the putative portal protein of the Rhodobacter prophage). A further argument is that relationships may be preserved in the three-dimensional structure of tailed phage capsids and tail proteins but not any longer in sequence (Ackermann, 1999). This hypothesis cannot be tested yet because not a single one of these phage proteins has been studied for its spatial structure. Again, our comparisons provide an indirect illustration for the validity of this hypothesis. The major head proteins from Sfi21 and HK97 do not share any detectable sequence similarity, yet they yield an extremely similar coiled-coil prediction and they are cleaved at a nearly identical aa position. Both arguments speak in favor of a rather similar spatial structure. Ackermann (1999) is skeptical about the use of aa sequence alignments for addressing the question of phage evolution simply because the evidence of relatedness of phage proteins has not yet been found. This is, however, not an argument against aa alignments but an encouragement to sequence more phage genomes from carefully chosen bacterial genera. An increasing number of missing links are obtained as the number of sequenced phage genomes increases (Hendrix et al., 1999). This will soon allow an integrated, sequence-based picture of phage evolution.

MATERIALS AND METHODS

Bacterial strains, phages, and media

*Streptococcus thermophilus* Sfi1 was grown at 40°C in an M17 broth (Difco Laboratories, Detroit, MI) supplemented with 0.5% lactose. Phage Sfi21 was isolated from a batch of Sfi21 cells during starter preparation and independently from a lysogenic cell Sfi19 (Brüssow and Bruttin, 1995). For phage preparation, calcium chloride was added to the medium to a final concentration of 10 mM. Phages were purified and concentrated as described previously (Brüssow and Bruttin, 1995).

Cloning and sequencing

Sfi21 DNA was first purified using the Wizard DNA purification kit (Promega; Madison, WI) and then cut with restriction enzymes *EcoRI, HindIII, NsiI, XbaI, or Sau3A* (Boehringer Mannheim, Mannheim, Germany) and cloned into vector pUC19 or the *E. coli*-lactococcal-streptococcal shuttle vector pNZ124 (Platteeuw et al., 1994).

DNA sequencing was started with the universal forward and reverse primers of pUC19 or pNZ124 and continued with synthetic oligonucleotide (17-mer) primers (Microsynth, Switzerland). Both strands of the cloned DNA were sequenced by the Sanger method of dideoxynucleotide's mediated chain termination using either the fmol DNA Sequencing System of Promega (Madison, WI) as described previously (Brüttin et al., 1997) or the Amersham Lab-station sequencing kit based on the Thermo Sequenase-labeled primer cycle-sequencing with 7-deaza-dGTP (RPN2437). The thermal cycler was programmed at 25 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min. Sequencing was done on a Licor 6000L automated sequencer with fluorescence-labeled universal reverse and forward pUC19 primers.

PCR was used to span regions not obtained through random cloning. PCR products were generated using the synthetic oligonucleotide pair designed according to the established phage Sfi21 DNA sequence, purified phage DNA, and Super *Taq* polymerase (Stehelin, Basel, Switzerland). PCR products were purified using the QiAquick*Spin* PCR purification kit.

The *cos*-sites were digested using purified phage DNA and T4 ligase according to the manufacturer's protocol (Boehringer Mannheim). The *cos*-overhangs were digested by the 3'-to-5'-end exonuclease activity of the T4 polymerase in the appropriate buffer (Boehringer Mannheim) and incubated for 5 min at room temperature.

Protein techniques

Extracellular phage particles were purified by PEG precipitation and CsCl density gradient centrifugation as described previously (Brüssow and Bruttin, 1995). The purified phage particles were denatured at 95°C for 10 min in Laemmli's sample buffer (Bio-Rad, La Jolla, CA) and then loaded onto a 12% SDS–polyacrylamide minigel running at 200 V for 60 min. The gel was stained with Coomassie Brilliant Blue (Bio-Rad). Broad-range prestained SDS–PAGE standards (Bio-Rad) were used to estimate the size of the stained proteins.

For N-terminal sequencing, the polyacrylamide gel was transferred to a PVDF membrane (Immobilon; Millipore, Bedford, MA) using a Bio-Rad Trans-Blot cell in transfer buffer (50 mM, pH 9). The proteins were then visualized with Coomassie Brilliant Blue (0.1% Coomassie blue, 40% methanol, 1% acetic acid), and the protein of interest was cut from the membrane for sequencing by
Edman degradation with an Applied Biosystems 473A pulsed liquid protein sequencer (PE Applied Biosystems, Foster City, CA).

Sequence analysis

The Genetics Computer Group sequence analysis package (University of Wisconsin, Madison, WI) was used to assemble and analyze the sequences. Nucleotide and predicted aa sequences were compared with those in the databases [GenBank, Release 110; EMBL (Abridged), Release 58; PIR-Protein, Release 59; SWISS-PROT, Release 37; PROSITE, Release 15.0] using FastA (Pearson and Lipman, 1988) and BLAST (Altschul et al., 1990) programs. Sequence alignments were performed using the CLUSTAL W 1.74 method (Thompson et al., 1994) and the BLOCKMAKER program (Henikoff et al., 1995). The phylogenetic tree was constructed using the Computational Biochemistry Research Group Server (Zurich, Switzerland).

Accession numbers

The complete genome sequence of Sfi21 was reported recently (Lucchini et al., 1999) and was deposited under the accession number AF115103.

ACKNOWLEDGMENTS

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 5002-04454/1) and S. Foley for reading the manuscript.

Note added in proof. A very similar module for capsid assembly was recently described for the temperate phage T231 which infects the high GC content gram-positive bacterium Streptomyces (Smith et al., 1999).

REFERENCES


Kaneo, J., Kimura, T., Narita, S., and Tomita, T. (1999). Complete nucleotide sequence and molecular characterization of the temper-


10 Packaging and morphogenesis genes from *Lactobacillus johnsonii* and *E.coli* phages reveal different modes of evolution of *Siphoviridae*

Frank Desiere¹, R. David Pridmore¹, Stephan Beck², Bart Barrel², and Harald Brüssow¹

¹ Nestlé Research Center, Nestec Ltd., Vers-chez-les-Blanc, CH Lausanne 26, Switzerland
² The Sanger Center, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

**INTRODUCTION**

Bacteriophage attack has always been a major problem in industrial fermentation, especially in the dairy industry (Mercenier, 1990; Peitersen, 1991). Phages are a threat to the dairy industry where they contaminate the milk fermentation line and cause fermentation delays or even loss of the product. The economical importance of dairy phages led to intensive research efforts on phages attacking bacterial starters used in the dairy industry like *Streptococcus thermophilus*, *Lactococcus lactis* and various species of Lactobacilli. Our laboratory is interested in the genetic diversity of bacteriophages infecting these lactic acid bacteria used in the production of yoghurt and Italian and Swiss-type cheese. *Streptococci*, *Lactococci* and *Lactobacilli* are evolutionary related bacteria and they constitute (together with, for example, the genera *Bacillus* and *Staphylococcus*) sister groups of the low GC-content branch from the gram-positive bacteria.

A large number of distinct streptococcal phages have been collected in many dairy laboratories. *Streptococcus thermophilus* phages are peculiar since all currently known isolates (>200) belong to a single DNA homology group which covers both virulent and temperate phages (Brüssow *et al.*, 1994a; Brüssow *et al.*, 1994b). Biochemical analysis defined two subgroups in *Streptococcus thermophilus* bacteriophages, *cos*-site and *pac*-site phages (Le Marrec *et al.*, 1997). At least one representative of each group of *S. thermophilus* phages (Sfi19/Sfi21 virulent/temperate *cos*-site phage, 01205/Sfi11 virulent/temperate *pac*-site phage) has been entirely sequenced (Desiere *et al.*, 1999; Lucchini *et al.*, 1999a; Lucchini *et al.*, 1999b; Stanley *et al.*, 1997; Tremblay and Moineau, 1999). Comparative sequence analysis revealed the basic mechanisms that create diversity in this group of bacteriophages (Lucchini *et al.*, 1999b). These processes included modular exchanges of large functional groups of genes (morphogenesis gene cluster) (Lucchini *et al.*, 1998), of individual genes or even of part of genes encoding putative single protein domains (e.g. repressor, anti-repressor, anti-receptor) (Lucchini *et al.*, 1999). Independent of these processes point mutations contributed to the diversification of the phage genomes, but the degree of accumulation of point mutations differed significantly for distinct genome regions of the phages.

¹ To whom correspondence should be addressed: Harald.Bruessow@rdls.nestle.com
The structural proteins from *S. thermophilus* phages showed sequence similarity only to phages infecting gram-positive bacteria. A hierarchy of sequence relatedness was observed that correlated approximately with the evolutionary distance between the bacterial hosts (*S. thermophilus* > *Lactococcus* > *Leuconostoc* ≈ *Bacillus* ≈ *Lactobacillus* > *Mycobacterium* ≈ *Streptomyces*) (Lucchini et al., 1998).

However, all conclusions were based on comparison using *Streptococcus thermophilus* phage genomes as reference points. It will thus be important to conduct similar comparative analyses by using other phage sequences as independent reference points. We therefore decided to sequence the structural gene cluster of a phage from *Lactobacillus johnsonii*, a species from which no phage sequences has yet been provided. In the case of lambdoid phages only two complete phage sequences are currently in the database. We analysed therefore the structural gene cluster of a lambdoid coliphage that was accidentally sequences many times in the Human genome project since it had apparently contaminated the *E. coli* cells used for cloning of human DNA at the Sanger Center. We concentrated the genome analysis on the morphogenesis genes since the phage structural genes were the most conserved segments of the phage genome.

To avoid a selection bias we screened a sequence from an ongoing *Lactobacillus johnsonii* genome sequencing project for the presence of prophages. We identified a structural gene cluster of a putative temperate *Siphoviridae* that confirmed the observation of our previous analysis of *Streptococcus thermophilus* phage sequences.

To provide a further test we screened the sequences from the human genome project at the Sanger Center for phage sequences. We reasoned that the *E. coli* cells used for cloning the eukaryotic DNA will be occasionally contaminated with phage sequences from temperate phages of the environment. We identified indeed a phage sequence with sequence relationships to several phage and prophage sequences from Gram-negative bacterial hosts. We analyzed both new phage sequences to address the question whether phages from both Gram-positive and Gram-negative hosts follow comparable rules in their modular organization.

**RESULTS**

**Sequence similarities**

Bioinformatics analysis of the L965 sequence allowed the establishment of a partial tentative genetic map for this prophage sequence. The likely DNA packaging and morphogenesis genes from the *L. johnsonii* prophage LJ965 comprised 23 open reading frames (orfs) all encoded on the same strand followed by a lysis cassette consisting of a putative holin and lysin gene (Fig. 1). Over the seventeen leftmost open reading frames (orfs) the predicted LJ965 gene map could be superposed on the gene map of the *pac-site S. thermophilus* phages Sti11 resulting in one-by-one correspondence of similar-sized genes. Sequence similarity was, however, limited to 5 genes from the DNA packaging and head morphogenesis module.
Figure 1: Gene map of *L. johnsonii* prophage LJ965 and comparison to the gene maps of *S. thermophilus* bacteriophage Sfi11, *Lb. delbrückii* phage LL-H, *Lb. plantarum* phage phig1e and *B. subtilis* phage SPP1. Open reading frames (Orfs) are denoted according to their length which is marked below the arrows in aa. Similarities between protein sequences are indicated by shading. The degree of similarity is given in the shaded area.
Sequence similarity with phage Sfi11 was only detectable at the protein level (Table 1). Over the rightmost 7 kb, LJ965 showed a similar organisation as *Lactobacillus delbrückii* phage LL-H (Mikkonen et al., 1996), which was further supported by significant sequence similarities between 3 proteins from the two phages predicted for this region. At the left end we identified orf 285 whose gp showed significant sequence similarity to a small subunit terminase from the *Bacillus subtilis* prophage PBSX (Wood et al., 1990). The adjacent orf 345 predicted a protein with similarity to the likely large subunit terminase of *pac-site S. thermophilus* phage Sfi11 (Table 1). This region covers the lysis module and part of the preceding tail fibre module.

