Virology **302,** 21–32 (2002)
doi:10.1006/viro.2002.1571

Transcription Analysis of *Streptococcus thermophilus* Phages in the Lysogenic State

Marco Ventura, Anne Bruttin, Carlos Canchaya, and Harald Brüssow[1]

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

Received November 22, 2001; returned to author for revision January 30, 2002; accepted May 9, 2002

The transcription of prophage genes was studied in two lysogenic *Streptococcus thermophilus* cells by Northern blot and primer-extension experiments. In the lysogen containing the *cos*-site phage Sfi21 only two gene regions of the prophage were transcribed. Within the lysogeny module an 1.6-kb-long mRNA started at the promoter of the phage repressor gene and covered also the next two genes, including a superinfection exclusion (*sie*) gene. A second, quantitatively more prominent 1-kb-long transcript was initiated at the promoter of the *sie* gene. Another prophage transcript of 1.6-kb length covered a group of genes without database matches that were located between the lysin gene and the right attachment site. The rest of the prophage genome was transcriptionally silent. A very similar transcription pattern was observed for a *S. thermophilus* lysogen containing the *pac*-site phage O1205 as a prophage. Prophages from pathogenic streptococci encode virulence genes downstream of the lysin gene. We speculate that temperate phages from lactic streptococci also encode nonessential phage genes ("lysogenic conversion genes") in this region that increase the ecological fitness of the lysogen to further their own evolutionary success. A comparative genome analysis revealed that many temperate phages from low GC content Gram-positive bacteria encode a variable number of genes in that region and none was linked to known phage-related function. Prophages from pathogenic streptococci encode toxin genes in this region. In accordance with theoretical predictions on prophage–host genome interactions a prophage remnant was detected in *S. thermophilus* that had lost most of the prophage DNA while transcribed prophage genes were spared from the deletion process.

INTRODUCTION

Many bacterial genome-sequencing projects have revealed an important contribution of prophage sequences to the host chromosome. For example, the sequencing of the *Escherichia coli* O157 Sakai genome identified 18 prophages or prophage remnants integrated into the bacterial DNA (Hayashi *et al.,* 2001). The whole set of prophage sequences accounted for half of the strain-specific DNA when the O157 strain was compared to the *E. coli* reference laboratory strain K-12 (Ohnishi *et al.,* 2001). This clearly implies that bacteriophages are a major contributor to the genomic diversification of *E. coli.* In addition, prophages from *E. coli* O157 encode a number of potential virulence factors. Prophages might thus also play an important role in the evolution of bacterial pathogenicity. This is not an isolated case: prophages contribute also to the virulence of *Salmonella* strains (Figueroa-Bossi and Bossi, 1999; Figueroa-Bossi *et al.,* 2001), another important food pathogen, and numerous toxins from pathogenic bacteria are prophage-encoded (Bishai and Murphy, 1988).

Our laboratory is interested in prophages and their influence on the phenotype of lactic acid bacteria (low GC content Gram-positive bacteria). Lactic acid bacteria play two important roles for humans. On one side, *Streptococcus pyogenes* and *Streptococcus pneumoniae* are major bacterial pathogens of humans. Most clinical isolates from both species are lysogenic or even polylysogenic (Hynes *et al.,* 1995; Ramirez *et al.,* 1999; Yu and Ferretti, 1991). It was also shown that *Str. pyogenes* prophages contribute important virulence factors to the lysogenic host (Goshorn and Schlievert, 1989; McShan *et al.,* 1997; Weeks and Ferretti, 1984). Detailed analysis of the prophage genomes integrated into the completely sequenced *Str. pyogenes* strain SF370 identified potential virulence genes in all three prophages (Desiere *et al.,* 2001; Ferretti *et al.,* 2001). On the other side, lactic acid bacteria are the major bacterial starter organisms for food fermentation and are increasingly discussed as probiotic organisms, i.e., gut bacteria conferring health benefit to humans and animals. Due to their practical importance for the dairy industry as the major cause of industrial fermentation failures, phages from lactic acid bacteria became a focus of comparative phage genomics (Brüssow, 2001; Brüssow and Desiere, 2001; Brüssow and Hendrix, 2002). These efforts made phages from lactic streptococci, lactococci, and lactobacilli the best documented phage group in the database. In addition, several genomes of dairy bacteria have now been sequenced and many contained important contributions of prophage sequences. For example, the genome of *Lactococcus lactis* strain IL1403, a major starter organism for cheese fermentation, contains six prophage elements (Chopin *et al.,* 2001) and two different *Lactobacillus* species discussed as probiotic bacteria contained two and three prophage sequences (D. Pridmore and R. Siezen,

[1] To whom correspondence and reprint requests should be addressed.

0042-6822/02 $35.00

personal communication). It is currently unknown to what extent prophages contribute to the phenotype of dairy bacteria. To answer this question it is not sufficient to investigate the prophage sequences for genes that could change the phenotype of the lysogen. We need data on the expression of prophage genes in the lysogenic cell and how their transcription is regulated. However, experimental data on this subject are scarce for lysogenic dairy bacteria (Boyce *et al.,* 1995a).

To fill this gap we investigated the transcription pattern of two *Streptococcus thermophilus* prophage genomes in the lysogenic state. We have chosen this industrially important yogurt and cheese starter for the investigation because *Str. thermophilus* phages are argueably the best characterized phage system with respect to genomics data. In addition, a detailed transcription map was developed for the temperate *cos*-site *Str. thermophilus* phage Sfi21 in the lytic mode of infection (Ventura *et al.,* 2002). Here we report that only two genome segments of *Str. thermophilus* phages are transcribed in the lysogenic state. Notably, one region corresponds to the genome segment from *Str. pyogenes* prophages that encodes virulence factors (Desiere *et al.,* 2001; Ferretti *et al.,* 2001). In the *Str. thermophilus* prophage remnant Sfi16 transcribed genes are spared from massive prophage DNA deletion, suggesting a selective value for these genes as predicted by theoretical considerations (Desiere *et al.,* 2001; Lawrence *et al.,* 2001).

RESULTS

Transcription of the Sfi21 prophage genes

An indicator strain was lysogenized with the *cos*-site temperate *Str. thermophilus* phage Sfi21. The lysogen showed a low spontaneous prophage induction rate: only 10^3 pfu/ml of infectious phage was released. Systematic Northern blot analysis demonstrated that few Sfi21 prophage genes were expressed in the lysogenic cell. The *cI*-like repressor (orf 127) gene was transcribed as a 1.6-kb-long mRNA (Fig. 1A). Primer-extension experiments (Fig. 2A) demonstrated that the *in vivo* 5′-end of this transcript was in fact within orf 127, suggesting the use of a second in-frame start codon for the phage repressor (Fig. 3). This start site would predict a 116-aa-long repressor. The *in vivo* 5′-end of the mRNA is preceeded by a possible promoter, but no ribosomal binding site (Fig. 3). According to its length estimate the transcript would cover the *cI*-repressor gene, orf 122 (encoding a protein with a metalloproteinase motif) and orf 203 (encoding an experimentally demonstrated superinfection immunity exclusion gene (*sie*), Bruttin *et al.,* 1997a) (Fig. 4, bottom). In fact, when an orf 203-specific probe was used, the 1.6-kb mRNA species was again detected. In addition, a smaller, but much stronger hybridization signal for a 1-kb-long transcript was observed (Fig. 1B). Primer-extension experiments located the 5′-end of this 1-kb transcript downstream of orf 203 (Fig. 2B). A possible promoter and a ribosomal binding site (RBS) (Fig. 3) preceded the start codon.

No transcription of the *cro*-like repressor (Fig. 1J) or the *ant*-like repressor genes was detected in the lysogenic cell (Fig. 1E). No transcription of early, middle, and late phage genes described for lytic infection of *Str. thermophilus* cells (Ventura *et al.,* 2002) was detected in the lysogenic cell. This is illustrated in Fig. 1 for orf 143 to 130 ("early" transcriptional regulation genes, Fig. 1F), orf 157 to 271 ("middle" DNA replication genes, Fig. 1K), orf 623 to 384 ("late" DNA packaging genes, Fig. 1H), orf 397 to 117a (late head and head-to-tail joining genes, Fig. 1G), orf 1560 (late tail tape measure gene, Fig. 1L), and orf 117b to 288 (late lysis genes, Fig. 1I), respectively. A notable exception were the early genes orf 110 to 140b located downstream of the phage lysin gene. This prophage genome region gave rise to a 1.6-kb-long mRNA. Primer extension experiments (Fig. 2C) identified possible −35 and −10 promoter sites and a RBS in front of the orf 110 start codon (Fig. 3). The length of the transcript predicts termination between orf 140 b and the right attachment site.

