Widespread Distribution of a Group I Intron and Its Three Deletion Derivatives in the Lysin Gene of *Streptococcus thermophilus* Bacteriophages

SOPHIE FOLEY, ANNE BRUTTIN, AND HARALD BRÜSSOW*

Nestlé Research Centre, Nestec Ltd., CH-1000 Lausanne 26, Switzerland

Of 62 *Streptococcus thermophilus* bacteriophages isolated from various ecological settings, half contain a lysin gene interrupted by a group IA2 intron. Phage mRNA splicing was demonstrated. Five phages possess a variant form of the intron resulting from three distinct deletion events located in the intron-harbored open reading frame (orf 253). The predicted orf 253 gene sequence showed a significantly lower GC content than the surrounding intron and lysin gene sequences, and the predicted protein shared a motif with endonucleases found in phages from both gram-positive and gram-negative bacteria. A comparison of the phage lysin genes revealed a clear division between intron-containing and intron-free alleles, leading to the establishment of a 14-bp consensus sequence associated with intron possession. The conserved intron was not found elsewhere in the phage or *S. thermophilus* bacterial genomes. Folding of the intron RNA revealed secondary structure elements shared with other phage introns: first, a 38-bp insertion between regions P3 and P4 that can be folded into two stem-loop structures (shared with introns from *Bacillus* phage SPO1 and relatives); second, a conserved P7.2 region (shared with all phage introns); third, the location of the stop codon from orf 253 in the P8 stem (shared with coliphage T4 and *Bacillus* phage SPO1 introns); fourth, orf 253, which has sequence similarity with the H-N-H motif of putative endonuclease genes found in introns from *Lactococcus*, *Lactobacillus*, and *Bacillus* phages.

Introns are regions which are transcribed and subsequently excised from the primary transcript by RNA splicing to generate the mature RNA. Group I and II introns, in which the folded structure of the intron participates directly in the splicing reaction, can be distinguished by their secondary structures and the mechanisms that they use to catalyze their own splicing. Group I introns, which are particularly well characterized, catalyze their own splicing by a series of guanosine-initiated *trans*-esterification reactions (for a review, see reference 20). Introns are remarkable not only in their role in encoding ribozymes but also in their ability to act as mobile genetic elements capable of efficiently inserting themselves into cognate intronless alleles. This process, known as "homing," is mediated by different types of intron-encoded proteins which place the intron in the correct sequence context for efficient splicing (20).

The evolutionary history and biological role of introns have been the subject of much debate, and their true significance remains to be defined. The controversial debate has been nourished by the peculiar phylogenetic distribution of introns. Group I introns have been found within genes encoding mRNA, rRNA, and tRNA in diverse genetic systems, including eucaryotic (plant, fungus, and yeast) mitochondria, nuclei, and chloroplasts; bacteriophages infecting gram-positive and gram-negative bacteria; and eubacterial genomes (see references 20 and 27 for a review and a compilation, respectively). Among the eubacteria, the only group I introns described to date occur in several genera of cyanobacteria and purple proteobacteria (3, 32). Although group I introns have been described for several T-even bacteriophages of *Escherichia coli* (30, 36), six

Bacillus subtilis bacteriophages (1, 17, 18, 22), and one bacteriophage each of *Lactococcus lactis* (29), *Lactobacillus delbrueckii* (28), and *Staphylococcus aureus* (21), none have been reported so far for the genomes of the respective host bacteria. Interestingly, apart from the T-even bacteriophages, all of the intron-containing phages described to date infect a group of evolutionarily related low-GC-content gram-positive bacteria. It has also been shown that the temperate phages of these bacteria are evolutionarily related (23, 24). Since *Streptococcus* and *Lactococcus* are closely related bacterial genera and since the analysis of introns from further phages of this group could constrain theories on the phylogenetic distribution and origin of introns, we investigated *Streptococcus thermophilus* phages (6) for the presence of introns.

Due to their economical importance for the dairy industry, large ecologically characterized *S. thermophilus* bacteriophage collections are available (5). Furthermore, five *S. thermophilus* bacteriophage genomes have been completely sequenced (25, 26, 37, 39). This fact, in addition to the fact that all *S. thermophilus* bacteriophages are related in terms of DNA homology (5), makes them well suited for systematic intron searches. So far, the distribution of introns within closely related bacteriophages has been studied only for T-even (31) and *B. subtilis* HMU phages (18).

This study describes the identification of a group I intron within the lysin gene of half of the investigated *S. thermophilus* bacteriophages. The relationship of this intron to other phage introns is discussed.

MATERIALS AND METHODS

Bacterial strains, bacteriophages, and media. *S. thermophilus* strains were routinely subcultured at 42°C in either LM17 (M17 supplemented with 0.5% lactose) (38) or Belliker (Elliker plus 1% beef extract) medium (Difco manual, Difco Laboratories, Detroit, Mich.). The *S. thermophilus* phages used in this study were obtained from the Nestlé Research Centre phage collection. The