**Table 1: The DNA packaging and morphogenesis genes from the *Lactobacillus johnsonii* prophage LJ965.**

<table>
<thead>
<tr>
<th>Orf</th>
<th>Similarity</th>
<th>Putative Function</th>
<th>Identical aa / over aligned aa (%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>285</td>
<td>B. subtilis prophage PBSX, small subunit terminase</td>
<td>terS</td>
<td>62 / 229 (27)</td>
<td>10^-5</td>
</tr>
<tr>
<td></td>
<td>B. subtilis, xre region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>345</td>
<td>S. thermophilus phage O1205, ORF26 gp, l. sub. terminase</td>
<td>terL</td>
<td>72 / 241 (28)</td>
<td>10^-16</td>
</tr>
<tr>
<td>500</td>
<td>S. thermophilus phage Sfi11, Orf 502 gp</td>
<td>Portal</td>
<td>127 / 446 (28)</td>
<td>10^-24</td>
</tr>
<tr>
<td></td>
<td>Lactococcus lactis protein Y1901</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>360</td>
<td>S. thermophilus phage O1205, ORF 284 gp</td>
<td>Minor head protein</td>
<td>38 / 151 (21)</td>
<td>10^-7</td>
</tr>
<tr>
<td></td>
<td>S. thermophilus phage O1205, ORF 28 gp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>S. thermophilus phage O1205, ORF 29 gp</td>
<td>Scaffold</td>
<td>49 / 169 (28)</td>
<td>10^-9</td>
</tr>
<tr>
<td></td>
<td>S. thermophilus phage Sfi11, Orf 193 gp</td>
<td>protein</td>
<td>48 / 168 (28)</td>
<td>10^-9</td>
</tr>
<tr>
<td></td>
<td>Yersinia plasmid unknown</td>
<td></td>
<td>54 / 179 (30)</td>
<td>10^-10</td>
</tr>
<tr>
<td>121</td>
<td>S. thermophilus phage Sfi11, Orf 348 gp</td>
<td>Major head protein</td>
<td>88 / 311 (28)</td>
<td>10^-25</td>
</tr>
<tr>
<td></td>
<td>S. thermophilus phage O1205 gp, ORF 31, m. str. protein</td>
<td></td>
<td>87 / 311 (27)</td>
<td>10^-22</td>
</tr>
<tr>
<td></td>
<td>Streptomyces phage V WB, head protein</td>
<td></td>
<td>72 / 305 (23)</td>
<td>10^-9</td>
</tr>
<tr>
<td></td>
<td>???</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1433</td>
<td>Myosin</td>
<td></td>
<td>127 / 671 (18)</td>
<td>10^-15</td>
</tr>
<tr>
<td>482</td>
<td>S. thermophilus phage Sfi21, Orf 1560 gp</td>
<td></td>
<td>52 / 207 (44)</td>
<td>10^-38</td>
</tr>
<tr>
<td></td>
<td>L. delbrueckii phage LL-H, Orf 360 gp</td>
<td></td>
<td>102 / 257 (39)</td>
<td>10^-14</td>
</tr>
<tr>
<td></td>
<td>Lactobacillus phage g1e, Orf 1608, minor structural protein</td>
<td></td>
<td>64 / 164 (39)</td>
<td>10^-27</td>
</tr>
<tr>
<td></td>
<td>B. subtilis prophage skin, YbO protein</td>
<td></td>
<td>84 / 278 (30)</td>
<td>10^-10</td>
</tr>
<tr>
<td></td>
<td>S. thermophilus phage PBSX, Orf 1332 gp</td>
<td></td>
<td>77 / 225 (34)</td>
<td>10^-11</td>
</tr>
<tr>
<td></td>
<td>Staphylococcus aureus phage PVL, orf 16 gp</td>
<td></td>
<td>78 / 303 (25)</td>
<td>10^-11</td>
</tr>
<tr>
<td>109</td>
<td>S. thermophilus phage Sfi11, Orf 348 gp</td>
<td></td>
<td>88 / 311 (28)</td>
<td>10^-25</td>
</tr>
<tr>
<td></td>
<td>Streptomyces phage V WB, head protein</td>
<td></td>
<td>87 / 311 (27)</td>
<td>10^-22</td>
</tr>
<tr>
<td></td>
<td>Staphylococcus aureus phage PVL, orf 16 gp</td>
<td></td>
<td>72 / 305 (23)</td>
<td>10^-9</td>
</tr>
<tr>
<td></td>
<td>???</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>540</td>
<td>S. thermophilus phage Sfi11, Orf 348 gp</td>
<td></td>
<td>88 / 311 (28)</td>
<td>10^-25</td>
</tr>
<tr>
<td></td>
<td>Streptomyces phage V WB, head protein</td>
<td></td>
<td>87 / 311 (27)</td>
<td>10^-22</td>
</tr>
<tr>
<td></td>
<td>Staphylococcus aureus phage PVL, orf 16 gp</td>
<td></td>
<td>72 / 305 (23)</td>
<td>10^-9</td>
</tr>
<tr>
<td>977</td>
<td>L. delbrueckii phage LL-H, minor structural protein gp58</td>
<td></td>
<td>38 / 118</td>
<td>10^-7</td>
</tr>
<tr>
<td>86</td>
<td>S. thermophilus phage Sfi11, Orf 140 gp</td>
<td></td>
<td>32 / 100 (32)</td>
<td>10^-5</td>
</tr>
<tr>
<td>135</td>
<td>S. thermophilus phage Sfi11, Orf 140 gp</td>
<td></td>
<td>32 / 100 (32)</td>
<td>10^-5</td>
</tr>
<tr>
<td>115</td>
<td>HolinLL-H?</td>
<td>hol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>376</td>
<td>Lysin LL-H?</td>
<td>lys</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Prophage LJ965: Bioinformatics analysis**

Orf 500 gp resembled an established portal protein of *Bacillus subtilis* phage SPP1 (Alonso et al., 1997). Orf 360 gp and orf 214 gp, respectively, demonstrated sequence similarity with a likely minor head protein and a possible scaffold protein encoded at corresponding map positions in *pac-site S. thermophilus* phages (Table 1, Lucchini et al., 1998). Orf 121 gp showed no database matches, while orf 349 gp showed sequence similarity to the major head protein from *pac-site S. thermophilus* phages encoded at an identical map position. Significant sequence similarity were also detected with the major head protein from the *Streptomyces* phage VWB (Anne et al., 1995) (Table 1). *Streptomyces* belongs to the high GC content branch of gram-positive bacteria. The LJ965, *Streptococcus* and *Streptomyces* proteins were 348 or 349 aa long which is close to the length of the major head protein of phage lambda encoded at a corresponding map position (gene E, 341 aa). Interestingly, orf 349 gp from the LJ965 prophage shared detectable sequence similarity with the lambda protein (20 %
aa identity, P=0.02). In fact, a multiple alignment of the four viral head proteins was possible despite the large evolutionary distances separating their bacterial hosts (Fig. 2). Nine aa positions were absolutely identical between these phage proteins and 48 further positions were identical in at least three phages. It is notable that the P-values linking the phage sequences showed a gradient which varied according to the evolutionary distance of the bacterial hosts to *Lactobacillus*. Downstream of the putative major head gene, a very characteristic pattern of 8 relatively small genes was followed by a very large and a medium sized gene in both the LJ965 sequence and *pac*-site *S. thermophilus* phages. The similarity in codon length was striking (Fig. 1). In *pac*-site *S. thermophilus* phages, the region is predicted to cover the tail morphogenesis module (Lucchini et al., 1998; Stanley et al., 1997). Indirect support for tail morphogenesis genes in the LJ965 sequence was obtained by bioinformatic analysis using a sensitive PAM 250 scoring matrix (Altschul, 1991; Altschul, 1993; Henikoff and Henikoff, 1992). In this search off 105 gp showed weak similarity with a *B. subtilis* phage SPP1 head completion protein (gp 15 located downstream of the major head protein gp 13 from SPP1; 24 % aa identity) and a prohead core protein precursor from coliphage T4 (Keller et al., 1984) (protein GP68). In a triple alignment 39 aa positions were identical in at least two proteins (Fig. 3).
Figure 2: Multiple alignment of the major head protein from prophage LJ965, phage Stf11, bacteriophage lambda and *Streptomyces* phage VWB.
Figure 3: Multiple alignment of LJ965 orf 105 gp, *B. subtilis* bacteriophage SPP1 gp15 and *E. coli* bacteriophage T4 gp68.

The comparison with the phage lambda map suggested orf 159 as a candidate gene for the major tail protein. Interestingly, orf 159 gp showed 23 % aa identity (P=0.03) over its entire length with GP13, the major tail protein from coliphage N-15, while orf 136 gp resembled a dihydrofolate reductase from *S. pneumoniae* (26 % aa identity over 120 aa; P=10⁻⁷). In phage T4 dihydrofolate reductase appears to be part of the base plate structure (Hanggi and Zachau, 1980).

Except for similarities with myosins no links were established for orf 1434 gp which showed a strong coiled coil prediction over its N-terminal third (Figure 4). Both properties were also reported for the orf 1560 gp, a putative minor tail protein of the *cos*-site *S. thermophilus* phage Sfi21 (Desiere *et al.*, 1998). Orf 1560 is located at an identical topological position of the gene map: two genes downstream of the major tail gene. Interestingly, the adjacent orf 482 from LJ965 prophage encoded a protein with similarity to two domains (aa positions 880 to 1080 and 1290 to 1500) from orf 1560 gp of phage Sfi21. The similarity with the second domain was highly significant (Table 1). The latter domain was also found in two *Lactobacillus* phages, two *B. subtilis* prophages and a *Staphylococcus* phage protein (Table 1). Orf 540 gp showed borderline similarity with a *Yersinia pestis* prophage protein corresponding to the phage lambda J protein (20% identity over 135 aa; P=0.003).

Orf 977 gp gave a significant match with a likely minor tail protein from *Lactobacillus* phage LL-H (Table 1). Orf 86 resembled an orf from *S. pneumoniae* phage Dp-1 (30 % aa identity) located four genes upstream of the lysin gene (Sheehan *et al.*, 1997). Orf 135 finally showed significant sequence similarity with a LL-H gene located three genes upstream of the lysin gene (Table 1).

Further pac-site *Lactobacillus* phages have been sequenced from *Lactobacillus delbrückii* (virulent phage LL-H) (Mikkonen *et al.*, 1996) and *Lactobacillus plantarum* (temperate phage phi-g1e) (Kodaira *et al.*, 1997). Over the leftmost 14 orfs both
phages showed a coparallel gene map. The relatedness of both phages is documented by numerous similarities between the predicted proteins (24-46% aa identity) (Figure 1). Over this region, no similarity with the *L. johnsonii* prophage LJ965 was detected at the sequence level although all three phages showed a clearly similar gene map. Starting with orf 1608 in phig1e, all three *Lactobacillus* phages diverged in their predicted gene maps for 8 genes. Sequence similarity between all 3 phages was limited to the C-terminal domain of the orf 1608 gp of phig1e. A putative holin/lysin gene was located downstream of this region, but only the LL-H and LJ965 sequences were sequence related (Fig. 1).

*Lactobacillus plantarum* phage phig1e and *Bacillus subtilis* phage SPP1 showed related terminase proteins (small and large subunit), scaffold and major head protein.

The SC3 sequence

A 24-kb long contig, retrieved from the human genome project at the Sanger Centre, demonstrated sequence similarity with the genome structure of a number of phage or rearranged prophage sequences from Gram-negative bacteria (Fig. 4). Over the left five orfs the SC3 sequence resembled the gene map of the *Salmonella* phage P22. This resemblance was underlined by significant sequence similarity of *orf 182* with a small terminase subunit of *Salmonella* phage LP-7. Interestingly the adjacent *orf 527* predicted a protein with weak similarity to a large subunit terminase (Table 2).
Figure 5: Gene map of SC3 prophage and comparison to the gene maps of bacteriophage N15, phage P22, *Haemophilus influenza* prophage flu and *Yersinia pestis* prophage Yp96. Open reading frames (Orfs) are denoted according to their length which is marked below the arrows in aa. Similarities between protein sequences are indicated by shading. The degree of similarity is given in the shaded area. The GC-content of Yp96 is given in the bottom graph.
Table 2: The packaging and head and tail morphogenesis gene cluster in phage SC3.