When mapped on the prophage genome, all prophage transcripts were located near the left and the right attachment sites, while the large central region of the prophage genome was transcriptionally silent (Fig. 4). All PCR probes used in the present study were positively tested in lytically infected cells and none gave a hybridization signal with noninfected, nonlysogenic control *Str. thermophilus* cells (Ventura *et al.,* 2002).

Transcription of the O1205 prophage genes

We obtained the lysogenic *Str. thermophilus* strain CNRZ1205 containing the *pac*-site temperate phage O1205 (Stanley *et al.,* 1997) from D. van Sinderen (Cork, Ireland). Phages Sfi21 and O1205 differ completely over their structural genes (Brüssow and Desiere, 2001), while the lysogeny modules from both phages are related in a mosaic fashion: seven genes showed greater than 72% aa sequence identity, while orfs 52, 53, 2 (likely superinfection exclusion), orf 4 (*cI*-like repressor), orf 5 (*cro*-like repressor), and orf 7 were sequence-unrelated (Fig. 4).

Orf 4, the *cI* repressor gene, was expressed as a 1.4-kb-long transcript (Fig. 5A). Primer-extension experiments (Fig. 2D) demonstrated again that this transcript was translated from an internal start codon within orf 4 (Fig. 3). A possible promoter and a RBS were identified. The length of the transcript predicts cotranscription of orf 4 with orfs 3 and 2 (Fig. 4). Indeed, a 1.4-kb-long transcript was also revealed when using an orf-2-specific DNA probe (Fig. 5B). In contrast to the Sfi21 lysogen, only a weak 0.9-kb transcript specific for orf 2 was detected. An *int*-specific probe revealed, as in the case of Sfi21 lysogen, a transcript of a molecular weight similar to that of the *sie* transcript (Figs. 1B, 1C, 5B, and 5C, respectively). The *cro*-like repressor and the lysin genes were

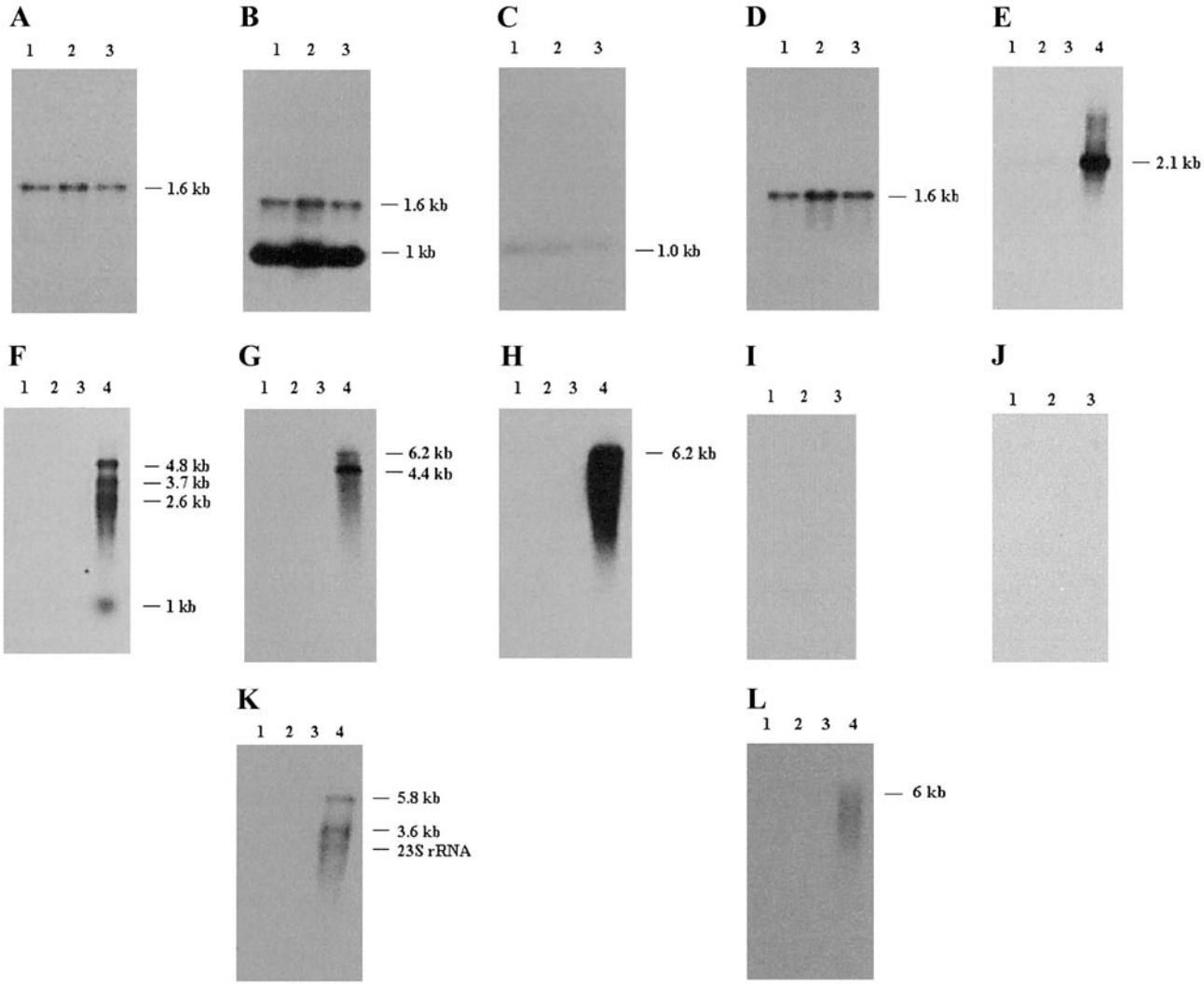

FIG. 1. Northern blot hybridization of RNA isolated from the *Str. thermophilus* lysogen containing the *cos*-site phage Sfi21. RNA was isolated from lysogenic cells grown to an OD of 0.4 (lane 1), 0.8 (lane 2), or 1.2 (lane 3). Where indicated, lane 4 refers to a cell lytically infected with phage Sfi21. The molecular weight calculated for the hybridization signal is provided. The following probes were used for the indicated panel. A: orf 127/*cI*-like repressor; B: orf 203/superinfection exclusion (*sie*) gene; C: orf 359/integrase *int;* D: orf 110–140b/orfs between lysin gene and attachment site; E: orf 287/anti-repressor *ant;* F: orf 143–orf 130/early transcription regulation genes; G: orf 397–orf 117a/head and head–tail-joining genes; H: orf 623–orf 384/DNA packaging genes; I: orf 117 to 288/lysis cassette; J: orf 75/*cro*-like repressor; K: orf 157–orf 271/DNA replication genes; L: orf 1560/tail measure gene.

not or only very weakly transcribed (Figs. 5D and 5E). When the orf 52 to 57 region, located downstream of the lysin gene, was used as a probe, a prominent 1.8-kb-long transcript was detected (Fig. 5G). By comparative genomics, orf 52 and 53 shared similarity with the virulent *pac*-site *Str. thermophilus* phages (Lucchini *et al.,* 1999a) and orf 54 to 57 shared similarity with the temperate *cos*-site phage Sfi21 (Lucchini *et al.,* 1999b). When orf 52 and 53 were used as probe, only the minor transcripts of 1.3- and 1.2-kb lengths were detected (Fig. 5F), suggesting that the major 1.8-kb transcript was initiated at orf 54. Indeed, primer extension experiments identified a strong radioactive signal upstream of orf 54 and a weaker signal farther ahead of orf 52 (Figs. 2F, 2E, and 3). A consensus promoter and a RBS preceeded the orf 54 transcript. In contrast, no consensus promoter structure preceeded orf 52 and a stem–loop structure separated the experimentally determined 5′-end of the mRNA and translation start site from orf 52 (Fig. 3). As in the case of phage SFi21 the transcribed prophage O1205 genes were mapped near the two attachment sites. Interestingly, next to the *attR* site phage O1205 showed a significantly lower GC content and an obvious GC skew when compared to the rest of the prophage (Fig. 4).