* Corresponding author. Mailing address: Nestlé Research Centre, Nestec Ltd., Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland. Phone: 41 21 785 8676. Fax: 41 21 785 8925. E-mail: Harald.Bruessow @rdls.nestle.com.
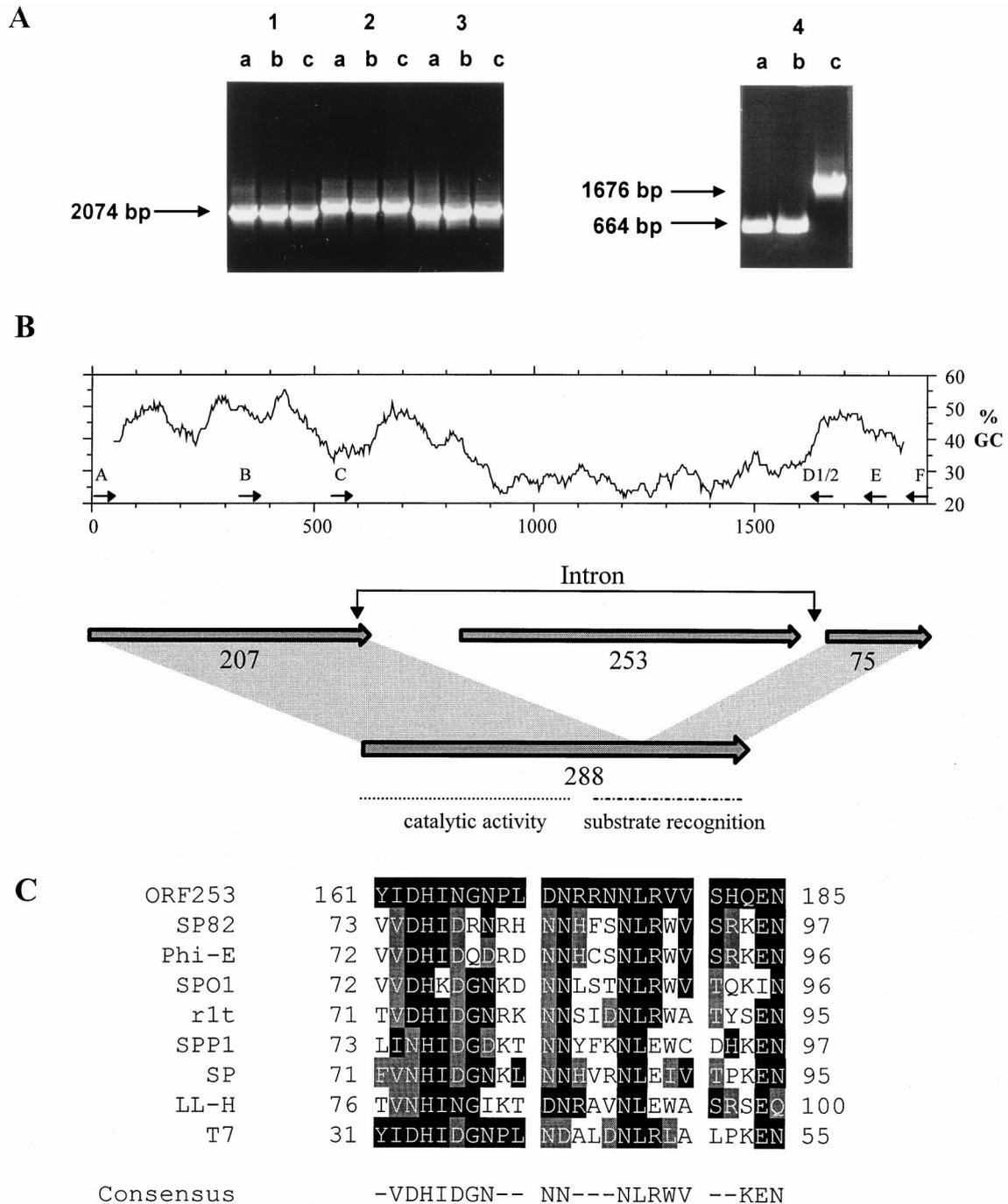
FIG. 1. (A) PCR screening of *S. thermophilus* bacteriophage genomes for insertion or deletion events. PCR amplification products obtained with primer pairs G1-G2 (block 1), H1-H2 (block 2), I1-I2 (block 3), and A-E (block 4) and phage DNAs from Sfi21 (lanes a), S3 (lanes b), and S3b (lanes c) are shown. Molecular sizes of the PCR products are provided for blocks 1 and 4. (B) (Top) Percent GC content distribution of the investigated 1.9-kb region of phage S3b encompassing the lysin gene (window length, 100 bp). Primers referred to throughout the text are indicated by thin arrows above the ruler providing the nucleotide scale in base pairs. (Bottom) Prediction of open reading frames in the region of the phage S3b lysin gene. The open reading frames are indicated by shaded arrows, with their lengths given in amino acids. orf 288, which results from an orf 207-orf 75 fusion, is shown together with the domain structure deduced from knowledge of pneumococcal phage Dp-1 (33). (C) Multiple alignment of a 25-aa block from the S3b orf 253 gene product and the gene product of coliphage T7 gene 3.8 (V01146), the gene product of *Bacillus* phage SPP1 orf 36.1 (X67865), and the intron-encoded gene products of *Bacillus* phages SP82 (U04812), phi-E (U04813), SPO1 (P34081), and SP (L31962), *L. lactis* phage r1t (U38906), and *Lactobacillus* phage LL-H (L37351). Amino acid positions identical to those of S3b are shaded in black, while conserved amino acids are shaded in grey. The consensus sequence indicates identical amino acids present in five or more sequences.

phages were propagated on the appropriate *S. thermophilus* strain in LM17 broth as described previously (4).

DNA techniques. Phage purification, DNA extraction and purification, agarose gel electrophoresis, Southern blot and dot blot hybridizations, and DNA labelling were done as described previously (4).

DNA sequencing and analysis. Phage S3b was sequenced by use of an Amersham Labstation sequencing kit based on Thermo Sequenase-labelled primer cycle sequencing with 7-deaza-dGTP. Sequencing was done on a Licor 6000L automated sequencer with fluorescence-labelled primers.