<table>
<thead>
<tr>
<th>Orf</th>
<th>Similarity</th>
<th>Putative Function</th>
<th>Identical aa / over aa (%)</th>
<th>BLAST-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>182</td>
<td>DNA packaging protein GP3 (terminase)</td>
<td>terS</td>
<td>37 / 104 (35)</td>
<td>72</td>
<td>10^-12</td>
</tr>
<tr>
<td></td>
<td><em>Salmonella</em> phage LP-7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>527</td>
<td>Large terminase subunit /</td>
<td>terL</td>
<td>51 / 214 (23)</td>
<td>39</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td><em>Methanobacterium</em> phage psiM2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>427</td>
<td><em>Haemophilus influenzae</em> protein HI 1409</td>
<td>Portal</td>
<td>90 / 411 (21)</td>
<td>94</td>
<td>10^-18</td>
</tr>
<tr>
<td>253</td>
<td><em>Haemophilus influenzae</em> protein HI 1407</td>
<td>Scaffold</td>
<td>22 / 65 (33)</td>
<td>47</td>
<td>10^-4</td>
</tr>
<tr>
<td>370</td>
<td><em>Haemophilus influenzae</em> protein HI 1405</td>
<td>Major head</td>
<td>83 / 342 (24)</td>
<td>42</td>
<td>0.005</td>
</tr>
<tr>
<td>255</td>
<td><em>Yersinia pestis</em> 100kb plasmid pMT1</td>
<td>Tail fiber</td>
<td>29 / 74 (39)</td>
<td>44</td>
<td>0.001</td>
</tr>
<tr>
<td>957</td>
<td>N15 gp16</td>
<td>Tail component</td>
<td>230 / 727 (31)</td>
<td>353</td>
<td>10^-6</td>
</tr>
<tr>
<td></td>
<td>λ minor tail protein precursor H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Integrin homolog <em>S. cerevisiae</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Myosin squid (<em>Loligo</em>)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Salmonella</em> prophage H homolog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>N15 gp17</td>
<td>Tail component</td>
<td>39 / 115 (33)</td>
<td>68</td>
<td>10^-11</td>
</tr>
<tr>
<td></td>
<td><em>Salmonella</em> prophage M homolog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ minor tail protein M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>270</td>
<td>N15 gp18</td>
<td>Tail component</td>
<td>115 / 244 (47)</td>
<td>199</td>
<td>10^-5</td>
</tr>
<tr>
<td></td>
<td><em>Yersinia</em> plasmid L homolog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ minor tail protein L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Salmonella</em> prophage L homolog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>244</td>
<td>N15 gp19</td>
<td>Tail component</td>
<td>112 / 236 (47)</td>
<td>217</td>
<td>10^-56</td>
</tr>
<tr>
<td></td>
<td><em>Yersinia</em> plasmid K homolog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coxiella unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ tail assembly protein K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>N15 gp20</td>
<td>Tail component</td>
<td>77 / 191 (40)</td>
<td>131</td>
<td>10^-30</td>
</tr>
<tr>
<td></td>
<td>λ tail assembly protein I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Yersinia</em> plasmid unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1139</td>
<td>N15 gp21</td>
<td>Tail host specificity</td>
<td>463 / 1136 (41)</td>
<td>869</td>
<td>10^-100</td>
</tr>
<tr>
<td></td>
<td>Neurofilament squid (<em>Loligo</em>)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ host specificity protein J</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Yersinia</em> plasmid J homolog</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>N15 gp22</td>
<td></td>
<td>57 / 99 (57)</td>
<td>116</td>
<td>10^-25</td>
</tr>
<tr>
<td>229</td>
<td>N15 gp23</td>
<td></td>
<td>121 / 229 (52)</td>
<td>229</td>
<td>10^-56</td>
</tr>
<tr>
<td>75</td>
<td>N15 Cor protein</td>
<td></td>
<td>27 / 76 (35)</td>
<td>53</td>
<td>10^-7</td>
</tr>
</tbody>
</table>

The three following orfs 427, 253 and 370 showed similarity with three adjacent genes from *Haemophilus influenzae*. These *H. influenzae* genes belong to the recently described cryptic prophage dflu (Hendrix et al., 1999). The three dflu genes were also preceded by genes with a weak similarity to a phage small terminase subunit and significant similarity to gene 15 from phage P22 (P=10^-13) located directly downstream of the P22 terminase genes (Fig. 4). No database match was obtained for orf 158 gp. Orf 255 gp resembled orf 265 gp from the 100 kb plasmid pMT1 in *Yersinia pestis* (GenBank accession number: AAC13172, CDS 13, 19030..19776) (Table 2). This plasmid contains a number of genes with similarity to tail genes from lambdoid coliphages (see below). The *Yersinia* orf 265 gp itself resembled the major tail protein from coliphages N15, λ, and BF23 (all P = 10^-4). For the next nine orfs (orf 957 to orf 75) the predicted gene map from the SC3 sequence could be aligned with the gene map of phage lambda (Fig. 4) and six predicted proteins shared sequence similarity with phage λ proteins H, M, L, K, I and J (Table 2).
Table 3: Similarities between open reading frames from SC3, N15 and lambda.

<table>
<thead>
<tr>
<th>SC3 orf / λ (N15)</th>
<th>Putative Function</th>
<th>SC3</th>
<th>N15</th>
<th>lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orf 957 / H (gp16)</td>
<td>Tail component</td>
<td>N15: 230/727 (31)</td>
<td>λ: 491/854 (57)</td>
<td>N15: 491/854 (57)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 230/727 (31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 218/764 (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>St: 41/139 (29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 39/115 (33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 25/63 (39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>St: 63/109 (57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orf 270 / L (gp18)</td>
<td>Tail component</td>
<td>N15: 115/244 (47)</td>
<td>λ: 57/227 (25)</td>
<td>Yp: 40/112 (35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 115/244 (47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 61/242 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yp: 82/229 (37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orf 244 / K (gp19)</td>
<td>Tail component</td>
<td>N15: 112/236 (47)</td>
<td>λ: 77/235 (30)</td>
<td>Mt: 32/82 (39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 112/236 (47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 48/183 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>St: 65/199 (32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 77/191 (40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 68/195 (34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yp: 54/179 (30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orf 1139 / J (gp21)</td>
<td>Tail host specificity</td>
<td>N15: 469/1136 (41)</td>
<td>λ: 298/926 (32)</td>
<td>Yp: 276/905 (30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 469/1136 (41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 54/191 (33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yp: 60/186 (32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orf 110 / (gp22)</td>
<td></td>
<td>N15: 57/99 (57)</td>
<td>λ: 298/926 (32)</td>
<td>Yp: 18/36 (50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 57/99 (57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 28/69/91 (27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>squid: 40/554 (73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orf 229 / (gp23)</td>
<td></td>
<td>N15: 121/229 (52)</td>
<td>λ: 27/76 (35)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SC3: 121/229 (52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>λ: 26/74 (39)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


The proteins corresponded to three minor tail proteins (H, M, L), two tail assembly proteins (K, I) and the host specificity protein (J). An even closer match was seen between the SC3 sequence and coliphage N15. The region of similarity extended to nine genes (gene 16 to gene 24 from phage N15) and the degree of similarity was higher (maximal aa identity: 57 % with N15 vs. 32 % with λ). In fact, the two phage genomes could be aligned at the nucleotide level (data not shown).

Protein H, M and L homologs from SC3 sequence demonstrated, in addition, sequence similarity with a likely Salmonella prophage and the protein L, K, I and J homologs from SC3 showed similarity with proteins encoded by a 100 kb plasmid from Yersinia pestis (see above). Table 2 provides a list of all database matches for the predicted tail genes shared between SC3, N15 and λ: for protein H, phage λ and N15 were the closest relatives, while for proteins M and L, phage λ and a Salmonella prophage were the most related. For protein K, I, J homologs and the three orfs downstream of gene J phage N15 and SC3 proteins were the closely related, the aa sequence identity reached 57 %.

Surprisingly, the closest match for a SC3 sequence was not with another coliphage, but with a purported neurofilament-like protein from the Northern European squid (Loligo forbesi) (Adjaye et al., 1993). The protein J homolog from SC3 shared 41 and 32 % aa identity over its entire length with the corresponding N15 and λ proteins, respectively, while it shared 73 % aa identity with the "squid" protein. The alleged squid protein corresponded to a 500 aa long segment in the N-terminal half of the J homolog. Alignment of the SC3 and the "squid" protein revealed segments of near aa identity separated by 3 regions of sequence diversity. Alignment of the SC3 sequence with the coliphage and Yersinia proteins showed in contrast a homogeneous degree of similarity (41 % aa identity, p-value 10-173) (Fig. 6). The "squid" protein, but none of the prokaryotic J homologs showed a consensus motif found in intermediate filament (IF) proteins: the TYRKLLEGEE consensus sequence is found at the C-terminal end of
the predicted rod structure of most IF proteins and represents a conserved epitope in IF. However, while an authentic squid neurofilament and a nematode intermediate filament demonstrated a clear coiled-coil prediction over the putative rod region, no coiled-coil was predicted for the J homologs of the bacteriophages or the purported squid protein (data not shown).

41.2% identity in 927 residues overlap

Yp96

98
DISCUSSION

The experimental basis for the modular theory were heteroduplex studies of lambdoid phages in the late 1960's. Until now only two complete sequences are available for lambdoid coliphages from the database (λ and N15) but their sequence comparison has not yet been reported in the scientific literature. Both phages shared a conserved structural gene cluster (DNA packaging, head and tail morphogenesis modules), but differed for the rest of the genome. According to heteroduplex studies the structural gene cluster from lambdoid phages is relatively conserved, only two distinct forms, represented by λ and HK97, have been described. For phage HK97 only 3 genes have until now been submitted to the database. Partial sequence information for structural genes is available for coliphage 21.

In the current report we tried to address the question whether the evolution of phage DNA sequences followed different rules in gram-positive and gram-negative bacteria. To this purpose we analyzed the structural gene cluster from two new phage DNA sequences. One sequence represents a prophage sequence from a further Lactobacillus species, L. johnsonii, for which no phage sequences are yet available. The other sequence corresponds to the structural gene cluster of an unidentified lambdoid phage that apparently infected the E. coli strain used for cloning of human genome DNA in a large sequencing project.

Over the left two thirds of the partial gene map the L. johnsonii prophage LJ965 was strikingly similar to that of pac-site S. thermophilus phage Sfi11. Over the right third of the partial LJ965 prophage map similarity with part of the tail gene cluster of L. delbrueckii phage LL-H was detected. Sequence similarity with S. thermophilus phages was limited to the putative head morphogenesis genes and part of the DNA packaging region. The region covering possible tail genes showed a comparable gene map, but no sequence similarity. Previously, we have demonstrated a similar one-by-one gene correspondence between pac-site S. thermophilus phage Sfi11 and the Lactococcus lactis phage TP901-1 (Lucchini et al., 1998). TP901-1 and Sfi11 showed sequence similarity for the major tail gene and four genes upstream and two genes downstream of the tail gene. As already seen in previous bioinformatic analysis of S. thermophilus phages, the genomes from Siphoviridae of lactic acid bacteria can apparently be subdivided into groups of associated genes. The groups of associated genes did not correspond to functional genetic units. For example, the likely small and large subunit terminase from prophage LJ965 showed similarity to a B. subtilis and a S. thermophilus phage protein, respectively. A likely minor head gene from LJ965 showed no similarity to the corresponding Sfi11 gene, while genes to the left and the right were related to Sfi11 genes. The LJ965 and TP901-1 similarity regions in Sfi11 were separated by a gene showing weak links only to a B. subtilis phage protein.
In the L965/Sfi11 and the TP901-1/Sfi11 comparison the sequence similarity was limited to the aa level and did not exceed 38%. If the similarities between the compared phage genomes reflect horizontal gene transfer events, then they must have occurred in a relatively distant past, allowing for a substantial accumulation of point mutations and insertion and deletion events in divergent lines of phages that ceased to exchange genetic material. Unfortunately, we do not possess a molecular clock for phage DNA sequences. We can thus not estimate what time periods separate the related Sfi11/LJ965 or Sfi11/TP901-1 sequences.

A previous analysis of *S. thermophilus* phage genomes suggested a hierarchy of sequence relationships, where the degree of sequence relationships between phages correlated approximately with the evolutionary distance that separated their bacterial hosts. On the basis of these data, one would have predicted that the closest relative of the prophage LJ965 is another *Lactobacillus* phage. This was not the case. This observation is, however, not at odds with our previous report. First, too few *Lactobacillus* phage sequences are in the database. Second, graded sequence similarities were also observed in *Lactobacillus* phages. Sequence similarity over structural genes that extended to the nucleotide level were observed between phages LL-H and mv4 (84 to 91% bp identity) which infect the same host species, *L. delbrueckii*. *Lactobacillus* phages that infect different host species like phages LL-H and phig1e shared still substantial sequence relationship (Fig. 7). However, sequence similarity was only detectable at the aa level and did not exceed 49%. In addition, the degree of sequence similarity between phig1e and LL-H proteins was variable: genes showing no or low similarity were interspersed with genes showing significant similarity (Fig. 7).