Comparative genomics of the lysin-to-integrase region

No bioinformatic links could be established for the transcribed Sfi21 and O1205 prophage genes located near the *attR* site. Comparative genome analysis of temperate *Siphoviridae* from low GC content Gram-positive

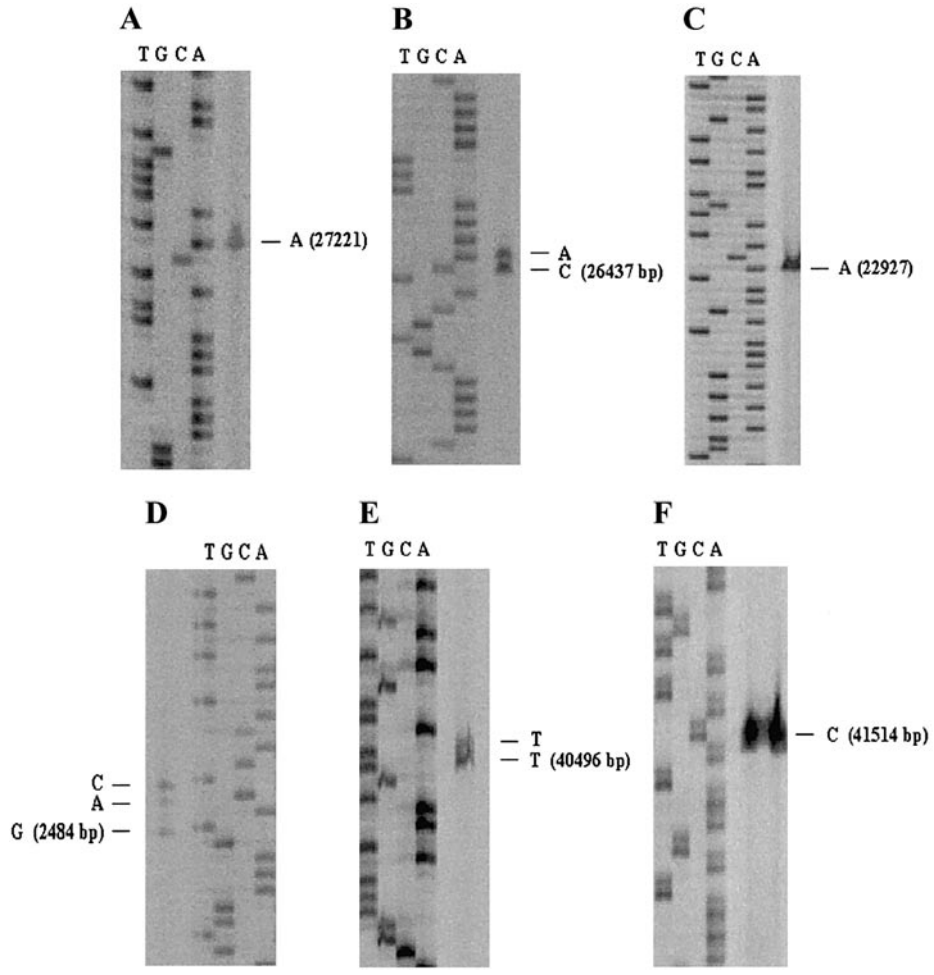

FIG. 2. Primer-extension experiments with RNA isolated from *Str. thermophilus* cells containing prophage Sfi21 (A–C) and O1205 (D–F). The oligonucleotides were targeted to the 5′-ends from orf 127 (A), orf 203 (B), and orf 110 (C) from Sfi21 phage and orf 4 (D), orf 52 (E), and orf 54 (F). The major extended product is located next to a sequencing gel of the region and is identified by the bp position from the corresponding database entry.

01205

54 TTTTTTGCAAAAAAATATAAATTTATTTATCAAAAGTGTTGACGTATGTCAACATGTAAGATATAATGTATACATAAGATAAAGAGAGAGGTATACAAAATG
4 CAAAGTCAAGTCTTTTTGTTCATTTTTTTGAACTTTTTTTACTTTTTTCTTGAACTTTATGAAAATATACTATATAATACACTCATGAAAGGTGATGTAATTATG
52 TTTGTTTGCTTTATCATAGATAAGTGTTATGATAGTTATTGTAATATTTGGCATCGTTTCAATAAATCTCTATAACTGTACCCTGAGTTTTGGTCAGGGTTTTTA
TTTTTTTGGTTTTGATAATTCTTAGATATATAATATTTACATG
2 TTACTAGTTGATTATTTAAAATCTACTGATATCTACGATTTTAATTGGGTTCGCTTTGCTGCACAGTATAATATTTCCACTACTTGGGGCGAAGCGATGATACAAGATG

SFi21

110 TGCAAAAAAATATAAATTTCTTTATCAAAAACGTTGACGTATGTCAACAAGTAAGATATAATGTATACATAAGATAAAGAGAGAGGTAAACAAAATG
127 AAAAAAATATGAAAAAAGTTTATTTTGGGAACATATTGTTTATTTTTGGGAACTGTTGTATAATGTTTATACTATTAAATAAAGGAAAAAAATGTATG
203 TGAAAAACGAATACTTAAAGTTAATAGAAACTTGAATAAACAATTCGTAGATATTTACAAAATTTAAAACTACGTGCAAAACTGAATCACGTTAAAAGCTGATGGAAGGATTCATATTATG

SFi16

56 TTGTGGTGGTTTATTTCCAATTCAATTTGAGAAAGGTTGCCTGATAGTGCTATTTCGTTCATTGTTTCCCTTTCTATTTTTGATATAATAGTTTTAAAAAAACGAGGTACTGTATG

FIG. 3. Overview of the experimentally determined 5′-ends of the transcripts from lysogenic *Str. thermophilus* cells containing prophage O1205, prophage Sfi21, or prophage remnant Sfi16. The end points of the primer extension products are marked by an *. A double and a single line underline putative −35 and −10 promoter sites, respectively. The most likely start codon is given at the right end of the sequence; possible ribosomal binding sites are boxed. A wavy line underlines a stem–loop structure ahead of orf 52.