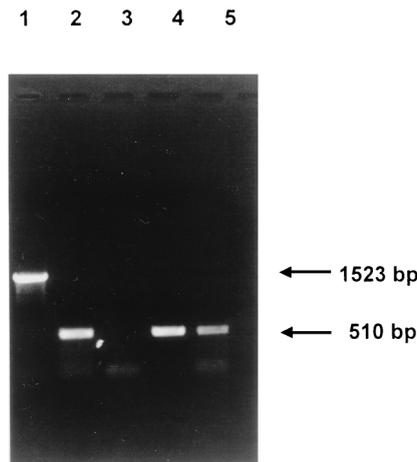PCR products were sequenced on both strands by dideoxy chain termination

FIG. 2. In vivo splicing of S3b intron RNA. Lanes 1 and 4, RT-PCR products obtained with phage S3b and Sfi21 DNAs, respectively, as templates; lanes 2 and 5, RT-PCR products obtained with RNA isolated from S3b- and Sfi21-infected cells, respectively, 15 min after infection; lane 3, PCR product (no reverse transcriptase) obtained with RNA isolated from S3b-infected cells 15 min after infection. Primers B and F (Fig. 1B) were used. The sizes of the products obtained are indicated in base pairs.

with an *fmol* DNA sequencing system (Promega, Madison, Wis.). The sequencing primers were end labelled with $[\gamma\text{-}^{33}P]$ATP according to the manufacturer's instructions. The thermal cycler (Perkin-Elmer) was programmed for 30 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min. The sequence obtained was analyzed as described by Lucchini et al. (23).

PCR. DNA samples were amplified in a Perkin-Elmer thermal cycler programmed for 30 cycles each consisting of 94°C for 30 s, 55°C for 30 s, and 72°C for 1.5 min. Synthetic primers were designed according to established *S. thermophilus* phage DNA sequences and used together with the relevant DNA template and *Taq* polymerase (Fermentas). PCR products were gel purified with Ultrafree-MC Centrifugal Filter Units (Millipore) by following the manufacturer's instructions. The sequences of relevant primers used for PCR and reverse transcription (RT)-PCR are as follows: A (5′ TGT CCC ACA ATC TCT TGT 3′), B (5′ TTG TGA CTA CTC AAC TCA AGG AGC 3′), C (5′ GAA GCC AAT GAA GTC AAA TAC G 3′), D1 (5′ GGT CTG CTC CAT CTG GAA GGT CGT T 3′), D2 (5′ GGT CAG CTC CGT CTG GAA GGT CGT T 3′), E (5′ GGT CTG CTC CAT CTG GAA 3′), and F (5′ GTG GTC TAT TGG TAG TAG TTT ACC 3′). Oligonucleotides G1 and G2, H1 and H2 and I1 and I2 anneal to positions 29400 and 31476, 31159 and 33515, and 33367 and 35460, respectively, of the published phage Sfi21 sequence (GenBank accession no. AF115103). They are all 18 nucleotides long. Oligonucleotides J (5′ GGC AAT ACC GTG CCA AGT C 3′) and K (5′ CCC AAC TTG GAT TCT AGC 3′) were used to generate the S3b intron probe for dot blot hybridization.

RT-PCR. Fifty milliliters of Belliker medium was inoculated with the relevant *S. thermophilus* strain and grown to an optical density at 600 nm of 0.2. $CaCl_2$ (final concentration, 10 mM) was added together with the phage (10^5 to 10^6 PFU) to be tested. Following 15 min of incubation at 42°C, the cultures were harvested by centrifugation. The pellets obtained were washed with diethyl pyrocarbonate-treated water and resuspended in 200 μl of RNase-free 10 mM Tris–1 mM EDTA solution (pH 7.5). The cells were ruptured by agitation for 5 min with 150 μl of RNase-free glass beads (Sigma; 106 μm) in a bead beater at 4°C. The lysate was rapidly centrifuged at 4°C, and the supernatant was recovered.

RNA was isolated from the supernatant with an RNeasy kit (Qiagen) by following the manufacturer's instructions. cDNA was synthesized and amplified with an Access RT-PCR system (Promega) and the following thermal cycling profile: 48°C for 45 min; 94°C for 2 min; 40 cycles at 94°C for 30 s, 55°C for 1 min, and 68°C for 2 min; 68°C for 7 min; and a 4°C soak.

Nucleotide sequence accession numbers. The sequence data for phages S3b, ST3, J, S92, Sfi16A, and ST64 have been submitted to the GenBank database under accession no. AF148561, AF148565, AF148566, AF148563, AF148564, and AF148562, respectively.

RESULTS

Identification of an interrupted lysin gene in phage S3b. Several *S. thermophilus* phages have been entirely sequenced (25, 26, 37, 39). For one of them, phage Sfi21, deletion deriv-

atives and variants have been reported (7). The availability of the Sfi21 genome sequence allowed a rapid characterization of these phages by use of PCR and primers selected such that the entire genome was amplified as overlapping fragments of approximately 2 kb (data not shown). The use of primers A and E, which are located within the Sfi21 putative lysin gene (orf 288) (Fig. 1B), resulted in an amplified DNA product of approximately 1.6 kb for variant phage S3b, in contrast to the expected 664-bp product for Sfi21 and S3 (Fig. 1A). This region of S3b was sequenced; analysis led to the prediction of three similarly oriented open reading frames with coding potential for 207, 253, and 75 amino acids (aa) (Fig. 1B). The proteins predicted for both orf 207 and orf 75 demonstrated high sequence similarity to the putative enzymatic and substrate recognition domains, respectively, of bacteriophage lysins (Fig. 1B).

Database searches yielded no significant matches with the noncoding regions, while the predicted orf 253 gene product showed similarity over the N- and C-terminal halves to an unattributed 54-aa protein of *S. thermophilus* phage DT1 and to the predicted protein for gene 3.8 of coliphage T7, respectively. Closer scrutiny revealed a 25-aa region corresponding to the recently described H-N-H motif found in group I intron endonucleases of phages infecting gram-positive bacteria (34) (Fig. 1C). These data suggest that there is an intron in the lysin genes of *S. thermophilus* phages S3b and DT1.

Identification of the presence of a group I intron. In order to test for in vivo splicing of RNA transcripts, reverse transcription (RT)-PCR was performed with oligonucleotides which specifically anneal to orf 207 and orf 75 of S3b. RNA was isolated from *S. thermophilus* Sfi1 15 min following infection by S3b and used for the synthesis of cDNA and subsequent PCR amplification with primers B and F (Fig. 1B). Phage Sfi21-infected *S. thermophilus* Sfi1 was also included in this assay for comparison. As a control, Sfi21 and S3b phage DNAs were PCR amplified with the same primers. A 510-bp RT-PCR product was obtained for S3b-infected Sfi1, in contrast to the 1,523-bp fragment obtained for S3b phage DNA (Fig. 2A). This result indicates that RNA splicing has occurred, resulting in the excision of 1,013 nucleotides at the level of the RNA precursor. Furthermore, the S3b RT-PCR product is identical in size to that obtained from Sfi21-infected Sfi1.