Apparently, prophage LJ965 defined a second structural gene cluster in temperate *Lactobacillus* phages. It is unknown how many different structural gene clusters exist for temperate phages in this bacterial genus. Partial DNA sequences are available for the DNA packaging region of *L. casei* phage A2. A2 lacked sequence similarity with other *Lactobacillus* phages, but interestingly shared sequence similarity with cos-site *S. thermophilus* phages at the aa sequence level.

The standard theory of phage evolution, the modular theory, has been developed on the basis of heteroduplex mapping of lambdoid coliphages in the pre-DNA sequencing era. For a comparative evolutionary analysis it would be desirable to have more sequence information for lambdoid phages than currently available.

We reasoned that large sequencing projects that use *E. coli* during the cloning their target DNA should accidentally sequence phage DNA derived from the phages in the laboratory environment. Therefore we screened the database of the human genome project at the Sanger Centre for phage-like DNA sequences. Three such putative phage sequences were retrieved from the human genome project at the Sanger Center. We were unable to trace the contamination from the environmental source to the sequenced clones. One sequence SC3 was obtained so frequently that an electronic filter against phage sequences was introduced at the Sanger Center. The
largest contig corresponded precisely to the structural gene cluster of a phage lambda-like sequence. Over 6 putative tail-genes (tape-measure gene H to adhesion gene J) the similarity with phage lambda was both in gene order and in sequence (25 to 30 % aa identity) However, similarity of the SC3 sequence was more pronounced with E. coli phage N15: it extended over 9 adjacent genes and the sequence similarity was higher (31 to 57 %).

The genetic organization of the SC3 sequence resembled that of lambdoid phages over the DNA packaging and morphogenesis genes. Sequence similarity that extended to the bp level was detected with coliphage N15, similarity with phage λ proteins H to J was limited to the aa level. Upstream of the gene H homolog, the SC3 sequence showed no similarity to proteins from lambdoid coliphages, while similarities to a Salmonella phage and likely prophage proteins from Haemophilus or Yersinia were detected. Currently only three distinct alleles exemplified by coliphages λ and HK97 and Salmonella phage P22, have been described for the morphogenesis module of lambdoid phages. Few if any modular exchanges have been reported for the structural gene cluster of the lambdoid phages.

At the sequence level, SC3 appears to be the result of a recombination event between an as yet undefined phage contributing DNA packaging, head and the proximal part of the tail genes and a N15-like phage contributing the distal tail genes starting with gene H. Interestingly, the comparison of the λ and N15 sequences suggested a complementary exchange reaction: the dot plot demonstrated bp similarity over the leftmost 12 kb of the two genomes and diversification starting with the H gene. Beyond the H gene the phage λ and N15 genes demonstrated only 30 to 36 % identity at the aa level until gene J, thereafter both phage genomes lost any sequence similarity. A much higher level of identity was found for the protein A to T homologs (60 to 94 % aa identity, two exceptions). Interestingly, prophage sequences found on a plasmid of Yersinia pestis showed a comparable level of aa identity. We are not aware of a situation in dairy phage where the range of differences between related phages infecting the same species overlapped the range of differences between phages infecting different bacterial genera. A possible interpretation are more efficient ways for horizontal phage DNA transfer between different genera of Enterobacteriaceae than lactic acid bacteria (e.g. bacteriophage P1 and P2). Currently an increasing number of complete phage genome sequences become available. It will therefore soon be possible to decide whether the observations of graded relatedness made in phages from lactic acid bacteria are peculiar to this group of bacteria.

The protein J homolog from SC3 demonstrated 73 % aa identity over the entire length of a purported squid protein of neuronal origin (Adjaye et al., 1993). The authors screened a cDNA library of the optical lobes from squid with antibodies against purified squid neurofilaments. Hybridization analysis showed that the positive clones correspond to a single gene that gives rise to a transcript of 2.6 kb. An antibody raised against this protein stained specifically neuronal tissue from the optical lobe of the squid, cross-reacted with neurofilaments of a marine worm, but failed to react with vertebrate neurofilaments. It is unlikely that this protein represents an authentic squid protein. The purported squid protein, but not the predicted SC3 homolog shared with an authentic squid neurofilament, an Ascaris and several mammalian intermediate filament a very conserved motif at the C-terminus of the rod region. We propose that this shared epitope is responsible for the two-way immunological cross-reactions. The cross-hybridization of the cloned DNA with squid DNA and mRNA is more difficult to explain. However, the cDNA contained in addition to the open reading frame 1.8 kb non-coding DNA. This DNA showed no similarity with phage DNA and could be responsible for the hybridization signal. We propose that during cloning of squid cDNA
into λ.gt11 a recombination event took place between the lambda cloning vector and a SC3-like coliphage from the environment (the cloning site is near to the J gene). If this interpretation is correct, SC3-like phages might be relatively common environmental contaminants.

MATERIAL AND METHODS

Sequence analysis was done using the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711. Additional database searches have been conducted using the BLAST server at NCBI (Altschul et al., 1997) and FASTA (Lipman and Pearson, 1985). Sequence alignments were done using the MultAlin program (Corpet, 1988) and the SIM alignment tool (Huang et al., 1990). The sequences can be retrieved under the GenBank accession numbers, xxxxxx (LJ965), A (SC3) and A (Yp96). These sequence data were produced from the Y. pestis genome sequencing project at the Sanger Centre and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/yp/.

ACKNOWLEDGEMENTS

We thank the Swiss National Science Foundation for financial support of Frank Desiere in the framework of its Biotechnology Module (Grant 5002-044545/1, AU-054586).
INTRODUCTION

(Hendrix et al., 1999; Hendrix et al., 1999) The question about the relationship between bacteriophages can now be addressed on the molecular level thanks to the availability of nearly 35 completely sequenced bacteriophage genomes. Full genome comparisons are however rare and were done only to compare bacteriophages infecting the same host species. The recent report on the comparison of the genome structure of Mycobacteriophage D29 with Mycobacteriophage L5 (Ford et al., 1998) yielded important information about phage evolution within the same genus. Extensive sequence similarity has been observed and the comparison of the two genomes showed that they are punctuated by a large number of insertions, deletions and substitutions of genes which were consistent with the genetic mosaicism that was described for the lambdoid phages (Campbell, 1994). This comparison has extended some of the lambdoid evolution principles to the unrelated group of viruses, the mycobacteriophages. The comparative analysis of five completely sequenced Streptococcus thermophilus bacteriophage genomes (Lucchini et al., 1999b) demonstrated that their diversification was achieved by insertion/deletion, recombination and accumulation of point mutations. Both investigation showed the close relationship between virulent and temperate phages.

For an understanding of the evolutionary relationship between bacteriophages an analysis of sequence relationships between phages infecting different bacterial genera will be crucial. In the conserved replication module of Sfi21 we found previously high amino-acid (aa) sequence similarities to phage BK5-T in several genes and the topological gene constellation seemed to be highly conserved (Desiere et al., 1997). We have furthermore found previously a hierarchy of sequence relatedness of genes that correlated approximately with the evolutionary distance between the bacterial hosts (S. thermophilus >> Lactococcus > Leuconostoc = Bacillus = Lactobacillus >> Mycobacterium= Streptomyces) (Desiere et al., 1999). The observation suggested that temperate Siphoviridae from low GC content Gram positive bacteria have evolved from
a common ancestor. To test this hypothesis a detailed comparison of the phages at the DNA sequence level is essential. Here we compare the complete genomes of lactococcal phage BK5-T (Boyce et al., 1995), Streptococcus thermophilus bacteriophage Sfi21 (Desiere et al., 1999) and Staphylococcus aureus phage PVL (Kaneko et al., 1998).

The Lactococcus lactis temperate bacteriophage BK5-T is a small isometric-headed temperate phage that infects Lactococcus lactis subsp. cremoris. It is a type phage in the lactococcal phage classification (Jarvis et al., 1991) and it carries cohesive ends.

**RESULTS**

The three bacteriophages have a comparable genome size and their average GC content are consistent with that of the corresponding hosts (Table 1). Sfi21 possesses 15 nucleotide long cohesive ends with a 3' overhang that reconstitutes a cos-site (CCGCCACAAAGGTGTC) with two-fold hyphenated rotational symmetry (Desiere et al., 1999), bacteriophage PVL carries 3' staggered cohesive ends with 9 bases (CGGAGCAGA) and BK5-T has been reported to possess a cos-site as well (Mahanivong et al., 1999).

<table>
<thead>
<tr>
<th>Phage</th>
<th>Genome size [bp]</th>
<th>G+C content Phage/Host [%]</th>
<th>Orfs</th>
<th>Access. Number</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sfi21</td>
<td>40,739</td>
<td>37.5 / 38.1*</td>
<td>53</td>
<td>AF115103</td>
<td>(Desiere et al., 1999)</td>
</tr>
<tr>
<td>BK5-T</td>
<td>40,003</td>
<td>34.9 / 36.4</td>
<td>53</td>
<td></td>
<td>(Mahanivong et al., 1999)</td>
</tr>
<tr>
<td>PVL</td>
<td>41,401</td>
<td>33.5 / 32.9</td>
<td>63</td>
<td>AB009866</td>
<td>(Kaneko et al., 1998)</td>
</tr>
</tbody>
</table>

Fig. 1: Gene maps of *Streptococcus thermophilus* bacteriophage Sfi21, *Lactococcus lactis* bacteriophage BK5-T and *Staphylococcus aureus* bacteriophage PVL. The open reading frames are indicated as arrows according to their orientation and marked by their length in amino acids. Probable functions identified by bioinformatic analysis are indicated. For Sfi21, the phage genome is divided in different functional genome regions according to previous bioinformatic and comparative evolutionary analysis. White color indicates lack of information about the function of the orf. Putative rho-independant terminators are indicated with an hairpin, and overlaps between start and stop codons are indicated with a triangle. Similarities to *L. lactis* bacteriophage r11t are indicated with black bars. Aa similarities between orfs of the phages are connected with shading.
The comparison of the three phage genomes shows that they are similarly organized with respect to their modular structure (Figure 1).

The DNA packaging and morphogenesis modules cover about half of the genomes (20 kb). Over the leftmost 15 kb from both Sfi21 and BK5-T, the gene organization and the number of genes is extremely similar and the size of the orfs and their position are almost identical. Database matches were found to all but 3 for the 16 genes which are located in this genome region. For 8 of these we could predict a putative gene function according to bioinformatics analysis (Figure 1). At the aa level, all the BK5-T proteins from the small subunit terminase to orf 1904 gp shared detectable aa similarities ranging from 27 % for BK5-T orf 548 gene product (gp) to 62 % for the large subunit of the terminase of both phages (Table 2).