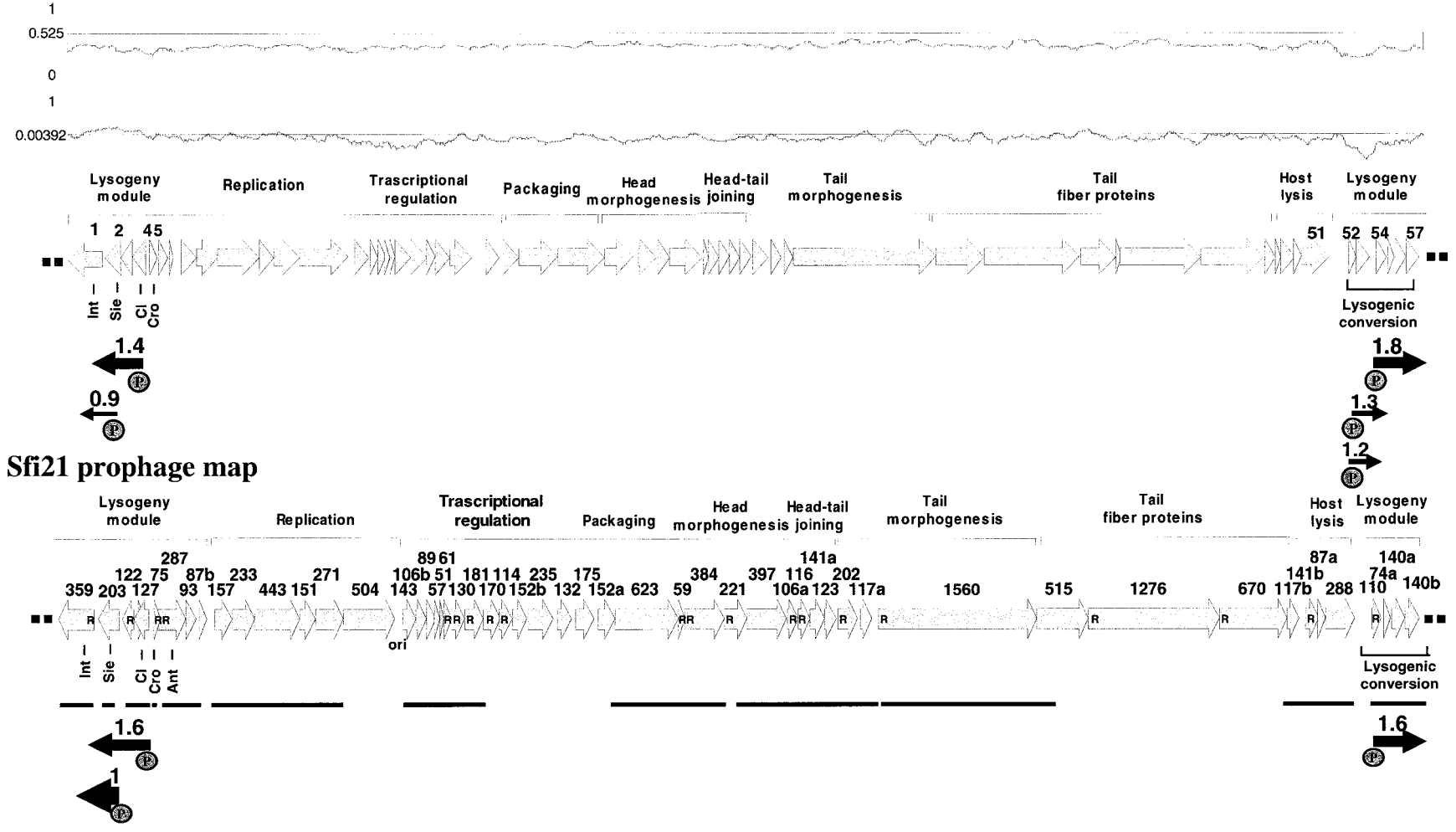

FIG. 4. Localization of the transcripts from the two lysogenic *Str. thermophilus* cells on the prophage genome map of phage O1205 (top) and phage Sfi21 (bottom). From top to bottom: GC content, GC skew, the gene map with modular organization and selected gene annotation for phage O1205 in the prophage constellation (the two black dots indicate bacterial DNA sequences into which the prophage had integrated), and the location of the phage transcripts from the lysogenic cell are given as arrows. The transcripts are indicated with molecular weight in kb; the width of the arrow indicates the intensity of the hybridization signal. P indicates that the 5′-end of the transcripts was determined in primer extension experiments. At the bottom are the corresponding prophage Sfi21 data. The calculation of the GC skew was done using a window size of 100 (DNAStarGeneQuest).

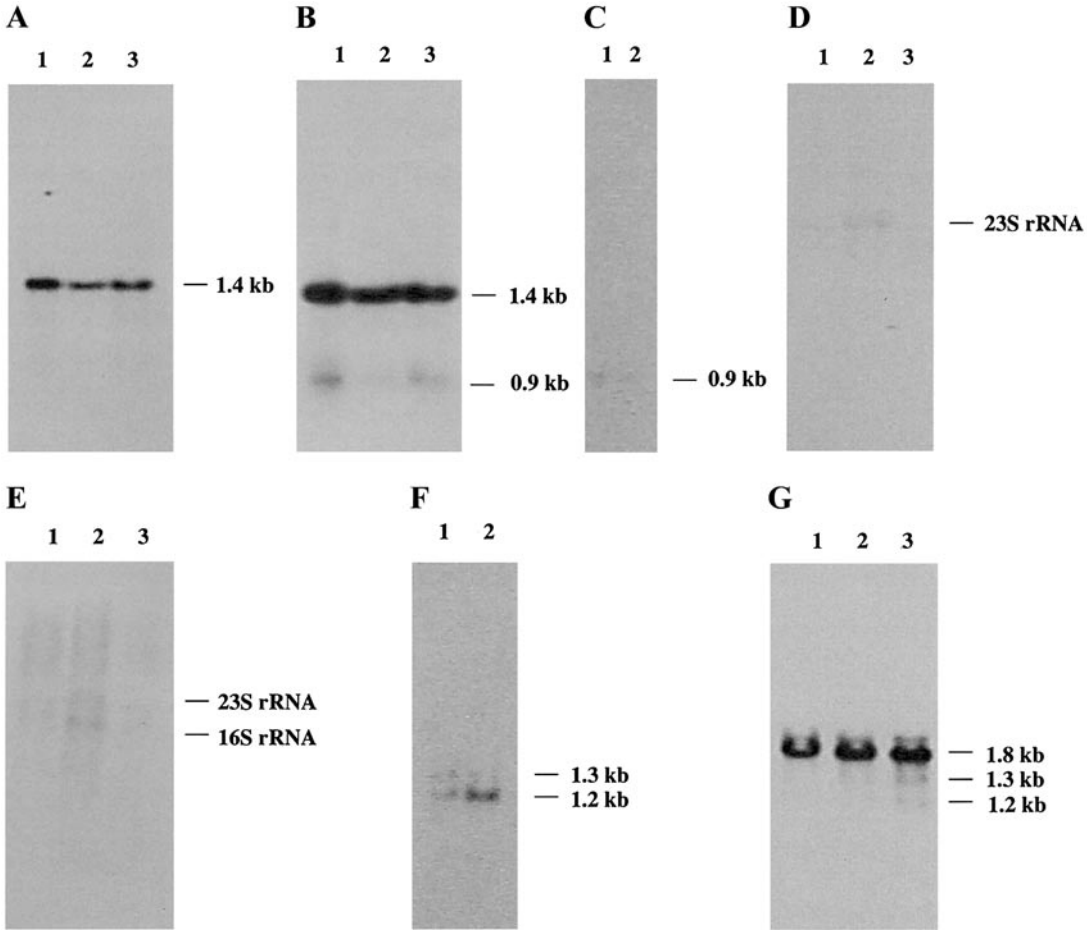

FIG. 5. Northern blot hybridization of RNA isolated from the *Str. thermophilus* cell CNRZ 1205 containing the *pac*-site *Str. thermophilus* phage O1205. RNA was isolated from lysogenic cells grown to an OD of 0.4 (lane 1), 0.8 (lane 2), or 1.2 (lane 3). The molecular weight calculated for the hybridization signal is provided. The following probes were used for the indicated panel. A: orf 4/*cI*-like repressor; B: orf 2/*sie;* C: orf 1/integrase; D: orf 5/*cro*-like repressor; E: orf 51/lysin; F: orf 52-orf 53/downstream of lysin gene; G: orf 52-orf 57/between the lysin gene and attachment site.

bacteria showed that the genome region between the phage lysin and the phage integrase contained, with the exception of phage MM1, one to six genes (Fig. 6). None of these genes was linked to known phage-specific functions or entries in the database. Only the temperate phages from *Str. pyogenes* or *Staphylococcus aureus,* two prominent pathogenic bacteria, encoded established or suspected virulence factors for the bacterial pathogen at this genome position. From the literature, several temperate phages from *Sta. aureus* encode two different combinations of leukocyte toxins (LukS and LukF or LukM and LukF) or an exfoliative toxin in this region (Kaneko *et al.,* 1998). Temperate *Str. pyogenes* encode superantigens (speC, H, I) or mitogenic factors (mf2, 3) in this region (Desiere *et al.,* 2001; Ferretti *et al.,* 2001). Their distinct GC content also suggests that these are not genuine phage genes.

Prophage remnant Sfi16

On the basis of theoretical reasoning it was predicted that prophage DNA will experience inactivating mutations, followed by massive loss of prophage DNA (Desiere *et al.,* 2001; Lawrence *et al.,* 2001). Phage genes transcribed in the lysogenic cell should be spared from this deletion process since this DNA might encode functions that are of selective advantage to the lysogen (Desiere *et al.,* 2001). To test these predictions we screened *Str. thermophilus* isolates from our strain collection for prophage remnants. This was done by a two-step hybridization procedure. In a first step, a dot blot containing the chromosomal DNA from 47 *Str. thermophilus* strains was hybridized against the combined DNA from a *cos*- and a *pac*-site temperate *Str. thermophilus* phage (Sfi21 and TPJ-34). Two strains gave a clearly positive result. The prophages could be induced by mitomycin C and visualized in the electron microscope, and one Sfi21-like prophage could be lytically propagated (data not shown). We confirmed with the same test a lysogeny rate of about 10% in another strain collection (Fayard *et al.,* 1993).