In order to determine the exact intron boundaries, the RT-PCR product was gel purified and sequenced. Sequence analysis revealed that splicing has occurred, as is typical of group I introns, following uridine and guanosine residues at positions 617 and 1630, respectively, resulting in the excision of a 1,013-bp intron (Fig. 3A). This event results in the generation of an open reading frame with a coding potential for a 288-aa protein. Although a new codon (GUA rather than GUU) is created at the splice junction, the coding potential for valine is conserved. The predicted orf 288 gene product demonstrates >85% identity with the putative lysins of *S. thermophilus* phages (Sfi21, Sfi19, Sfi18, Sfi11, DT1, and O1205) present in the database.

Secondary structure analysis of the phage S3b intron. By exploiting the secondary structure predictions based on a comparative sequence analysis (27) and the recently established three-dimensional structure of a group I intron (16), it was possible to predict a secondary structure for the *S. thermophilus* phage intron (Fig. 3A) (T. R. Cech, personal communication). The *S. thermophilus* phage intron possesses a conserved core of seven base-paired stems (P3 to P9), the integrity of which is essential for the self-splicing activity of introns (9, 10). The additional stem-loop structures P7.1, P7.2a, and P7.2b (located between P7 and P3) indicate the presence of a sub-

FIG. 3. (A) Secondary structure prediction for the S3b intron represented according to the structural convention of Burke et al. (8). Large arrows indicate the 5′ and 3′ splice sites. The exon and intron sequences are denoted by lower- and uppercase letters, respectively. Underlined sequences denoted with P10 indicate the regions which can anneal to form P10. Predicted tertiary interactions are indicated by dotted lines (P4/J6-7 and J3-4/P6) and by circled nucleotides (P5/P9). The shaded nucleotides in P7 represent the putative guanosine binding site. The start and stop codons of orf 253 are boxed. Sequence differences within the structural regions of phage Sfi16A (nucleotide [nt] positions 764 and 789), S92 (nt positions 764 and 791), ST64 (nt positions 764 and 789), and DT1 (nt position 791) introns are indicated by small arrows. The two possible locations for the G residue (nt position 800) are indicated. Nucleotide positions are numbered according to the S3b sequence

group IA2 intron (27). An internal guide sequence (Fig. 3A) which could bring P1 and P10 sequences in proximity to facilitate the splicing process was identified (14). Tertiary interactions were also predicted (P5/P9; joining segments J3-4/P6 and J6-7/P4) (Fig. 3A) (T. R. Cech, personal communication).

Several interesting features were noted for this intron. The *S. thermophilus* phage intron lacks the P2 hairpin, a property shared with phage T4 *sunY* (41). The intron possesses a small catalytic core of about 230 nucleotides starting from the internal guide sequence. A similarly small core has recently been reported for the group I intron of *S. aureus* phage Twort (21). The orf 253 termination codon from phage S3b is located within the intron core, in the 3′ portion of the P8 stem, which is critical for catalysis (Fig. 3A). Similar observations have been made for the three T4 introns and the SPO1 intron (17, 35). Due to the location of the stop codon, a link between translation and the splicing and expression of genes containing these introns has been suggested (35).

Two further regions of the S3b intron showed a relationship to the intron in the DNA polymerase gene of *B. subtilis* phage SPO1 and its derivatives (18). First, the region between P3 and P4, which is characteristically only a few nucleotides long, has a long extension in both SPO1 and S3b. This structure has two stems flanked by short single-stranded regions. As shown in Fig. 3B, the size and sequence of these regions are identical between SPO1 and S3b, indicating a possible relationship between the respective introns. Second, an alternative structure that resembles the structure suggested for the phage SPO1 intron can be proposed for P7.2 (Fig. 3C).

The GC content of 41% for the noncoding intron sequences did not differ from the value of 43% for the surrounding phage sequences. In contrast, orf 253, located in the intron, had a remarkably low GC content of 28%, clearly indicating different origins for the intron coding and noncoding segments.

Distribution of introns within lysin genes of *S. thermophilus* phages. Our *S. thermophilus* phage collection, which comprises isolates from different ecological settings in France, Italy, Germany, Switzerland, and Austria, was screened for the presence of an interrupted lysin gene. Of 61 phages tested, 31 yielded a PCR product larger than expected for a noninterrupted lysin gene. Twenty-seven of these phages yielded a fragment similar in size to that obtained for S3b, while smaller PCR amplification products were obtained for phages Sfi16A, S92, S93, and ST64 (Fig. 4A). It should be noted that the presence or absence of the lysin intron could not be correlated with the ecological context of the streptococcal phages (data not shown).

Phages which yielded a PCR result indicative of an intron sequence were all found positive by dot blot hybridization when primers J and K were used to generate S3b intron DNA as a probe (Fig. 4B). Conversely, all phages which yielded PCR products that comigrated with the phage Sfi21 signal were negative with this probe in dot blot hybridization experiments. Southern hybridizations were performed with a selected number of phages possessing the S3b-type intron with S3b intron DNA as a probe. When a restriction enzyme which does not cut within the intron sequence was used, a single hybridization signal which corresponded to the lysin-located intron was obtained (Fig. 4C). These experiments indicated that the S3b-type intron is not located in areas of the *S. thermophilus* bacteriophage genome other than the lysin gene.

The genomic DNAs of 33 *S. thermophilus* strains (including raw milk isolates and yogurt and mozzarella cheese starter cultures) were examined for evidence of the presence of an S3b-type intron by dot blot analysis. No hybridization signals were obtained when a probe consisting of S3b intron DNA was used (Fig. 4B), indicating the absence of this group I intron in these strains.