Table 2: Amino acid sequence similarities of the genes of bacteriophage BK5-T

<table>
<thead>
<tr>
<th>BK5-T orf</th>
<th>Best database match</th>
<th>Sfi21 orf</th>
<th>Putative function</th>
<th>P-value (aa sim.)</th>
<th>Similarity with PVL</th>
</tr>
</thead>
<tbody>
<tr>
<td>157</td>
<td>Lc. lactis cremoris orf4 24/59 (40%)</td>
<td>152</td>
<td>Terminase s sub</td>
<td>3e-28 (44%)</td>
<td></td>
</tr>
<tr>
<td>660</td>
<td>gp5-like Lc. lactis cremoris (74%)</td>
<td>623</td>
<td>Terminase l sub.</td>
<td>0.0 (62%)</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>portal gp3 phage HK9750/278 (17%)</td>
<td>384</td>
<td>Portal protein</td>
<td>e-122 (55%)</td>
<td></td>
</tr>
<tr>
<td>237</td>
<td>Clp protease Streptomyces coelicolor</td>
<td>221</td>
<td>Protease</td>
<td>2e-57 (51%)</td>
<td></td>
</tr>
<tr>
<td>404</td>
<td>-</td>
<td>397</td>
<td>Major head pro.</td>
<td>e-126 (59%)</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>orf K phage L10 5e-05, 23/80 (28%)</td>
<td>106</td>
<td>-</td>
<td>0.001 (39%)</td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>orf A phage L10 1e-19, 55/139 (30%)</td>
<td>141</td>
<td>-</td>
<td>6e-09 (26%)</td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>orf L phage L10 1e-13, 41/110 (37%)</td>
<td>123</td>
<td>-</td>
<td>1e-13 (37%)</td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>orf E phage L10 2e-35 , 83/192 (43%)</td>
<td>203</td>
<td>Major tail protein</td>
<td>1e-34 (45%)</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>-</td>
<td>117</td>
<td>-</td>
<td>35/121 (28%)</td>
<td></td>
</tr>
<tr>
<td>163</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>Min. capsid pr. phigle 4e-69 (31%)</td>
<td>1560</td>
<td>-</td>
<td>2e-41 (42%)</td>
<td>8e-20 (22%)</td>
</tr>
<tr>
<td>548</td>
<td>O1205orf 512 3e-10, 82/372 (22%)</td>
<td>515</td>
<td>-</td>
<td>2e-34 (27%)</td>
<td>-</td>
</tr>
<tr>
<td>1904</td>
<td>O1205 orf45 4e-60, 105/143 (73%)</td>
<td>1276</td>
<td>-</td>
<td>5e-83 (37%)</td>
<td>-</td>
</tr>
<tr>
<td>78</td>
<td>-</td>
<td>117</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75</td>
<td>-</td>
<td>114</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>95</td>
<td>YAF2_BAC118e-08, 32/81 (39%)</td>
<td>87</td>
<td>Helix</td>
<td>-</td>
<td>1e-06 (37%)</td>
</tr>
<tr>
<td>259</td>
<td>(Zf3946) amidase Dp-1 2e-26</td>
<td>288</td>
<td>Lysin</td>
<td>34/118 (28%)</td>
<td></td>
</tr>
<tr>
<td>536</td>
<td>4 orfs</td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>374</td>
<td>Various integrases</td>
<td></td>
<td>Integrase</td>
<td>3e-53 (32%)</td>
<td>8e-122 (21%)</td>
</tr>
<tr>
<td>285</td>
<td>(U95827) putative extracellular protein</td>
<td>203</td>
<td>Immunity</td>
<td>0.91 (28%)</td>
<td></td>
</tr>
<tr>
<td>297</td>
<td>(L28019) repressor Tuc2009 e-115</td>
<td>1116</td>
<td>cl-repressor</td>
<td>2e-08 (25%)</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>(Y14232) hypoth. pro. Tp101-1 5e-10</td>
<td>111</td>
<td>Cro repressor</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>266</td>
<td>(AP000007) Pyrococcus horikoshii 0.008</td>
<td>1146</td>
<td>Antirepressor</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>(U38906) ORF5 r1t e-145</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>113</td>
<td>(U38906) ORF6 r1t 2e-57</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>143</td>
<td>Repl. Pro. Clostridium plasmid pCB101</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>72</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>80</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>89</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>234</td>
<td>(AF077306) gp233 Sfi21 4e-73</td>
<td>233</td>
<td>-</td>
<td>3e-77 (60%)</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>173</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>235</td>
<td>(U78608) Rep pro. Bacillus popilliae 3.9</td>
<td></td>
<td>SSB</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>133</td>
<td>(U38906) ORF14 r1t 2e-70</td>
<td></td>
<td>orf 48 0.040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>(U38906) ORF16 r1t 2e-37</td>
<td></td>
<td>dUTPase</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>(U38906) dUTPase r1t 2e-35</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>122</td>
<td>(AF011378) unknown skt 1e-52</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>57</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>133</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>172</td>
<td>(AB009866) orf63 phi PVL 3e-09</td>
<td>175</td>
<td>-</td>
<td>2e-24 (39%)</td>
<td>8e-08 (28%)</td>
</tr>
</tbody>
</table>

Only BK5-T orf 163 represents an exception in not having any corresponding aa sequence in Sfi21 and no match in the database. Orf 163 is at the position of the phage
lambda T gene, which encodes a protein that participates in the initiation of the morphogenesis of the tail structure. This protein seems to lack in Sfi21 and PVL. In some lambdoid phages however, the G/T proteins seem to be linked by translational frameshifting. Other regions without aa sequence similarity are more localized to an extremely variable region in the center of orf 1560 gp and the C-terminus of orf BK5-T orf 1904 gp.

The next functional entity is the lysis region with the holin and lysin function. In Sfi21 and in other Streptococcus thermophilus phages there are two holins preceding the lysin gene. One holin shows a dual start site which possibly allows the balanced expression of a holin and an anti-holin from the same gene. In BK5-T, however, we could only a single holin. The similarity between the Sfi21 and the BK5-T is restricted to the N-terminal half of the lysin, which also shows database similarities to the N-terminus of the Streptococcus pneumoniae bacteriophage Dp-1 lysin (Lopez et al., 1992) Bacillus subtilis yzkA (similar to cell wall-binding protein). The C-terminal part of the protein orf 259 gp of phage BK5-T shows no significant database matches. These findings confirm the modular character of the lysis enzymes as previously described for pneumococcal phages (Sheehan et al., 1997).

The lysogeny module is located in all three phages between the lysis and the DNA replication module. For Sfi21 the following gene order was established: int (integrase), imm (superinfection immunity), mpm (metalloproteinase motif protein), cl (phage repressor) and in opposite orientation cro (phage repressor), ant (anti-repressor). A roughly comparable gene order was also observed in BK5-T and PVL; but only two proteins showed sequence similarities between Sfi21 and BK5-T. A detailed comparison of the lysogeny modules of numerous bacteriophages have been reported recently (Lucchini et al., 1999). The DNA region between the lysis and the lysogeny module seems to be a stuffer region containing non-essential genes. The 3 investigated phages showed no similarity over this region. The first 4 genes of the lysogeny module of Sfi21 are replaced in BK5-T with a single orf (orf 536) which shows low homology to recombination/mobilization protein from Staphylococcus cohnii. In PVL this location is occupied with the Panton-Valentine leukocidin genes LukF and LukS, which are dermonecrotic and leukocytolytic toxins partly responsible for the virulence of Staphylococcus aureus (Szmigielski et al., 1999).

The DNA replication module is the next functional unit, which covers approximately 10 kbp followed by an unattributed gene region (possibly involved in transcriptional control) next to the cos-site. The different replication modules of the respective phages have about the same size. However, Sfi21 and BK5-T showed a distinct replication strategy except for sharing two genes, orf 157 and orf 233. Sfi21 possesses a replication module with a putative primase and helicase. The primase is located next to the origin of replication (Foley et al., 1998). In the replication module of BK5-T and PVL we find a single-stranded-binding protein (SSB) and an undefined replication protein and further downstream a dUTPase. Actually this alternate replication function is similar to the one of L. lactis bacteriophage r11 (van Sinderen et al., 1996) which is underlined by several other orfs showing matches to genes from bacteriophage r11 (see Figure 1).

Comparison of the Sfi21 and BK5-T bacteriophage sequences using the SIM program revealed astonishingly high similarities at the nucleotide level. Over about 10 kb covering the packaging and morphogenesis region (Figure 2) nucleotide identity was as high as 60 % for the putative small subunit terminases (BK5-T orf 157, Sfi21 orf 152) and the putative major head proteins (orf 404, 397). The subsequent 4 orf were less conserved and showed only similarity at the aa level. The major tail protein showed again 60 % similarity on the nucleotide level.
When the complete genome was aligned using the Dotplot program, further identities were identified between Sfi21 and BK5-T (Figure 3). Similarities were observed in the putative tail fiber gene orf 1276 (98/119 nt, 82%) (Lucchini et al., 1998) the lysis gene, in the integrase gene, the anti-repressor and three downstream genes. A comparison of the nucleotide sequences of bacteriophages BK5-T or Sfi21 with bacteriophage PVL revealed in contrast no similarities at the nt level.

Comparison of the BK5-T genome to other lactococcal phages revealed only very few sequence similarities on the nucleotide level. Bacteriophage rlt shared only one segment with BK5-T in the right genome region covering the complete BK5-T orf 122 (308/345 nt, 89%) (left figure). Bacteriophage skl in contrast shared more DNA segments with BK5-T (right picture). In the lysogeny module the integrase showed sequence similarity of 96%. The similarities went further through to the anti-repressor (97% similarity). More similarities were observed in the right genome region (orf 133 to orf 121). Phage bl170 (not shown) showed the same similarity as phage rlt.

Figure 2: SIM alignment of the first 10 kb of the genomes of Sfi21 and BK5-T on the nucleotide level. Arrows give the size (in aa) and orientation of open reading frames. Their putative function is indicated where determined previously due to bioinformatics analysis.
Fig. 3: Dot plot calculated from the DNA genome sequence of Sfi21 (x-axis) and BK5-T (y-axis). The comparison window was 25 bp and the stringency 16 bp. The colored gene maps are given on the respective axis. Regions of sequence similarities are marked.
DISCUSSION

We have compared the sequences of Streptococcus thermophilus bacteriophage Sfi21 with Lactococcus lactis spp. cremoris bacteriophage BK5-T and Staphylococcus aureus phage PVL. The three bacteriophages grow on hosts belonging to different bacterial genera and no phages are known to cross the bacterial genus barrier. Remarkably, 21 out of 50 proteins shared aa similarity with Sfi21 while 9 proteins shared similarities with PVL genes. The similarities between BK5-T and Sfi21 extended to the nt level. Such a close relationship between bacteriophages infecting two different genera has to our knowledge not yet been described in the literature.

The mosaic character of these similarities, where regions of similarity are interspersed with regions with non-similarity, are compatible with the modular theory of phage evolution as stated by Botstein (Botstein, 1980) and suggest recombinational events as a major mechanism in bacteriophage evolution. However, the relatively low degree of nt sequence identity between BK5-T and Sfi21 excludes recent lateral gene exchanges. Unfortunately we do not possess a molecular clock whether for phage proteins nor reliable data when streptococci and lactococci have separated in evolution. We are therefore unable to decide whether BK5-T and Sfi21 derive from a common ancestor phage that existed before the separation of these two bacterial genera.

The most conserved part of the BK5-T and Sfi21 genome is the packaging and morphogenesis region. Apparently the selective pressure to maintain this region is bigger here than in other genome regions. The structural constraints to build a functional bacteriophage particle might be the reason for this conservation, where the driving force to maintain a common particular interrelationship between genes may come from the selective pressure to cluster genes whose products interact (Stahl and Trautner, 1966). There are however also differences within this module, e.g. the similarities on the nt level of the 4 genes between the major head and the major tail protein have exchanged through recombination with DNA from another phage during evolution but similarities on the aa level were nevertheless still clearly visible. Aa similarities can be detected over evolutionary much greater distance than nt similarities. Selective pressure can maintain a certain aa sequence while this pressure is less effective for the nt sequence because of the redundancy of the genetic code.

Insertions and/or deletions have shaped the genomes of both phages during evolution: BK5-T orf 163 is here the clearest illustration. The deletions/insertions have changed the colinearity of both phage genomes only slightly. In fact, deletions have been characterized previously in both bacteriophage Sfi21 and BK5-T. These deletions where located in the lysogeny region from Sfi21 and in the tail fiber region from BK5-T and Sfi21.

The comparison of bacteriophage genomes of different genera living in the same ecological environment, i.e. the dairy environment, are important to study bacteriophage evolution. The presented study is to our knowledge the first detailed comparison of two phages from different genera. This investigation yielded important knowledge about bacteriophage evolution in general:

- The relationship between the two phages is so close that we have to postulate a common gene-pool from which these phages were derived.
- The modular organization of the two genomes is very much conserved.
- The gene map is especially well conserved when the gene products interact with each other.
- Recombination has taken place between different phages resulting in a mosaic character of the genome.
• Different genome regions of a single phage might belong to different evolutionary lineages.

The question about common ancestry of bacteriophages has also been addressed by Hendrix (Hendrix, 1998). From investigation about virion assembly in dsDNA phages and from comparisons of genomic sequences and genome organization, it became clear that the investigated phages (HK97, P22, Lambda, T7, T4, L5, P2, phi29) shared a common ancestor, even when all evidence for similarity has been lost from the nucleotide sequences of their genomes and the aa sequences of their structural proteins. The elucidation of three-dimensional structures of phage virion proteins will be instrumental in the analysis of phylogenetic relationships between viruses that can not any longer be addressed by sequence analysis.

The classical taxonomic view of the natural hierarchical classification cannot resolve the relationship between the bacteriophages in our study. The phenomenon of lateral gene transfer (LGT) must be taken into concern in order to explain the complex interrelationship between bacteriophages. Similar observations were recently made in comparative analysis of bacterial genomes. It was stated that the evolutionary history of prokaryotes in general cannot properly be represented in a classical phylogenetic tree. Doolittle predicted that the familiar arborate (tree-like) phylogenetic scheme will be soon replaced by reticulate (net-like) representation of phylogenetic relationships (Doolittle, 1999). It has also been suggested that the numerous documented instances of LGT among bacteria and archaea change the tree of life to rather a net (Hilario and Gogarten, 1993).