DNA from *Str. thermophilus* strain Sfi16 gave only a weakly positive hybridization result. On a Southern blot

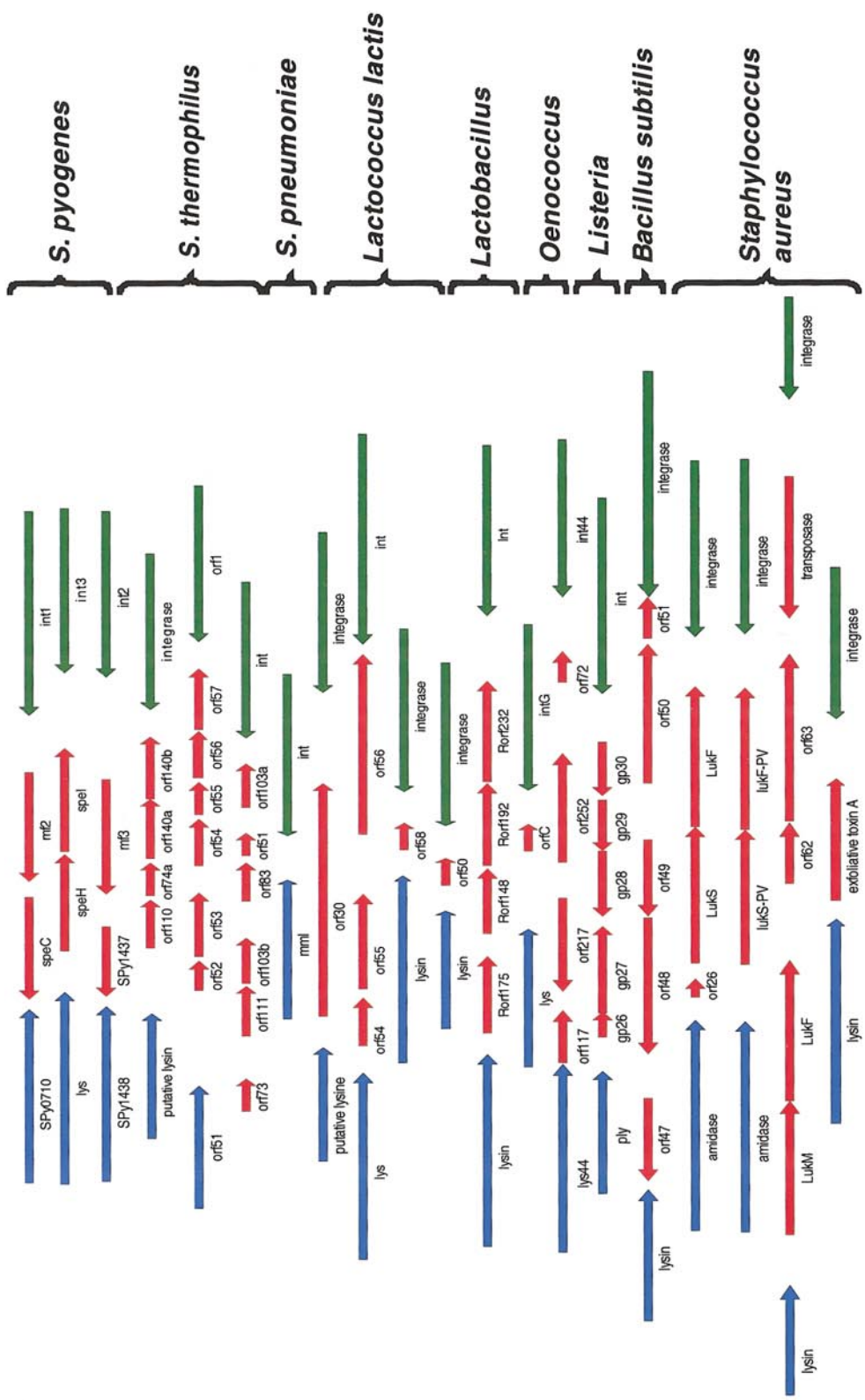

FIG. 6. Compilation of currently available partial genome maps of temperate *Siphoviridae* from low GC content Gram-positive bacteria. The depicted genome regions are flanked on the left by the phage lysin (blue) and on the right by the phage integrase gene (green). Candidates for lysogenic conversion genes are highlighted in red. The code name of the prophages or temperate phages is indicated at the left of the gene map; the host of the specified phage is indicated at the right of the map.

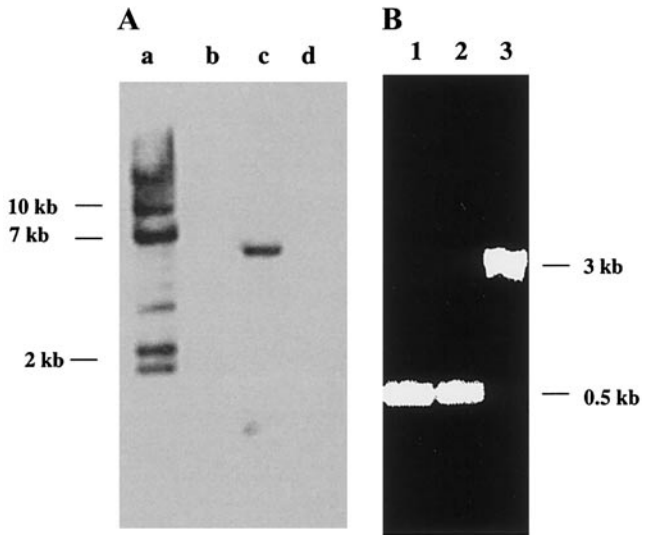

FIG. 7. Prophage remnant Sfi16. A. *Eco*RI-digested chromosomal DNA from a *Str. thermophilus* strain containing the complete Sfi21 prophage (lane a), two prophage-free *Str. thermophilus* strains (lanes b and d), and *Str. thermophilus* strain Sfi16 containing a prophage remnant (lane c). Labeled total phage Sfi21 DNA was used as a hybridization probe. B. PCR amplification of two prophage-free *Str. thermophilus* strains (lanes a and b) and the Sfi16 strain containing the prophage remnant Sfi16 (lane c) with primers flanking the *attB* site.

Sfi16 DNA yielded a single hybridization signal with the labeled Sfi21 DNA probe (Fig. 7A). Primers were placed at both sides of the bacterial *attB* site: chromosomal DNA from all nonlysogenic cells gave the expected 0.5-kb PCR product while DNA from strain Sfi16 yielded an approx. 3-kb-long amplification product (Fig. 7B). Sequence analysis of this 3-kb product revealed an authentic *attL* site overlapping the phage integrase (Bruttin *et al.,* 1997b). The phage integrase gene differed from phage Sfi21 integrase at only 15-bp positions. A total of 1.5 kb of putative prophage DNA followed upstream of the phage integrase before joining again bacterial DNA 82 bp downstream of *attR* resulting in the loss of this attachment site. This 1.5-kb DNA segment encodes three orfs; one of them (orf 97) was 90% nt sequence identical with orf 2 from the virulent *Str. thermophilus* phage 7201 (Fig. 8A). However, orf 56 and 168 showed no database matches. Northern blot hybridization (Fig. 8B) demonstrated that orf 56 and 168 gave rise to a 1.3- and a 1.1-kb-long mRNA. The *int* gene, but not orf 97, was transcribed. In phage Sfi21 the orf upstream of the *int* gene encodes a superinfection exclusion gene that is strongly transcribed in the lysogenic state (Fig. 1B and Bruttin *et al.,* 1997a). To test for a possible *sie* function associated with the prophage element, we cloned the entire prophage element except for the *int* gene into the high-copy-number plasmid pNZ124. An indicator cell susceptible to 25 distinct virulent *Str. thermophilus* phages was transformed with this construct, but no phage resistance phenotype was associated with this cell.