Total RNA was isolated from *S. thermophilus* cultures infected with phages S92 and ST64. RT-PCR analysis yielded products which were smaller than those obtained with the respective phage DNAs as templates and which comigrated with the RT-PCR product obtained from Sfi21-infected cells (Fig. 5). This result confirms the presence of an intron in both phages.

The variant introns have deletions in the intron open reading frame. The variant introns detected in phages Sfi16A, S92, and ST64 were sequenced. Phages ST3 and J, which have a lysin gene interruption similar in size to that of phage S3b, were also chosen randomly for sequence comparison. Analysis revealed that the introns of phages J and ST3 are identical, while those of ST3 and S3b differ by one nucleotide substitution. Sequence alignment indicated that phages Sfi16A, S92, and ST64 possess introns of 519, 443, and 316 bp, respectively. The phage DT1 intron differs from that of phage S92 by one base-pair substitution. The size differences between the variant introns were caused by deletion events within orf 253 (Fig. 3D). Since RNA splicing was observed for the variant introns (Fig. 5), the orf 253 gene product is dispensable for this process.

An alignment of the ST64 and S3b introns revealed that each extremity of the S3b intron region, absent in ST64, was preceded by an identical 6-bp motif (Fig. 3D). Since only one of the repeats was retained in ST64, the deletion event probably resulted from slippage of the DNA polymerase. A similar observation was made by Eddy and Gold (15) for the nonmobile *nrdB* intron of bacteriophage T4. No such repeats flanked the deletion sites in the phage S93, S92, DT1, and Sfi16A introns. However, all of these phages share the same deletion start site, located in codon 55 of orf 253, and S93, S92, and DT1 share the same deletion stop site (Fig. 3D). Furthermore, the four phages are not ecologically or genetically related in that they possess nonoverlapping host ranges and distinct restriction patterns and are of diverse geographical origins. Phages S92, S93, and DT1 are mozzarella isolates from Italy, Italy, and Canada, respectively, while Sfi16A originates from a French yogurt.

Apart from the deletion event, alignment of the variant introns with that of S3b revealed the presence of nearly identical intron sequences. The highest diversification was seen between S3b and Sfi16A, which differed in 15 nucleotide positions. The secondary structures of the variant introns do not differ greatly from that of S3b, since the majority of the differences are located in the looped-out region of P8. In contrast to the high degree of sequence conservation within the intron sequences, substantial sequence diversity was detected over the

(accession no. AF148561). The numbering begins with the start codon of orf 207. (B) Folding of the joining region between P3 and P4 (J3-4) of the S3b intron and the intron of *B. subtilis* phage SPO1. The sequence difference between both the phage ST64 and Sfi16A introns and the S3b intron is boxed. A belongs to phage S3b. (C) Alternative folding of a P7.2 stem for the phage S3b intron in comparison with the P7.2 stem for the phage SPO1 intron. (D) Comparison of the deletion points of the variant introns of phages ST64, DT1, S92, S93, and Sfi16A. The sequence given is that of S3b. Bent arrows indicate the region which is deleted in the respective phages. The 6-bp direct repeat flanking the deletion in the phage ST64 intron is boxed, and the stop codon of orf 253 is underlined. The GTAG sequence matching the intron insertion site (see Fig. 6) is doubly underlined in variant introns of phages DT1 and Sfi16A.
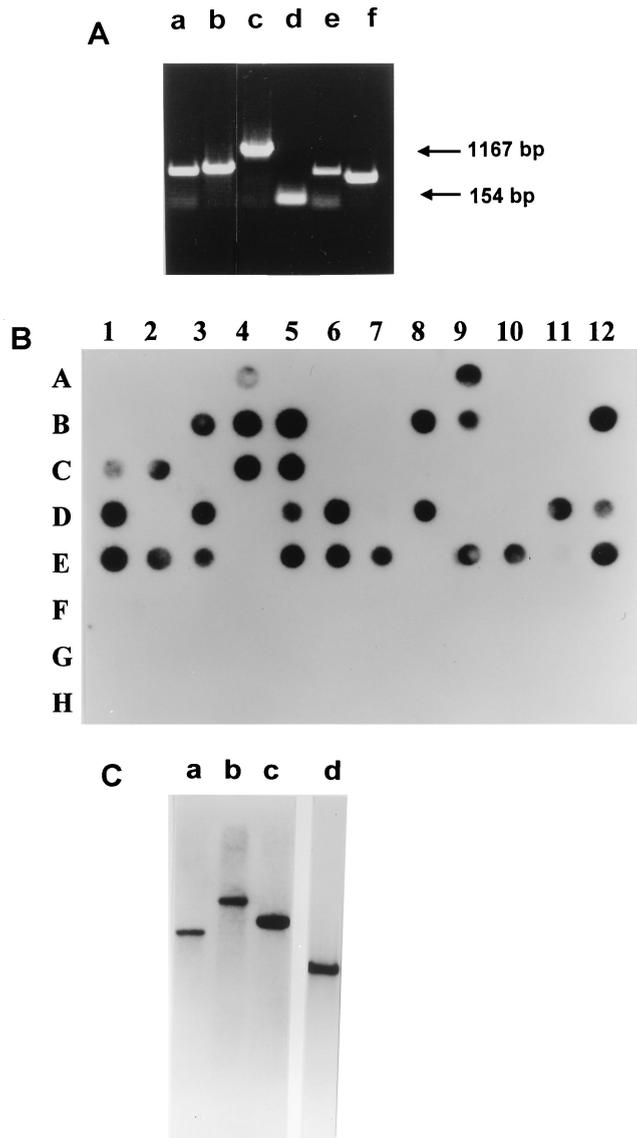
FIG. 4. Screening of *S. thermophilus* bacteriophages and starter cultures for the presence of the S3b-type intron. (A) PCR amplification with primer pair C-D1 (or, alternatively, C-D2 in the case of nonannealing of D1) and the following phage DNAs as templates: S93, Sfi16A, S3b, Sfi21, S92, and ST64 (lanes a to f, respectively). The sizes of the DNA products obtained for S3b and Sfi21 are shown in base pairs. (B) Dot blot of *S. thermophilus* bacteria and bacteriophage DNA with S3b intron DNA as a probe. (A) Lanes 1 to 12, Sfi phages 21, 11, 13, 16A, 18, 19, and 121 and HN phages 1, 4, 6, 7, and 9. (B) Lanes 1 to 12, ST phages 1, 2, 3, 9, 12, 13, 20, 30, 40, 41, 42, and 44. (C) Lanes 1 to 5, ST phages 64, 72, 84, 128, and J1; lanes 7 to 12, TN starter cultures 2, 17, 36, 44, 21, and 62. (D) Lanes 1 to 12, S phages 3b, 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, and 12. (E) Lanes 1 to 10 and 12, phages 13, 14, 15, 17, 18, 19, 56, 66, 77, 89, and 94. (F) Lanes 1 and 2, S phages 96 and 97; lanes 6 to 12, TN starter cultures 49, 70, 39, 41, 42, 45, and 58. (G) Lanes 1 to 12, Sfi starter cultures 1, 2, 3, 6, 8, 10, 11, 13, 16, 18, 19, and 20. (H) Lanes 1 to 5, Sfi starter cultures 25, 26, 39, 41, and IL; lanes 7 to 9, S starter cultures 3, 17, and 97. (C) Southern hybridization with the phage S3 intron DNA as a probe. Lanes a to c, *Pvu*II-digested genomic DNAs from phages S18, S89, and S94; lane d, *Hin*dIII-digested DNA from phage S3b. Molecular sizes for the signals in lanes a to d were approximately 8.5, 12, 7, and 3.5 kb, respectively.