MATERIAL AND METHODS

Sequence analysis was done using the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 and also the Extension EGCG Package, Peter Rice, The Sanger Center Hinxton Hall, Cambridge, CB10 1RQ, England. Additional code by Peter Rice, The Sanger Center, Hinxton, England and other members of the EGCG team. Additional database searches have been conducted using the BLAST server at NCBI (Altschul et al., 1997). The complete bacteriophage genomes can be retrieved under the GenBank accession numbers AF115103 (Sfi21), AF176025 (BK5-T) and AB009866 (PVL).

ACKNOWLEDGEMENTS

We thank the Swiss National Science Foundation for financial support of Frank Desiere in the framework of its Biotechnology Module (Grant 5002-044545/1, AU-054586).
Seite Leer / Blank leaf
12 Summary and Conclusions

12.1 Ecology of *Streptococcus thermophilus* phages

Lactic Acid Bacteria are used as starters in the fermentation of various food products (cereals, vegetables, meat, milk). Fermentation problems linked to phage attack on the starter are most prominent in the dairy industry. Since a large yogurt factory processes up to 50,000 liters of milk every day, bacteriophage attack and the possible loss of the product can be very costly. The economical implications are even more serious in cheese production where many dairies process half a million liters of milk daily. *Streptococcus thermophilus* is used to produce cheese and yogurt and the genus *Streptococcus* belongs together with other dairy starters like *Lactococcus* and *Lactobacillus* to an evolutionary related group of the low GC branch of Gram-positive Eubacteria. To elucidate the ecology of *Streptococcus thermophilus* bacteriophages and the evolution of these and related bacteriophages we conducted this study. We decided to start with an ecological study that explored the origin of bacteriophages and their diversity in a factory. We decided subsequently to characterize the genome of the temperate *Streptococcus thermophilus* phage Sfi21 in molecular detail to screen it for possible phage inhibitory genetic elements and to better understand the evolutionary history of dairy phages.

Our *S. thermophilus* phage collection covers about 30 years of factory surveys. The low frequency of phage re-isolation over longer time periods could indicate that very many distinct phage strains exist in the environment and, consequently, the probability of phage re-isolation is low (Brüssow et al., 1994a). Alternatively, a phage, defined by a specific restriction pattern, could experience within relative short time periods rearrangements of its genome by deletion, duplication and acquisition of new phage DNA modules. As we frequently observed more than one phage type in a single industrial sample, infection of a single host with two phages could give rise to recombinant phages. Superinfection of lysogenic starters with virulent phages might also be a potential source for recombinants (see 12.4).

To explore the ecology of phage infections a longitudinal survey of *S. thermophilus* phages was conducted in a mozzarella factory (see chapter 4) (Bruttin et al., 1997) which uses a complex, undefined mixture of starter cells and open, continuous fermentation tanks which are very susceptible to invasion by raw milk phages. If a single phage invades the factory and experiences genetic changes during its residence, one would predict the following population structure for the factory phages A: a closely related family of phages with similar restriction patterns and overlapping host ranges. In addition, one would expect an accumulation of point mutations or consequently changes of the restriction pattern during the survey period. Alternatively, if the diversity of phages arose before their entry into the factory, one would predict a different phage population structure B: phages with a wide range of restriction patterns, clearly distinct host ranges and no changes in nucleotide sequence during the two years of the survey.

Our factory data supported model B: we observed 12 distinct lytic groups of phages showing 11 distinct restriction patterns attacking starter cells which exhibited distinct ribotypes. All but one phage type were unknown to our phage collection. Multiple phage isolates belonging to the same lytic group showed identical restriction patterns except in two groups where restriction modification systems seemed to be involved in restriction pattern variations. A 0.5 kb phage DNA fragment from a protein-coding gene from phages belonging to 10 different lytic groups was PCR amplified and sequenced.
Independent phage isolates from the same lytic group showed no nucleotide sequence changes during the survey period.

Proof of a link between factory phages and phages from the environment is difficult, as the phage concentration in raw milk is very low, while cheese factories tend to be infested with moderate to high levels of phages ($10^4$ to $10^7$ pfu/ml). The dominant factory phage population will thus prevent the establishment and consequently detection of in the factory. Finally, one will only see phages for which susceptible cells exist in the starter cultures. To study the link between environmental phages and the factory phage population we conducted an intervention study in a mozzarella factory.

The complex, undefined starter system was replaced by a defined starter system composed of starters unable to propagate the dominant phage population from the factory. As expected this intervention led first to a marked decrease in the dominant phage population. The factory was then followed longitudinally and screened regularly for the appearance of new phages. The first phage attacking the new host cells was detected 5 days after the intervention, while the second, third and fourth phages were observed after 11, 20 and 26 days. During the next 3 months no new phages were observed. Restriction analysis demonstrated that the new phages were not derivatives of the phage population existing previously in the factory, but were identical to phages that were found at low titers (10-130 pfu/ml) in the raw milk delivered to the factory during the intervention period.

This study indicated that cheese factories are infected by phages coming from the environment, i.e. raw milk. The diversity of the phages is reflecting their natural diversity and not their evolution during factory residence.

12.2 The genome of Streptococcus thermophilus bacteriophage Sfi21

We sequenced the Streptococcus thermophilus bacteriophage Sfi21 to characterize it in molecular detail since the knowledge on complete genomes of bacteriophages is scarce, in particular for Streptococcus thermophilus bacteriophages no sequence information was available. Sfi21 has a genome of 40,739 bp in size with cohesive ends (cos). The genome consists of 53 putative orfs longer than 50 codons (see figure 1). Two of these orfs were of unusual length for prokaryotic genomes in that they exceed 1200 codons. All but for orfs started with an ATG start codon. 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. Sfi21 possesses 15-nucleotide-long cohesive ends with a 3' overhang that reconstitutes a cos-site with a hyphenated rotational symmetry (see chapter 10).

Most of the nucleotide sequence has been determined from random shot-gun pUC19 clones and non-clonable, lethal regions within the genome have been sequenced on the appropriate PCR product or by sequencing directly on the bacteriophage DNA. The complete nucleotide sequence has been deposited in GenBank under the accession number AF115103. The number one nucleotide has been arbitrarily defined as the first nucleotide on the right of the cos-site sequence, that lies close to the morphogenesis gene segment. The size of the genome is within the range of all 5 complete genomes of Streptococcus thermophilus bacteriophages, which were reported recently (see table 1). They vary from 34,820 bp for the lytic cos-site phage DT1 to 43,075 bp for the temperate pac-site phage 01205. Its G/C content of 38 % is similar to that of the other phages and also identical to the 38 % of the host Streptococcus thermophilus (CUTG: Codon Usage Tabulated from GenBank, http://www.dna.affrc.go.jp/~nakamura/codon.html, current database size: 114 CDS, 30966 codons).

In contrast to virulent phages, not all orfs of Sfi21 were encoded on the same DNA strand. The exception were 4 adjacent genes covering the core genetic information for the lysogenic life style (see 12.2.1).
### Table 1: *Streptococcus thermophilus* bacteriophages; complete genomes

<table>
<thead>
<tr>
<th>Phage</th>
<th>Genome size [bp]</th>
<th>Life-style</th>
<th>Pack. mode</th>
<th>orfs</th>
<th>G/C-content</th>
<th>Origin</th>
<th>Reference Acc. num.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sfi21</td>
<td>40,739</td>
<td>Temperate</td>
<td>cos</td>
<td>53</td>
<td>38.1%</td>
<td>Lysogenic French yogurt starter</td>
<td>(Desiere et al., 1999), AF115103</td>
</tr>
<tr>
<td>Sfi19</td>
<td>37,392</td>
<td>Lytic</td>
<td>cos</td>
<td>44</td>
<td>38.3%</td>
<td>French yogurt isolate</td>
<td>(Lucchini et al., 1999a), AF115102</td>
</tr>
<tr>
<td>Sfi11</td>
<td>39,807</td>
<td>Lytic</td>
<td>pac</td>
<td>52</td>
<td>38.4%</td>
<td>French yogurt isolate</td>
<td>(Lucchini et al., 1999b), AF158600</td>
</tr>
<tr>
<td>01205</td>
<td>43,075</td>
<td>Temperate</td>
<td>pac</td>
<td>57</td>
<td>38%</td>
<td>Greek lysogenic yogurt starter</td>
<td>(Stanley et al., 1997), U88974</td>
</tr>
<tr>
<td>DT1</td>
<td>34,820</td>
<td>Lytic</td>
<td>cos</td>
<td>46</td>
<td>39.1%</td>
<td>Canadian mozzarella whey</td>
<td>(Tremblay and Moineau, 1999), AF085222</td>
</tr>
</tbody>
</table>

Otherwise, the genome of Sfi21 and the genome of the lytic group I cos-site Sfi19 (Lucchini et al., 1999a) are very similar. In fact, over major parts of the genome the two phages showed an identical gene order. The genes differed from each other only by point mutations (mainly at the third codon position) and small deletions/insertions. Overall the two phages differed by 10% bp changes. The changes were unevenly distributed: Over the packaging and morphogenesis modules the change rate was high (20%), while over the lysis and replication modules only few (<1%) bp changes were observed. Interestingly, both phage genomes differed only over the lysogeny module and the DNA replication modules. The virulent phages Sfi11 and Sfi19 showed a duplication of the phage origin of replication, while the temperate phages Sfi21 and O1205 possessed only the origin downstream of the primase gene (see 12.2.3). O1205 has a genome of 43,075 bp and contains a pac-site. It resembles Sfi21 in several respects: all but four genes of the core lysogeny module are coded on the same DNA strand; the lysogeny module is located at an identical genome position; the overall modular structure is identical. O1205 was closely related to virulent *S. thermophilus* phages, but in contrast to Sfi21 the closest similarity was with pac-site phages of lytic group II. Major differences between O1205 and lytic group II Sfi11 were located in three regions: the lysogeny module, the DNA replication region and the putative tail fiber gene. Over the remaining regions the two phages differed more or less uniformly by 10% bp changes except for the highly conserved DNA replication module (Lucchini et al., 1999b).

#### 12.2.1 Lysogeny module

In the lysogeny module of Sfi21 the following gene order was found (see chapter 5, 6): int (integrase), imm (superinfection immunity), mpm (metalloproteinase motif protein), cl (phage repressor) and in opposite orientation cro (phage repressor), ant (anti-repressor). This geometry creates a potential genetic switch structure between the two phage repressor genes. The lysogeny module is flanked on one side by the lysis cassette and on the other side by the DNA replication module. This gene constellation is found in many *Siphoviridae* infecting an evolutionary related group of gram-positive bacteria (*Streptococcus, Lactococcus, Lactobacillus, Bacillus*) (Lucchini et al., 1999). Numerous sequence similarities link the predicted proteins of *Siphoviridae* from these bacteria suggesting a common evolutionary origin also for their phages. Recombination events have, however, reshuffled the genes of this module: Sequence alignments of the lysogeny modules from different streptococcal phages revealed areas of high sequence conservation interspersed with regions of low sequence similarity. Several transition zones were found within genes. Contrary to the initial formulation of the modular theory of phage evolution the unit of the evolutionary exchange in streptococcal phages is not a group of functionally related genes, but can be as small as a single gene. Exchange reactions can even occur within genes, possibly between gene segments encoding distinct protein domains (Neve et al., 1998) (chapter 6).
Figure 1: Prediction of open reading frames (orf's) within the complete genome of *Streptococcus thermophilus* bacteriophage Sfi21. Orf's are marked with their length in amino-acids. Probable gene functions identified by bioinformatic analysis were noted. Gene functions identified by biological experiments were underlined. The phage genome was divided into functional units according to previous bioinformatic and comparative analysis. Possible rho independent terminators are indicated with a hairpin. Orf's with no filling lack information about function. Orfs preceded by a potential RBS are marked with a black dot in the arrow. Orfs starting with an unconventional initiation codon are indicated with a U. Triangles indicate overlapping start and stop codons. The location of the cos-site is indicated.
Several genes from the lysogeny module of Sfi21 were biologically characterized. The integrase mediates the integration of phage DNA into the chromosome of several *S. thermophilus* cells via site-specific recombination. Sequencing of the phage and bacterial attachment sites (*attP*, *attB*, *attR*, *attL*) identified a 40 bp identity region which, unusually, overlapped both the phage integrase gene for 18 bp and the terminal 11 bp of a host t-RNA gene. A phage DNA segment covering *attP* and the phage integrase contained all the genetic information for faithful integration of a non-replicative plasmid into the bacterial *attB* site. O1205 and TPJ-34 use a molecularly identical integration system (Bruttin et al., 1997).