DISCUSSION

In order to assure their ecological and evolutionary success, prophages must encode genetic functions that are of selective advantage to the lysogenic host. Without these genes, prophages represent only a burden to the cell and will lead to the elimination of the lysogen from the bacterial population and thereby to the loss of the prophage (Desiere *et al.,* 2001). One obvious group of prophage genes with positive selective value for the host cell is the phage immunity functions. In temperate streptococcal phages immunity functions come in two forms. One form is the *CI*-like repressor that protects the cell against superinfection with temperate phages sharing identical or related repressor DNA recognition sites as shown for phage Sfi21 (Foley *et al.,* submitted). In addition, the Sfi21 repressor down-regulates the transcription of the *cro*-gene when both were cloned on plasmids (unpublished results). The transcription analysis revealed the expected mRNA for the phage repressor gene in the Sfi21 lysogen, the lack of expression of the *cro*-repressor gene, and transcriptional silence for all prophage genome regions yielding middle and late transcripts in the lytic infection cycle. The experimental observations concur thus with the theoretical expectations from other phage systems. A second form of immunity is mediated by orf 203 in phage Sfi21 located directly upstream of the phage integrase gene. Expression of that gene on a high-copy-number plasmid protects the transformed cell against superinfection with a wide range of virulent phages (Bruttin *et al.,* 1997a). Orf 203 is transcribed on two mRNA species: it is cotranscribed with the *cI* gene and it is transcribed from its own promoter. The latter messenger is the most prominent phage mRNA in the Sfi21 lysogen. In an evolutionary context, expression of this superinfection exclusion gene (*sie*) benefits both host and prophage. The lysogenic cell experiences a valuable protection against infection with virulent *Str. thermophilus* phages, while the prophage is protected against unwanted competitor phages. The dual immunity system in phage Sfi21 is probably typical for many temperate phages of dairy bacteria. Genes with bioinformatic links to repressor and superinfection immunity genes, but unrelated to Sfi21 genes at the sequence level, are also found in corresponding positions in the temperate *pac*-site *Str. thermophilus* phage O1205 and both genes were also expressed in the O1205 lysogen. The gene in lactococcal phage Tuc2009 located upstream of the integrase gene is also transcribed in the lysogenic state and encodes a membrane-anchored protein which interferes with the DNA injection process of some superinfecting phages (McGrath *et al.,* 2002).

An *int*-specific probe yielded a weak hybridization signal in all *Str. thermophilus* lysogens. This could reflect a low-level transcription of this gene compatible with the low level of spontaneous induction of the prophage. However, a transcription start site upstream of the *int*

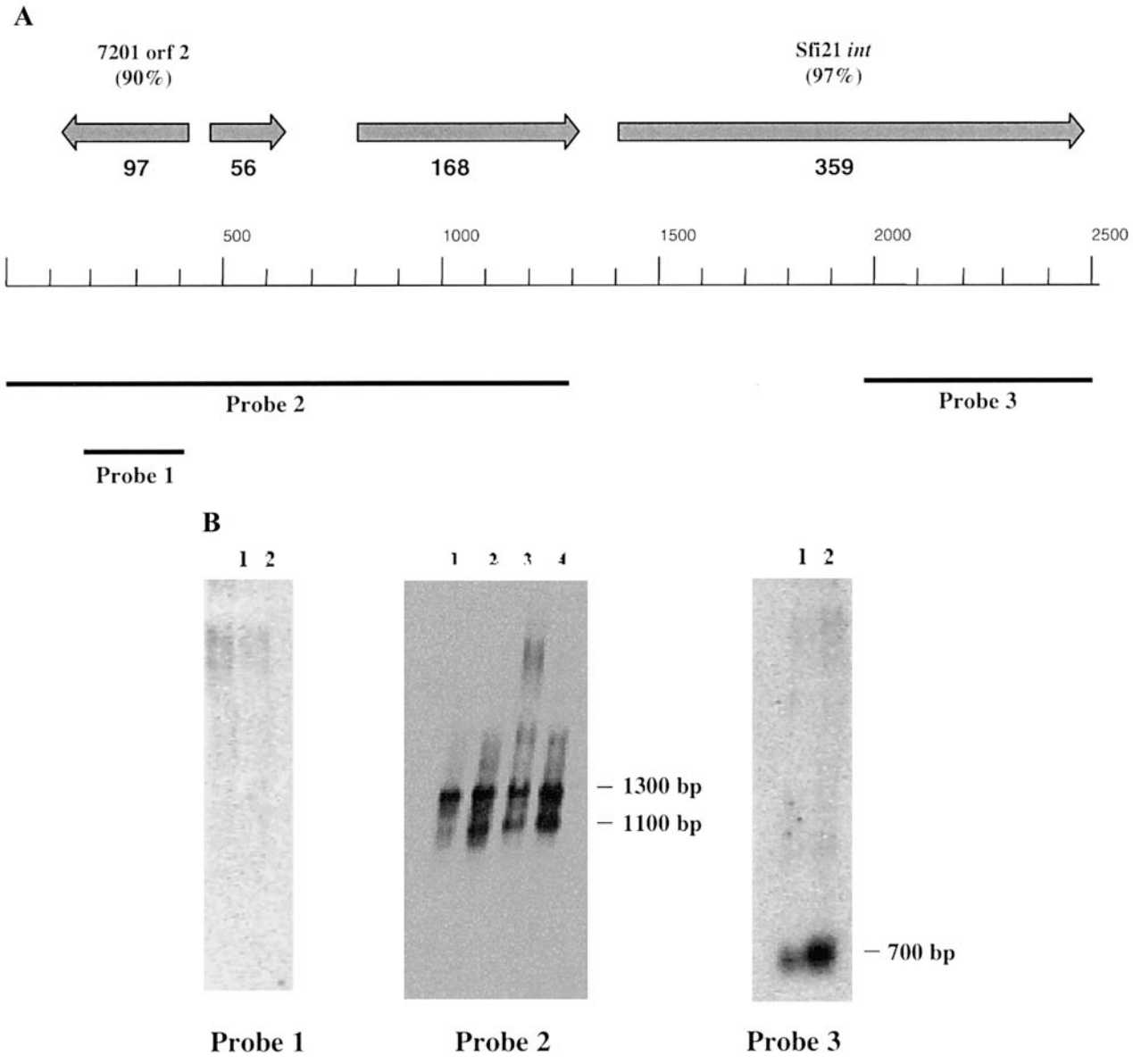

FIG. 8. A. Gene map of the prophage remnant Sfi16. The predicted orfs are indicated with their codon length and were annotated with their database matches. The positions of the probes used in Northern blot experiments are indicated under the gene map. B. Northern blot hybridization of RNA isolated from the *Str. thermophilus* cell Sfi16 with probes 1, 2, and 3. The molecular weight calculated for the hybridization signal is provided.

gene could not be identified in primer-extension experiments. We must therefore also consider the possibility of a read-through of the transcript started at the promoter of the *sie* gene into the 5′ half of the *int* gene. This interpretation is backed by the identical size of the mRNA revealed by *sie* and *int* probes.

In addition to the immunity functions, temperate phages from pathogenic lactic acid bacteria encode further genes that could contribute a selective advantage to the lysogen. In *Str. pyogenes* and *Sta. aureus* these genes encode toxins, superantigens, mitogenic factors, and DNases (Desiere *et al.,* 2001; Ferretti *et al.,* 2001; Kaneko *et al.,* 1998). The proteins are either known or suspected virulence factors for the pathogenic host bacterium that target the immune system either directly (leucotoxins of *Sta. aureus*) or indirectly by a nonspecific overstimulation of lymphocytes (superantigens of *Str. pyogenes*). The prophage-encoded DNase possibly assists the phage-encoded hyaluronidase in the spreading of streptococci along tissue planes. On the prophage map, these genes are encoded between the phage lysin and the attachment site. This peculiar genome location probably reflects the acquisition of bacterial genes during a faulty prophage excision event in an unusual bacterial host. The foreign origin of these genes is underlined in *Str. pyogenes* prophages by a distinctly lower GC content when compared to the prophage and bacterial host DNA.