adjacent lysin-encoding sequences (data not shown). For example, alignment of the 1-kb intron sequences revealed differences of 3 and 1 nucleotides for the S3b-DT1 and S3b-ST3 comparisons, respectively, while alignment of the 0.25-kb re-
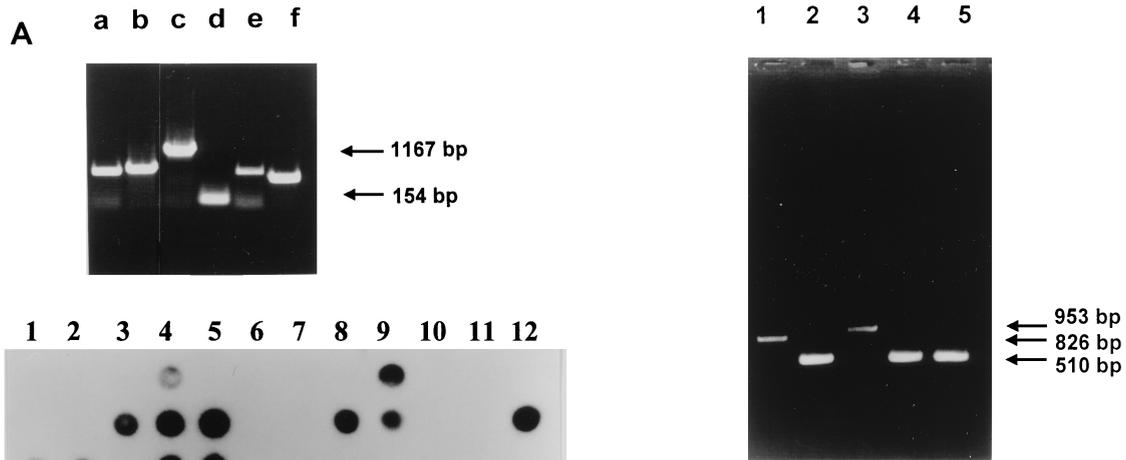

FIG. 5. *In vivo* splicing of phage ST64 and S92 intron RNAs. Lanes 2, 4, and 5, RT-PCR products obtained with RNA isolated from ST64-, S92-, and Sfi21-infected cells 15 min after infection; lanes 1 and 3, RT-PCR products obtained with phage ST64 and S92 DNAs as templates. The oligonucleotide pair B-E (Fig. 1B) was used. The sizes of the products obtained are indicated in base pairs.

gion 3′ of the intron revealed 24 and 54 nucleotide differences for the S3b-DT1 and S3b-ST3 comparisons, respectively.

Comparison of sequences surrounding the splice site. The DNA sequence of a 40-bp region surrounding the intron insertion site of 10 phages was compared with the equivalent region of 10 intron-free *S. thermophilus* lysin genes (Fig. 6). The sequence alignment demonstrated that phages possessing the intron and those lacking it differed clearly in a 14-bp segment centered around the intron insertion point. A screening of the five complete phage genome sequences available indicated that no unoccupied 14-bp sequence was present. The nearest matches were two segments with 12-bp identity (nucleotide substitutions at the −2 G and +1 A positions), possibly reflecting the critical nature of these nucleotide positions for intron homing and/or splicing (data not shown).

DISCUSSION

This study describes a group I intron present within the lysin gene of over half of the investigated *S. thermophilus* bacteriophages. No phenotypic or ecological differences were observed between *S. thermophilus* phages possessing an intron-containing lysin gene or an intron-free lysin gene. The possession of the intron is thus either without selective advantage or is used in subtle regulatory circuits which do not detectably affect the laboratory growth or environmental distribution of *S. thermophilus* phages. A comparison of the intron insertion site for intron-containing and intron-free lysin genes led to the identification of a 14-bp consensus sequence associated with intron possession. Since no unoccupied consensus sequence was identified, one might deduce that intron invasion has reached a saturation point in *S. thermophilus* phages, thereby suggesting the presence of a very invasive molecular parasite. However, the consensus sequence could also be a consequence of the homing event if it causes gene conversion of the flanking exon sequences.