When the indicator cell Sfi1 was lysogenized with Sfi21, the cell was protected against superinfection with the temperate phage and with 15 out of 19 virulent *S. thermophilus* phages (notably not Sfi19). The immunity functions could be attributed by cloning experiments to two phage genes, orf 203 and orf 127. Interestingly, the two genes showed complementary activities. Transformation of an indicator cell with the *cl* (orf 127) repressor protected the cell against superinfection with the homologous temperate phage and its deletion derivatives, but not against virulent phages. The *cl* repressor showed a helix-turn-helix DNA binding motif. Gel-shift experiments and DNase protection assays demonstrated binding to the putative genetic switch region between the *cl* and *cro* genes. Protection of the lysogenic cell against superinfection with virulent phages is achieved by the phage gene orf 203 located upstream of the integrase. Transformation of the indicator cell with this gene protected the cell against 12 virulent phages, but not against the homologous temperate phage (Foley et al., personal communication).

The *Cro* and *Ant* proteins showed both helix-turn-helix DNA binding motifs. Sequence similarity of *Cro* was observed to a repressor from a lambdoid coliphage 434 and the anti-repressor showed similarities to anti-repressor from a coliphage P1 and a protein from *Haemophilus influenzae* prophage *dflu* (Lucchini et al., 1999). However, a spontaneous Sfi21 deletion mutant (D1 in Figure 4 chapter 5 (Bruttin et al., 1997) showed that neither *ant* nor the adjacent downstream gene are necessary for establishment of lysogeny or induction of the prophage.

### 12.2.2 The lysis cassette

Three genes constitute the lysis cassette of *S. thermophilus* phages. In contrast to other phage systems, two holins (pore-forming proteins) precede the lysin gene. The holins could be identified due to their transmembrane structure consisting of 2 or 3 membrane-spanning loops (Smith et al., 1998). One holin showed a dual start site (i.e. one of the two start codons are used, located two or three codons apart, in the same frame, resulting in the synthesis of two N-terminally different isoforms from the same orf) suggesting a holin-antiholin control (Bläsi and Young, 1996). The two isoforms are thought to regulate the formation of the pore (Young, 1992).

The lysin gene suggested a modular design. The N-terminal domain encoding the amidase showed similarity to other phage lysins. The C-terminal domain encoding the substrate-binding part resembled host proteins. In an infectious variant phage the lysin gene was interrupted by a type I intron. The intervening DNA showed links to endonucleases found in self-splicing group I introns. It is was shown that both parts of the lysin messenger are rejoined on the mRNA-level to form a functional lysin (Foley et al., 1999).
12.2.3 The conserved replication module

Seven adjacent genes from the DNA replication module are the most conserved part of the *S. thermophilus* phage genomes (see chapter 7). More than 70% of the phages from different collections contained this DNA segment (Brüssow et al., 1994b). The very low sequence variability observed in this module makes it an attractive target for the design of engineered phage resistance mechanisms. This genome segment codes for four genes with nucleoside triphosphate binding motifs (Walker A motif) (Walker et al., 1982), one of which is a distant member of the helicase superfamily, while another showed sequence similarity to a DNA primase from the cryptic plasmid pWS58 of *Lactobacillus delbrückii* subsp. *lactis* and other primases. Downstream of the putative primase gene a non-coding DNA was identified which possesses conspicuous DNA repeats. This AT-rich DNA segment showed 80% sequence identity to a region following the repA gene of the cryptic *S. thermophilus* plasmid pST1 (Desiere et al., 1997; Janzen et al., 1992). In fact, when cloned into a low copy number plasmid it acted as an origin of replication driven by phage infection (Foley et al., 1998). The inhibitory activity was plasmid copy number dependent and operates by blocking the accumulation of phage DNA during infection. Protein(s) in Sfi21-, but not in Sfi19-infected cells interacted with the putative phage origin. In fact, Sfi19 was insensitive to the inhibition mediated by the Sfi21 origin. The Sfi19 origin differed in 14 nt positions from the Sfi21 origin. When cloned on a plasmid the Sfi19 origin protected the indicator cell against infection with Sfi19 and five further phages insensitive to the Sfi21 origin. Together the two origins protected the indicator cell against infection with 23 out of 25 phages tested. None of the phages was able to form plaques on the origin-containing starter. In combination with the superinfection immunity gene, it might provide the basis for the development of phage resistant starters for yogurt and mozzarella factories.

12.2.4 Structural Gene Cluster

The genetic organization of the DNA packaging and the head and tail morphogenesis region from Sfi21 resembled closely that of phages infecting evolutionary very distantly related hosts, the gram-negative *Eubacterium* *Escherichia* and the Archaeon *Methanobactertum* (see chapter 9). From the cos-site until the gene coding for the major tail protein, a surprising similarity in gene order was detected between Sfi21 and lambdoid coliphages. Comparable genes were found at identical positions; Sfi21 genes which could not be attributed showed a very similar size to the corresponding genes from coliphages. In fact, the Sfi21 gene map resembled that of coliphage HK97. The similarity was striking: both major head proteins showed an N-terminal coiled-coil structure, the mature major head proteins started at amino acid position 105 and 104 respectively, and both major head genes were preceded by genes encoding a possible protease and portal protein. These observations strongly suggest that the portal, protease and capsid proteins encoded by phage Sfi21 assemble to form phage heads using the same mechanism as that for HK97 (Hendrix and Duda, 1998). While the similar gene organization of the morphogenesis genes suggests a conserved functional module derived from a common ancestor there is also evidence for shuffling of individual genes between phages by horizontal exchange. The clpP-like protease of Sfi21 has a close relative in *Lactococcus lactis* phage BK5-T (Mahanivong et al., 1999) and in *Lactobacillus gasserii* temperate phage adh (Altermann et al., 1999). A clpP-like protease has recently been found in *Pseudomonas aeruginosa* phage D3, (acc. number AF147978, unpublished).
12.3 Relationship of Sfi21 to other bacteriophages

The structural proteins from *S. thermophilus* phages showed sequence similarity only to phages infecting gram-positive bacteria. A hierarchy of sequence relatedness was observed that correlated approximately with the evolutionary distance between the bacterial hosts (*S. thermophilus* ≫ *Lactococcus* > *Leuconostoc* ≈ *Bacillus* ≈ *Lactobacillus* ≫ *Mycobacterium* ≈ *Streptomyces*) (Lucchini et al, 1998). Sequence similarity in *Streptococcus thermophilus* phage Sfi21 was closest to phage BK5-T infecting *Lactococcus lactis*, which is also the closest evolutionary relative of *S. thermophilus* in this bacterial branch. When comparing the genomes of *Streptococcus thermophilus* phage Sfi21 and *Lactococcus lactis* BK5-T it becomes quite clear that exchange of genes or genetic modules has occurred. Sfi21 and BK5-T share a similar genome organization except for the replication region. BK5-T shows a replication region as in phage PVL and r11, where a single-strand-binding-protein (SSB) and a dUTPase could be identified. Sfi21 shows a distinct replication module with a helicase-primase-origin function. Nucleotide sequence similarity was only shared between streptococcal and lactococcal phages. With all other phages from low GC content Gram-positive bacteria, *Streptococcus thermophilus* phages showed only sequence similarity at the amino-acid level.

No sequence similarity was seen between Sfi21 and lambdoid phages, including HK97. Nevertheless, striking structural similarities were observed. Both major head proteins showed a comparable length, a comparable secondary structure prediction and proteolytic cleavage at a comparable position. Both head genes were preceded by a protease gene. A possible interpretation is that *Siphoviridae* from Eubacteria (and even Archaea) share a common, although very distant ancestor for this genome region.

12.4 Bacteriophage evolution

There are several mechanisms involved in the evolution of bacteriophage genomes: Point mutations, deletions/insertions (Ford et al., 1998), modular exchange reactions (Casjens et al., 1992) and recombinations (Tetart et al., 1998) (Ackermann and DuBow, 1987). What is the role of these mechanisms for the *Streptococcus thermophilus* phage genomes?

**Point mutations:** *Streptococcus thermophilus* phages have accumulated point mutations over time. Phages isolated repetitively over two to twenty years did not show sequence changes. Dairy factories are not evolution machines (see chapter 4). The diversification of *Streptococcus thermophilus* phages has not occurred in historical time periods. The lack of a molecular clock for phage DNA sequences precludes to translate the about 30% sequence diversity observed in *Streptococcus thermophilus* phages into a likely time period that separates these phages. Interestingly, the most distinct *Streptococcus thermophilus* phages come near but do not overlap the difference observed between streptococcal and lactococcal phages.

**Deletions/Insertions:** Comparisons of sequences of closely related phage genomes have shown that insertions or deletions involving genes or only gene-segments occurred frequently as has previously been demonstrated in the comparison of two completely sequences mycobacteriophages (Casjens et al., 1992; Desiere et al., 1998; Ford et al., 1998; Neve et al., 1998). In *S. thermophilus* we observed 3 different spontaneous deletions after serial passages of the phage Sfi21. Two were located in the lysogeny module and the phage integrase seemed to be implicated as a
recombinase. A third deletion occurred in a tail fiber gene and was flanked by a long direct repeat making slippage of the DNA polymerase the most likely mechanism.

**Modular exchange reactions:** Alignments of different *S. thermophilus* bacteriophage genomes suggested that phage genome segments are reshuffled by recombination events. Two hotspots for modular exchanges were identified; the lysogeny module and a putative phage anti-repressor gene (chapter 5). In contrast to the initial formulation of the modular theory of phage evolution by Botstein (Botstein, 1980) we observed much smaller segments of DNA as units of exchange; frequently they were as small as genes or even a gene segment encoding a likely protein domain.

**Recombination/Genetic Crosses:** Theoretically, virulent bacteriophage superinfecting a lysogenic cell should be able to recombine with the same resident prophage yielding chimera phages. Fayard *et al.* reported experiments that could be interpreted within this framework (Fayard *et al.*, 1993). We were, however, unable to reproduce these results in our laboratory.

**Horizontal gene transfer:** Similarities were found in phages from Gram-negative bacteria and eukaryotic invertebrate viruses. One entomopoxvirus protein shared an astonishing 40 % aa identity with the N-terminal domain of the BK5-T antirepressor. These findings suggest ancient modular lateral gene transfer. The importance of lateral gene transfer in prokaryotic genomes is becoming a focus of bacterial genomics (Doolittle, 1999).

The application of modern powerful bioinformatic methods on complete bacterial genomic sequences demonstrated that lateral gene transfer is a major and continuing force in archaeal and bacterial evolution. One example is *E. coli* where 18 % of the genes have been introduced in its genome since the species diverged from the *Salmonella* lineage 100 million years ago (Lawrence and Ochman, 1998). Other reports underline the extensive sharing of housekeeping, biosynthetic or catabolic genes between bacteria and archaea due to lateral gene transfer (Doolittle, 1999). It is therefore questionable that evolution proceed in the ancestor-descendent pattern that Darwin saw for multicellular animals and plants, and that there rather are additional mechanisms of evolution, such as horizontal gene-transfer, that will obscure the ancient relationship among the domains of life. For the prokaryotic world reticulate (net-like) phylogenetic schemes will probably replace the arborate (tree-like) phylogenetic schemes.
Table 2: Completely sequenced bacteriophages (as of October 1999)