Interestingly, the corresponding genome region in temperate phages from dairy bacteria is not empty: it

contains one to six genes. Based on the comparison with the prophages from pathogenic lactic acid bacteria we hypothesize that these genes fulfill similar functions, i.e., they increase the ecological fitness of the lysogen in its natural environment (milk, decaying plant material). Since this environment is drastically different from that of *Str. pyogenes* (body surfaces of humans), we must anticipate very different genes to fulfill this task. Due to their medical interest, the ecology (epidemiology) of pathogens like *Str. pyogenes* has been extensively investigated and knowledge on virulence factors abounds. In contrast, our knowledge on dairy bacteria derives from their growth in man-made industrial and laboratory environments. We know the environment in which these prophage–host relationships were developed over long time periods very poorly. We should stress that the evidence for a lysogenic conversion region between the lysin gene and the attachment site in phages from lactic acid bacteria is at the moment only circumstantial. None of the respective genes in dairy phages is known to fulfill a phage-specific function; actually all lack a database match. The possibly foreign origin of these genes is suggested by a distinct GC content and a GC skew in phage O1205. The importance of these genes for the lysogen is proposed by the observation that they represent one of the few prophage genes expressed in the lysogenic state. As in the case of the *sie* gene, these genes are also expressed early during lytic infection with phage Sfi21 (Ventura *et al.,* 2002).

A very similar transcription pattern was described for *L. lactis* prophage BK5-T (Boyce *et al.,* 1995a). In the BK5-T prophage the putative *cI*-like repressor and *sie* genes were transcribed as 1.8- and 0.8-kb-long transcripts and the *int* gene gave rise to a 1.3-kb-long transcript. In contrast to the situation in the *Str. thermophilus* prophages the transcription of the BK5-T *int* gene was unequivocal and relatively prominent (Boyce *et al.,* 1995a). This might reflect the higher spontaneous induction rate in BK5-T. Most notably, in the current context, orf 536 located between the lysin gene and the *att* site of the BK5-T prophage also yielded a prominent 1.8-kb-long transcript. The lack of bioinformatic links prevents any speculation with respect to the function of the gene. In BK5-T, orf 536 was lost in a spontaneous deletion mutant. This deletion did not affect the lytic growth nor the formation and maintenance of stable lysogens (Boyce *et al.,* 1995b). This experiment tells us that this gene is nonessential for the laboratory propagation of the phage, but this does not exclude an important function in the natural ecological niche of *L. lactis.* A possible case of lysogenic conversion phenotype is provided by the *Str. thermophilus* phage TPJ-34. The TPJ-34 lysogen grows in suspension, while the prophage-cured cells aggregate (H. Neve, personal communication). Re-lysogenization resulted again in growth in suspension. The phenotype is not due to the inactivation of a host gene during prophage integration: TPJ-34 integrates into the same locus as Sfi21, a tRNA gene which is left intact by the integration event (Neve *et al.,* 1998).

Our data have implications for another aspect of prophage biology. It has been postulated that the prophage–host DNA interaction is a genetic arms race (Desiere *et al.,* 2001; Lawrence *et al.,* 2001). Selection will impose deletion processes that prevent an undue accumulation of prophage genomes in bacterial strains. Deletions are predicted to result in prophage remnants. We have further postulated that the prophage genes that confer a selective advantage to the lysogen are spared from this deletion process (Desiere *et al.,* 2001). The DNA element next to the $tRNA_{Arg}$ of *Str. thermophilus* strain Sfi16 concurs with several of the predictions. It is a prophage element as demonstrated by the close sequence similarity of three of five orfs with *Str. thermophilus* phages and the conservation of an authentical *attL* site. From the lack of an *attR* site, we deduce that the deletion process started in the prophage DNA and ended in bacterial DNA. Very similar prophage remnants were identified in *Str. pyogenes* strain SF370 (Canchaya *et al.,* submitted). Interestingly, as in Sfi16, the *Str. pyogenes* prophage remnants maintained the integrase gene. This could indicate that the phage integrase is of selective value to the lysogen or that the deletion process has some directionality (starting from the lysis/structural gene region). Transcription analysis demonstrated that orf 56 and 168, but not orf 97, were transcribed. This observation is compatible with the hypothesis on the sparing of genes with selective value from the deletion process. Since the orf 97 homologue in phage 7201 is located directly downstream of the *cro*-repressor, we suspected that the genes from the SFi16 prophage remnant were derived from the lysogeny module of an ancestor prophage. Since the gene upstream of the integrase frequently encodes a *sie*-like gene in dairy phages, we postulated a *sie* function for this gene. The N-terminus of orf 168 is very hydrophobic which is compatible with a membrane association and a possible *sie* function (McGrath *et al.,* 2002). In fact, strain Sfi16 is one of the industrial starters with the narrowest phage susceptibility pattern in our collection. However, cloned Sfi16 prophage DNA did not protect an indicator cell against phage superinfection.

Few dairy strains have been investigated in their natural environment. This will limit our ability to test for lysogenic conversion phenotypes. The above-mentioned change in growth phenotype of *Str. thermophilus* cells containing the TPJ-34 prophage could be a suitable start point to link this phenotype with prophage gene expression (transcription analysis, knock-out experiments). Notably, microarray experiments have demonstrated that the change from planktonic to biofilm growth in *Pseudomonas* is associated with the activation of few genes, and surprisingly prophage genes were prominent in this group of genes (Whiteley *et al.,* 2001).

MATERIALS AND METHODS

The DNA segment surrounding the attachment site in the prophage-free *Str. thermophilus* strain Sfi1 and in the prophage remnant-containing strain Sfi16 was amplified by PCR using primers 3 and 4 described in Bruttin *et al.* (1997a). The two PCR products were sequenced on both strands by the Sanger method of dideoxy-mediated chain termination using the fmol DNA sequencing system of Promega (Madison, WI). The prophage remnant SFi16 sequence was deposited under Accession No. AY082374 in GenBank. Nucleotide and predicted amino acid sequences were compared to those in the latest Genbank release using BLAST (Altschul *et al.,* 1997) at the NCBI and FASTA (Lipman and Pearson, 1985).

Total RNA isolation

Total RNA was isolated by resuspending the frozen bacterial cell pellets in 1 ml of TRIzol Reagent (GibcoBRL, Gaithersburg, MD), adding 106 μm glass beads (Sigma Chemical Co., St. Louis, MO), and shearing the slurry with a Mini-Beadbeater-8 cell disruptor (Biospec Products, Bartlesville, UK) as described (Walker *et al.,* 1999). Standard procedures to minimize RNase contamination were used (Sambrook and Russel, 2001). Northern blot analysis of phage transcripts were carried out on 15-μg aliquots of total RNA separated on a 1.5% agarose–formaldehyde denaturing gel, transferred to a Zeta-Probe blotting membrane (Bio-Rad Laboratories, Richmond, CA) by the method of Sambrook and Russel (2001), and fixed by UV cross-linking using a Stratalinker 1800 (Stratagene).

Northern blots and hybridizations

Prehybridization and hybridization of the Northern blots were carried out at 65°C in 0.5 M $NAHPO_4$ (pH 7.2)–1.0 mM EDTA (pH 7.0)–7% sodium dodecyl sulfate (SDS). Following 18 h of hybridization, the membrane was rinsed twice (30 min) at 65°C in 0.1 M NaHPO4 (pH 7.2)–1.0 mM EDTA–1% SDS and twice (30 min) at 65°C in 0.1 M NaHPO4 (pH 7.2)–1.0 mM EDTA–0.1% SDS and exposed to X-OMAT autoradiography film (Eastman Kodak Co., Rochester, NY).

The probes for the Northern blot hybridization were labeled with [α^{32}P] using the Random primed DNA labeling system (Boehringer Mannheim, GmbH) and purified with Nuc Trap probe purification columns (Stratagene).

Primer extension analysis

The 5′ ends of RNA transcripts were determined in primer extension reactions conducted with 15 μg of total RNA mixed with 1 pmol of primer (IRD800 labeled) and 2 μl of Buffer H (2 M NaCl, 50 mM PIPES, pH 6.4). The mixture was denatured at 90°C for 5 min and then hybridized for 1 h at 42°C. After the addition of 5 μl 1 M Tris–HCl (pH 8.2), 10 μl 0.1 M DTT, 5 μl 0.12 M $MgCl_2$, 20 μl 2.5 mM dNTP mix, 0.4 μl (5 U) of reverse transcriptase (Sigma), and 49.6 μl of double-distilled water, the enzymatic reaction was incubated for 2 h at 42°C. The reaction was stopped by adding 250 μl ethanol/acetone mix (1:1) and incubated at −70°C for 15 min followed by a centrifugation at 10,000 rpm for 15 min. The pellets were dissolved in 4 μl of distilled water and mixed with 2.4 μl loading buffer from the sequencing kit (Thermo sequenase fluorescence labeled, Amersham). The cDNA was separated through 8% polyacrylamide–urea gels along with sequencing reactions which were conducted using the same primers employed for primer extensions and detected using the LiCor sequencer machine (MWG Biotech).