The intron itself was apparently molecularly parasitized by orf 253, which has sequence similarity to H-N-H-type endonuclease genes commonly found in introns from bacteriophages infecting gram-positive bacteria (34). This gene conferred intron mobility in *L. lactis* phage r1t (29). The invasive character of orf 253 was deduced from its distinct GC content. On the

```
                                        Intron
            55                                                    1650
S3b     CCAATTGGCTTTGACTGGGT ▼ AGAAAATGGAATCCCAGTTG
ST3     ...................    ....................
J       ...................    ....................
S92     ........G..C........    ..........TG........T
ST64    ...G....A..C........    ...................A
DT1     ........G..C........    ..........TG........T
S15     ..TG....A..C........    ................G...A
S19     ..TG....A..C........    ................G...T
S56     ..TG....G..C........    ......C..TG.........
ST72    ...................    ................G...T
Sfi21   ........G.....T.ATT.    G.....C........G...A
Sfi18   ........G.....T.ATT.    G.G...C........G...A
Sfi19   ........G.....T.ATT.    G.....C........G...A
Sfi11   ........G.....T.ATT.    G.....C........G...A
O1205   ........A..C..T...C.    TC.G..C..GG...CG...T
S3a     ........G.....T.ATT.    G.....C........G...A
S7      .....C..A.....T.ATT.    CG....T........TA.GC
S96     ..T..C..G...A.T.ATT.    G........AA....G...T
ST41    ..C..C..A..CA.T.ATT.    TC....T..GA.......A.
ST42    ........G.....T.ATT.    G........AA....G...A
```
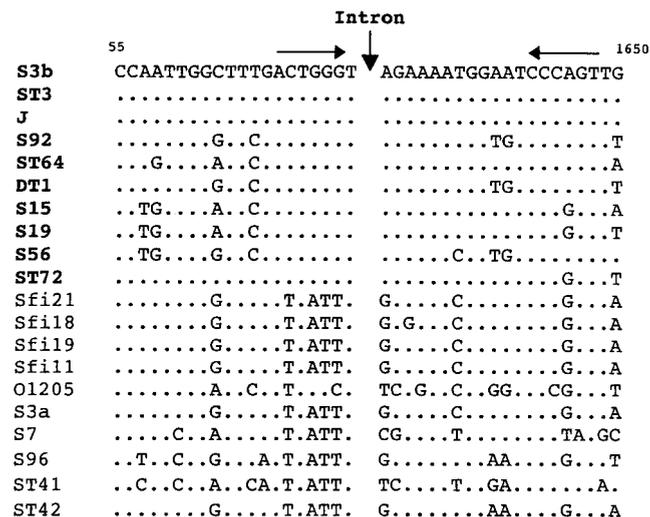
FIG. 6. Nucleotide sequence alignment of the region surrounding the splice site of the intron-containing lysin genes (in bold) with the equivalent region from intron-free lysin genes. Sequence differences relative to the phage S3b sequence are indicated. The intron insertion site is indicated by a vertical arrow. The horizontal arrows represent a 6-bp inverted repeat. The nucleotide positions indicated are based on the numbering of the S3b sequence (accession no. AF148561).

basis of theoretical consideration, one would expect pressure to remove the intron (13), and indeed five phages possess large deletions within orf 253. Examination of the DNA sequence indicated the presence of a deletion hot spot, i.e., codon position 55 in orf 253, in three of the five introns. The partial removal of the intron-bracketed orf 253, but maintenance of the intron in the phage genome is reminiscent of a similar process in coliphage T4 (15) and may be explained by the difficulty in removing the intron precisely without compromising lysin gene expression.

The initial discovery of self-splicing introns in coliphage T4 (11, 12) has attracted much interest among biologists interested in the evolutionary origin of introns (2, 19, 32, 40). The discovery of a group I intron in *B. subtilis* phage SPO1 was taken as evidence favoring the antiquity of introns in phages (17). To the discovery of additional type I introns in phages infecting the bacterial genera *Bacillus*, *Lactococcus*, *Lactobacillus*, and *Staphylococcus* can now be added the discovery of a type I intron in phages infecting the genus *Streptococcus*. All of these bacteriophages infect bacteria belonging to the low-GC-content branch of gram-positive bacteria. Comparative sequence analysis has demonstrated that temperate *Siphoviridae* from this branch of gram-positive bacteria are also evolutionarily related (23, 24).

The intron of *S. thermophilus* phages did not share significant overall sequence similarity with the introns from other phages. It is, however, premature to exclude a common origin for phage introns from that observation. In fact, the *S. thermophilus* phage intron looks like a typical subgroup IA2 intron, like many other phage introns. In addition, the secondary structure prediction provided strong evidence for a relationship to the introns in other phage systems. First, the intron of *S. thermophilus* phages possesses a 38-bp insert between the 5′ portions of P3 and P4 which can be folded into two stable stem-loop structures (Fig. 3C). In most group I introns, this region comprises a few nucleotides (typically 3 or 4). The only group IA introns which contain extra nucleotides in this region are those of *Bacillus* phages SPO1, SP82, and E, which possess

69 nucleotides in this interval folded into two thermodynamically stable stem-loop structures (18). Second, the location of the stop codon for orf 253 within the P8 stem is shared not only with the *B. subtilis* phage SPO1 intron but also with the three introns of *E. coli* bacteriophage T4 (17, 35). Third, the P7.2 region is highly similar in all phage introns described so far. It is always a perfect base-paired stem that is separated from P7.1 by the sequence GUA. Fourth, the intron-encoded proteins from *Lactococcus*, *Streptococcus*, *Lactobacillus*, and *Bacillus* phages have sequence similarity over the H-N-H motif. Fifth, the S3b intron lacks the P2 hairpin, a property shared with the phage T4 *sunY* intron.

Taken together, these observations suggest a certain conservation of structural features in phage introns. Whether this conservation reflects common ancestry or functional constraints of phage introns cannot currently be decided. We suspect that introns are much more widely distributed in phages than initially anticipated. The identification of phage introns from further genera of gram-positive bacteria could shed light on the evolutionary origin of phage introns.