<table>
<thead>
<tr>
<th>Phage</th>
<th>Host</th>
<th>Phage Family</th>
<th>Genome Size [bp]</th>
<th>Life-style</th>
<th>Packaging</th>
<th>Put. Genes</th>
<th>Reference Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sf21</td>
<td>S. thermophilus</td>
<td>Sipho</td>
<td>40,739</td>
<td>Temp</td>
<td>cos</td>
<td>53</td>
<td>(Desiere et al., 1999) AF115103</td>
</tr>
<tr>
<td>Sf19</td>
<td>S. thermophilus</td>
<td>Sipho</td>
<td>37,392</td>
<td>Lytic</td>
<td>pac</td>
<td>44</td>
<td>(Lucchini et al., 1999a) AF115102</td>
</tr>
<tr>
<td>Sf11</td>
<td>S. thermophilus</td>
<td>Sipho</td>
<td>39,807</td>
<td>Lytic</td>
<td>cos</td>
<td>52</td>
<td>(Lucchini et al., 1999b) AF158600</td>
</tr>
<tr>
<td>01205</td>
<td>S. thermophilus</td>
<td>Sipho</td>
<td>43,075</td>
<td>Temp</td>
<td>pac</td>
<td>57</td>
<td>(Stanley et al., 1997) U88974</td>
</tr>
<tr>
<td>DT1</td>
<td>S. thermophilus</td>
<td>Sipho</td>
<td>34,820</td>
<td>Lytic</td>
<td>cos</td>
<td>46</td>
<td>(Tremblay and Moineau, 1999) AF065222</td>
</tr>
<tr>
<td>BKS-T</td>
<td>Lactococcus lactis subsp. cremoris</td>
<td>Sipho</td>
<td>40,003</td>
<td>Temp</td>
<td>cos</td>
<td>64</td>
<td>(Mahanivong et al., 1999) AF176025</td>
</tr>
<tr>
<td>r1t</td>
<td>Lactococcus lactis subsp. cremoris</td>
<td>Sipho</td>
<td>33,350</td>
<td>Temp</td>
<td>cos</td>
<td>49</td>
<td>(van Sinderen et al., 1996) U38906</td>
</tr>
<tr>
<td>sk1</td>
<td>S. thermophilus</td>
<td>Sipho</td>
<td>28,451</td>
<td>Lytic</td>
<td>cos</td>
<td>54</td>
<td>(Chandry et al., 1997) AF011378</td>
</tr>
<tr>
<td>biL67</td>
<td>Lactococcus lactis subsp. cremoris</td>
<td>Sipho</td>
<td>22,195</td>
<td>Lytic</td>
<td>cos</td>
<td>37</td>
<td>(Schoulet et al., 1994) L33789</td>
</tr>
<tr>
<td>biL70</td>
<td>Lactococcus lactis subsp. cremoris</td>
<td>Sipho</td>
<td>31,754</td>
<td>Lytic</td>
<td>cos</td>
<td>63</td>
<td>(Altermann et al., 1999) AF009630</td>
</tr>
<tr>
<td>c2</td>
<td>Lactococcus lactis subsp. cremoris</td>
<td>Sipho</td>
<td>22,172</td>
<td>Lytic</td>
<td>cos</td>
<td>39</td>
<td>(Lubbers et al., 1995) L486685</td>
</tr>
<tr>
<td>LL-H</td>
<td>L. delbrückii subsp. lactis</td>
<td>Sipho</td>
<td>34,659</td>
<td>Temp</td>
<td>pac</td>
<td>48</td>
<td>(Mikkonen et al., 1996)</td>
</tr>
<tr>
<td>adh</td>
<td>Lactobacillus gasseri</td>
<td>Sipho</td>
<td>43,785</td>
<td>Temp</td>
<td>cos</td>
<td>63</td>
<td>(Altermann et al., 1999) AJ131519</td>
</tr>
<tr>
<td>phig1e</td>
<td>Lactobacillus plantarum</td>
<td>Sipho</td>
<td>42,259</td>
<td>Temp</td>
<td>pac</td>
<td>62</td>
<td>(Koda et al., 1997) X95106</td>
</tr>
<tr>
<td>phi-C31</td>
<td>Streptomyces</td>
<td>Sipho</td>
<td>41,489</td>
<td>Temp</td>
<td>cos</td>
<td>56</td>
<td>(Smith et al., 1999) A006589</td>
</tr>
<tr>
<td>psIM2</td>
<td>Methanob. thermoautotrophicum</td>
<td>Sipho</td>
<td>26,111</td>
<td>Temp</td>
<td>pac</td>
<td>31</td>
<td>(Pfister et al., 1998) AF065411</td>
</tr>
<tr>
<td>PVL</td>
<td>Staphylococcus aureus</td>
<td>Sipho</td>
<td>41,401</td>
<td>Temp</td>
<td>cos</td>
<td>63</td>
<td>(Kane et al., 1998) AB009886</td>
</tr>
<tr>
<td>Cp-1</td>
<td>Streptococcus pneumoniae</td>
<td>Podo</td>
<td>19,343</td>
<td>Lytic</td>
<td>pac</td>
<td>28</td>
<td>(Martin et al., 1996)</td>
</tr>
<tr>
<td>D29</td>
<td>Mycobacteriophage</td>
<td>Sipho</td>
<td>49,136</td>
<td>Lytic</td>
<td>cos</td>
<td>62</td>
<td>(Ford et al., 1998) Z47794</td>
</tr>
<tr>
<td>L5</td>
<td>Mycobacteriophage</td>
<td>Sipho</td>
<td>52,297</td>
<td>Lytic</td>
<td>cos</td>
<td>85</td>
<td>(Hafual and Sarkis, 1993) Z18946</td>
</tr>
<tr>
<td>TM4</td>
<td>Mycobacteriophage</td>
<td>Sipho</td>
<td>52,797</td>
<td>Lytic</td>
<td>cos</td>
<td>89</td>
<td>AF068845</td>
</tr>
<tr>
<td>HP-1</td>
<td>Haemophilus influenzae</td>
<td>Myo</td>
<td>32,355</td>
<td>Temp</td>
<td>cos</td>
<td>41</td>
<td>(Esposito et al., 1996) U12415</td>
</tr>
<tr>
<td>lambda</td>
<td>E. coli</td>
<td>Myo</td>
<td>48,502</td>
<td>Temp</td>
<td>cos</td>
<td>63</td>
<td>(Lebowitz et al., 1971) J02459</td>
</tr>
<tr>
<td>N15</td>
<td>E. coli</td>
<td>Myo</td>
<td>46,375</td>
<td></td>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td>E. coli</td>
<td>Podo</td>
<td>39,937</td>
<td></td>
<td></td>
<td>60</td>
<td>AF064539</td>
</tr>
<tr>
<td>25</td>
<td>E. coli</td>
<td>Podo</td>
<td>168,900</td>
<td></td>
<td></td>
<td>274</td>
<td>VT1145</td>
</tr>
<tr>
<td>186</td>
<td>E. coli</td>
<td>Podo</td>
<td>30,624</td>
<td>Temp</td>
<td>cos</td>
<td>46</td>
<td>U32222</td>
</tr>
<tr>
<td>28</td>
<td>E. coli</td>
<td>Podo</td>
<td>33,593</td>
<td>Temp</td>
<td>cos</td>
<td>42</td>
<td>AF063097</td>
</tr>
<tr>
<td>29</td>
<td>Escherichia coli</td>
<td>Podo</td>
<td>61,670</td>
<td>Temp</td>
<td>cos</td>
<td>60</td>
<td>(Plunkett et al., 1999) AF125820</td>
</tr>
<tr>
<td>30</td>
<td>Bacillus subtilis</td>
<td>Podo</td>
<td>44,007</td>
<td>pac</td>
<td>102</td>
<td></td>
<td>X97158</td>
</tr>
<tr>
<td>31</td>
<td>Bacillus subtilis</td>
<td>Podo</td>
<td>19,366</td>
<td></td>
<td></td>
<td>27</td>
<td>M11813</td>
</tr>
<tr>
<td>32</td>
<td>Bacillus subtilis</td>
<td>Podo</td>
<td>18,630</td>
<td></td>
<td></td>
<td>17</td>
<td>(Pencenkov and Paces, 1999) X93260</td>
</tr>
<tr>
<td>33</td>
<td>Bacillus subtilis</td>
<td>Podo</td>
<td>30,325</td>
<td>Temp</td>
<td>cos</td>
<td>51</td>
<td>AB081280</td>
</tr>
<tr>
<td>34</td>
<td>Pseudomonas aeruginosa</td>
<td>Sipho</td>
<td>35,580</td>
<td>Temp</td>
<td>cos</td>
<td>47</td>
<td>(Nakayama et al., 1999) AB008550</td>
</tr>
<tr>
<td>35</td>
<td>Escherichia coli</td>
<td>O157:H7</td>
<td>60,942</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>S. aureus</td>
<td>Fusella</td>
<td>15,465</td>
<td>Temp</td>
<td></td>
<td>31</td>
<td>AP000363</td>
</tr>
<tr>
<td>37</td>
<td>Listeria monocytogenes</td>
<td>Sipho</td>
<td>40,834</td>
<td>Temp</td>
<td>pac</td>
<td>72</td>
<td>(Palm et al., 1991) AJ242593</td>
</tr>
<tr>
<td>38</td>
<td>Mycoplasma arthritidis</td>
<td>?dsDNA</td>
<td>15,644</td>
<td>Temp</td>
<td></td>
<td>15</td>
<td>(Voelker and Dybvig, 1999)</td>
</tr>
</tbody>
</table>

*Sipho=Siphoviridae (phages with non-contractile tails), Myo=Myoviridae (phages with contractile tails), Podo=Podoviridae (phages with short tails), Fusello=Fuselloviridae, see chapter 3.1.*
12.5 Genome organization of tailed bacteriophages

Only about 35 complete bacteriophage genomes have been reported to date in the literature (Hendrix et al., 1999) (table 2), which is a remarkably low number in contrast to the 24 publicly available complete microbial genomes (TIGR Microbial Database http://www.tigr.org). This number certainly represents a very tiny fraction of the total number of phages in world. When comparing the genome organization of temperate *Siphoviridae* from low GC-content Gram-positive bacteria we found that they share a similar modular character (Lucchini et al., 1999). 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 regulation module (see Figure 2). This common overall organization sets this group apart from the two currently established genera of temperate *Siphoviridae* (the lambda genus and the L5-like genus) (Maniloff and Ackermann, 1998). One could propose therefore a new genus of Sfi21-like phages according to their common genome organization.

![Diagram](image.png)

Figure 2: Schematic overview of the genome organization of some dsDNA viruses. The bacteriophages are aligned to the left on their respective packaging region (Pack). The morphogenesis region is given (Morphogenesis), lysis functions (Lys, L), lysogeny functions (Lysogeny) and replication functions (Replication). Unknown sequences are indicated as empty square (gray).

It has only been recently stated in the literature due to a systematic comparison of nine bacterial and archaeal genomes that there is a low level of gene-order (and operon architecture) conservation present in bacterial genomes. Nevertheless, a number of gene pairs are conserved. The proteins encoded by conserved gene pairs appear to interact physically. This observation can therefore be used to predict functions of, and interactions between, gene products (Dandekar et al., 1998). In bacteriophages, gene
order is much more conserved. We documented a conservation of gene-order in the absence of similarities over the morphogenesis region of bacteriophages from gram-negative, gram-positive and archaea (see chapter 9). The order of genes would be kept as a heritage from the common ancestor. In this case the genes that interact would be subject to co-evolution, which preserve the productive interaction between these genes (in the morphogenesis region). Interestingly, temperate Siphoviridae from low GC content Gram-positive bacteria have a common overall genome structure that sets them apart from other temperate Siphoviridae. The packaging region is followed by the structural region, then the lysis and the lysogeny region and finally the replication region. Cladistic analysis would probably identify temperate Siphoviridae from low GC content bacteria as a homogeneous group and a younger offspring from the more ancestral ancestor common to Gram-positive, Gram-negative and archaea.
13 References

Reference List


14 Curriculum Vitae

Frank Desiere
Born on July 31st 1970 • German nationality
Fdesiere@yahoo.com

Qualifications

<table>
<thead>
<tr>
<th>Year</th>
<th>Qualification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996-1999</td>
<td>Ph.D. Microbiology, ETH Zurich, Switzerland. (To be accepted)</td>
</tr>
<tr>
<td>1999</td>
<td>European Doctorate in Biotechnology, (EDBT), European Federation of Biotechnology, European Molecular Biology Organization (EMBO), European Union Rectors' Conference. (To be accepted)</td>
</tr>
<tr>
<td>1989-1995</td>
<td>Diploma in Biotechnology, University of Stuttgart, Germany.</td>
</tr>
</tbody>
</table>

Academic work

<table>
<thead>
<tr>
<th>Year</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>Research Project, Armstrong Laboratory, Panama City, USA.</td>
</tr>
</tbody>
</table>

Practical experience

<table>
<thead>
<tr>
<th>Year</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>Research Project, University of Melbourne, Australia.</td>
</tr>
<tr>
<td>1996</td>
<td>Research Scientist, Nestlé Research Center, Team Microbiology, Lausanne, Switzerland.</td>
</tr>
<tr>
<td>1995</td>
<td>Summer Student Project, Procter &amp; Gamble Technical Center, Process Development Team Beverages Europe, Schwalbach, Germany.</td>
</tr>
<tr>
<td>1995</td>
<td>Researcher, Nestlé Research Center, Lausanne, Switzerland.</td>
</tr>
<tr>
<td>1992</td>
<td>Technician, Fraunhofer Institute for Interfacial Engineering and Biotechnology, Stuttgart, Germany.</td>
</tr>
<tr>
<td>1989</td>
<td>Temporary employment, Merck KGaA, Darmstadt, Germany. Logistical operations.</td>
</tr>
</tbody>
</table>
15 Publication list