The synthetic oligonucleotides used for each orf were as follows; the nucleotide position refers to the Sfi21 sequence reported in NC_000872: 110 (5′ position 23074, CCTTCTTTAGCATCAGCCCAAG, 3′ position 23053), 203 (5′ position 26282, GCAATATAGAATGAAACATTG, 3′ position 26303), 127 (5′ position 27111, CCCAGCTTTATTTGCTAATTCTG, 3′ position 27130).

The oligonucleotides used in primer extensions analysis of *Str. thermophilus* O1205 prophage for each orf were as follows; the nucleotide position refers to O1205 sequence reported in U88974: 52 (5′ position 40826, CATTTTTGTGACAAGAGCAAGGTTTAG, 3′ position 40798), 54 (5′ position 41653, GCAGCCCAAGCCATTTTCATAGC, 3′ position 41624), 4 (5′ position 2323, CCGCTAACATATTGAGAGAGG, 3′ position 2349), 2 (5′ position 1555, GCAACACCCACAAACATTAATAACC, 3′ position 1581). In Sfi16 we used the following oligonucleotide; the nucleotide position refers to the genome map reported in Fig. 8: 56 (5′ position 605, CTGTCCATGAGTTAGGTCGTTTTC, 3′ position 581).

Cloning

Two primers with restriction sites were designed in order to amplify the region of the prophage excluding the integrase. Primer 1 (*Eco*RI site): 5′-GCGAATTCCGATTACCTTTTCAGTGG-3′ and primer 2 (*Bgl*II site): 5′-GCAGATCTCGTTTGCGGTATGATGC-3′. The PCR product obtained was purified with the QIAquick gel extraction kit (Qiagen) and digested with the corresponding enzymes. The region was cloned in pNZ124 digested with the same enzymes. The ligation product was electroporated first in *E. coli* XL-1 and then in *Str. thermophilus* Sfi1.

ACKNOWLEDGMENTS

This work was supported in part by Research Grant 5002-057832 from the Swiss National Science Foundation. We thank Athena Zourari (Cork, Ireland) for providing us the lysogenic *S. thermophilus* strain CHRZ1205 and phage O1205.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new

generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Bishai, W. R., and Murphy, J. R. (1988). Toxic bacteriophage gene products. *In* "The Bacteriophages" (R. Calendar, Ed.), pp. 683–724, Plenum Press, New York/London.

Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1995a). Identification of prophage genes expressed in lysogens of the *Lactococcus lactis* bacteriophage BK5-T. *Appl. Environ. Microbiol.* **61,** 4099–4104.

Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1995b). Spontaneous deletion mutants of the *Lactococcus lactis* temperate bacteriophage BK5-T and localization of the BK5-T *attP* site. *Appl. Environ. Microbiol.* **61,** 4105–4109.

Bruttin, A., Desiere, F., Lucchini, S., Foley, S., and Brüssow, H. (1997a). Characterization of the lysogeny DNA module from the temperate *Streptococcus thermophilus* bacteriophage Sfi21. *Virology* **233,** 136–148.

Bruttin, A., Foley, S., and Brüssow, H. (1997b). The site-specific integration system of the temperate *Streptococcus thermophilus* bacteriophage Sfi21. *Virology* **237,** 148–158.

Brüssow, H. (2001). Phages of dairy bacteria. *Ann. Rev. Microbiol.* **55,** 283–303.

Brüssow, H., and Desiere, F. (2001). Comparative phage genomics and the evolution of Siphoviridae: Insights from dairy phages. *Mol. Microbiol.* **39,** 213–223.

Brüssow, H., and Hendrix, R. W. (2002). Phage genomics: Small is beautiful. *Cell* **108,** 13–16.

Canchaya, C., Desiere, F., McShan, W. M., Ferretti, J. J., Parkhill, J., and Brüssow, H. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. Submitted.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D., and Chopin, M. (2001). Analysis of six prophages in *Lactococcus lactis* IL1403: Different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* **29,** 644–651.

Desiere, F., McShan, W. M., van Sinderen, D., Ferretti, J. J., and Brussow, H. (2001). Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: Evolutionary implications for prophage–host interactions. *Virology* **288,** 325–341.

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

Ferretti, J. J., McShan, W. M., Ajdic, D., Savic, D. J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A. N., Kenton, S., Lai, H. S., Lin, S. P., Qian, Y., Jia, H. G., Najar, F. Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S. W., Roe, B. A., and McLaughlin, R. (2001). Complete genome sequence of an M1 strain of *Streptococcus pyogenes. Proc. Natl. Acad. Sci. USA* **98,** 4658–4663.

Figueroa-Bossi, N., and Bossi, L. (1999). Inducible prophages contribute to Salmonella virulence in mice. *Mol. Microbiol.* **33,** 167–176.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D., and Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. *Mol. Microbiol.* **39,** 260–271.

Goshorn, S. C., and Schlievert, P. M. (1989). Bacteriophage association of streptococcal pyrogenic exotoxin type C. *J. Bacteriol.* **171,** 3068–3073.

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C. G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., and Shinagawa, H. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* **8,** 11–22.

Hynes, W. L., Hancock, L., and Ferretti, J. J. (1995). Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes:* Evidence for a third hyaluronidase involved in extracellular enzymatic activity. *Infect. Immun.* **63,** 3015–3020.

Kaneko, J., Kimura, T., Narita, S., Tomita, T., and Kamio, Y. (1998). Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage PVL carrying Panton–Valentine leukocidin genes. *Gene* **215,** 57–67.

Lawrence, J. G., Hendrix, R. W., and Casjens, S. (2001). Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* **9,** 535–540.

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

Lucchini, S., Desiere, F., and Brüssow, H. (1999a). Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* **73,** 8647–8656.

Lucchini, S., Desiere, F., and Brüssow, H. (1999b). Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria. *Virology* **263,** 427–435.

McGrath, S., Fitzgerald, G. F., and Van Sinderen, D. (2002). Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. *Mol. Microbiol.* **43,** 509–520.

McShan, W. M., Tang, Y. F., and Ferretti, J. J. (1997). Bacteriophage T12 of *Streptococcus pyogenes* integrates into the gene encoding a serine tRNA. *Mol. Microbiol.* **23,** 719–728.

Neve, H., Zenz, K. I., Desiere, F., Koch, A., Heller, K. J., and Brüssow, H. (1998). Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: Implications for the modular theory of phage evolution. *Virology* **241,** 61–72.

Ohnishi, M., Kurokawa, K., and Hayashi, T. (2001). Diversification of *Escherichia coli* genomes: Are bacteriophages the major contributors? *Trends Microbiol.* **9,** 481–485.

Ramirez, M., Severina, E., and Tomasz, A. (1999). A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae. J. Bacteriol.* **181,** 3618–3625.

Sambrook, J., and Russel, D. W. (2001). "Molecular Cloning: Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Stanley, E., Fitzgerald, G. F., Le Marrec, C., Fayard, B., and van Sinderen, D. (1997). Sequence analysis and characterization of phi O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology* **143,** 3417–3429.

Ventura, M., Foley, S., Bruttin, A., Canchaya, C., and Brüssow, H. (2002). Transcription mapping as a tool in phage genomics: The case of the temperate *Streptococcus thermophilus* phage Sfi21. *Virology,* in press.

Walker, D. C., Girgis, H. S., and Klaenhammer, T. R. (1999). The groESL chaperone operon of Lactobacillus johnsonii. *Appl. Environ. Microbiol.* **65,** 3033–3041.

Weeks, C. R., and Ferretti, J. J. (1984). The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. *Infect. Immun.* **46,** 531–536.

Whiteley, M., Bangera, M. G., Bumgarner, R. E., Parsek, M. R., Teitzel, G. M., Lory, S., and Greenberg, E. P. (2001). Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413,** 860–864.

Yu, C. E., and Ferretti, J. J. (1991). Molecular characterization of new group A streptococcal bacteriophages containing the gene for streptococcal erythrogenic toxin A (speA). *Mol. Gen. Genet.* **231,** 161–168.