Another observation is worth noting. While all group I introns of bacteria and eucaryotic organelles have been localized to the anticodon position of tRNA genes, phage introns have been detected in a number of distinct genes (DNA polymerase, thymidylate synthase, ribonucleotide reductase, structural proteins, large-subunit terminase, and now lysin), demonstrating flexibility in the intron homing or invasion process between phages.

REFERENCES

1. **Bechhofer, D. H., K. K. Hue, and D. A. Shub.** 1994. An intron in the thymidylate synthase gene of *Bacillus* bacteriophage β22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame. Proc. Natl. Acad. Sci. USA **91:**11669–11673.
2. **Belfort, M.** 1991. Self-splicing introns in prokaryotes: migrant fossils? Cell **64:**9–11.
3. **Biniszkiewicz, D., E. Cesnaviciene, and D. A. Shub.** 1994. Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. EMBO J. **13:**4629–4635.
4. **Brüssow, H., and A. Bruttin.** 1995. Characterization of a temperate *Streptococcus thermophilus* bacteriophage and its genetic relationship with lytic phages. Virology **212:**632–640.
5. **Brüssow, H., A. Bruttin, F. Desiere, S. Lucchini, and S. Foley.** 1998. Molecular ecology and evolution of *Streptococcus thermophilus* bacteriophages—a review. Virus Genes **16:**95–109.
6. **Brüssow, H.** 1999. Phages of *Streptococcus thermophilus*, p. 1253–1262. *In* R. Webster and A. Granoff (ed.), Encyclopedia of virology, vol. 2. Academic Press Ltd., London, United Kingdom.
7. **Bruttin, A., and H. Brüssow.** 1996. Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. Virology **219:**96–104.
8. **Burke, J. M., M. Belfort, T. R. Cech, R. W. Davies, R. J. Schweyen, D. A. Shub, J. W. Szostak, and H. F. Tabak.** 1987. Structural conventions for group I introns. Nucleic Acids Res. **15:**7217–7221.
9. **Cech, T. R.** 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene **73:**259–271.
10. **Cech, T. R.** 1990. Self-splicing of group I introns. Annu. Rev. Biochem. **59:**543–568.
11. **Chu, F. K., G. F. Maley, F. Maley, and M. Belfort.** 1984. Intervening sequence in the thymidylate synthase gene of bacteriophage T4. Proc. Natl. Acad. Sci. USA **81:**3049–3053.
12. **Chu, F. K., G. F. Maley, D. K. West, M. Belfort, and F. Maley.** 1986. Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. Cell **45:**157–166.
13. **Darnell, J. E., and W. F. Doolittle.** 1986. Speculations on the early course of evolution. Proc. Natl. Acad. Sci. USA **83:**1271–1275.

14. **Davies, R. W., R. B. Waring, J. A. Ray, T. A. Brown, and C. Scazzocchio.** 1982. Making ends meet: a model for RNA splicing in fungal mitochondria. Nature **300:**719–724.

15. **Eddy, S. R., and L. Gold.** 1991. The phage T4 *nrdB* intron: a deletion mutant of a version found in the wild. Genes Dev. **5:**1032–1041.

16. **Golden, B. L., A. R. Gooding, E. R. Podell, and T. R. Cech.** 1998. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. Science **282:**259–264.

17. **Goodrich-Blair, H., V. Scarlato, J. M. Gott, M.-Q. Xu, and D. A. Shub.** 1990. A self-splicing group I intron in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. Cell **63:**417–424.

18. **Goodrich-Blair, H., and D. A. Shub.** 1994. The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. Nucleic Acids Res. **22:**3715–3721.

19. **Kuhsel, M. G., R. Strickland, and J. D. Palmer.** 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science **250:**1570–1573.

20. **Lambowitz, A. M., and M. Belfort.** 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. **62:**587–622.

21. **Landthaler, M., and D. A. Shub.** 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc. Natl. Acad. Sci. USA **96:**7005–7010.

22. **Lazarevic, V., B. Soldo, A. Dusterhoft, H. Hilbert, C. Mauel, and D. Karamata.** 1998. Introns and intein coding sequences in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPβ. Proc. Natl. Acad. Sci. USA **95:**1692–1697.

23. **Lucchini, S., F. Desiere, and H. Brüssow.** 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage φSfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts. Virology **246:**63–73.

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

25. **Lucchini, S., F. Desiere, and H. Brüssow.** 1999. The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: whole genome sequence comparisons of *cos*-site phages Sfi19 and Sfi21. Virology **260:**232–243.

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

27. **Michel, F., and E. Westhof.** 1990. Modelling of the three-dimensional archi-tecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. **216:**585–610.

28. **Mikkonen, M., and T. Alatossava.** 1995. A group I intron in the terminase gene of *Lactobacillus delbrueckii* subsp. *lactis* phage LL-H. Microbiology **141:**2183–2190.

29. **Nauta, A.** 1997. Molecular characterization and exploitation of the temperate *Lactococcus lactis* bacteriophage r1t. Ph.D. thesis. University of Groningen, Groningen, The Netherlands.

30. **Pedersen-Lane, J., and M. Belfort.** 1987. Variable occurrence of the *nrdB* intron in the T-even phages suggests intron mobility. Science **237:**182–183.

31. **Quirk, S. M., D. Bell-Pedersen, J. Tomaschewski, W. Rüger, and M. Belfort.** 1989. The inconsistent distribution of introns in the T-even phages indicates recent genetic exchanges. Nucleic Acids Res. **17:**301–315.

32. **Reinhold-Hurek, B., and D. A. Shub.** 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. Nature **357:**173–176.

33. **Sheehan, M. M., J. L. García, R. López, and P. García.** 1997. The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergenic origin. Mol. Microbiol. **24:**717–725.

34. **Shub, D. A., H. Goodrich-Blair, and S. R. Eddy.** 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. Trends Biochem. Sci. **19:**402–404.

35. **Shub, D. A., J. M. Gott, M.-Q. Xu, B. F. Lang, F. Michel, J. Tomaschewski, J. Pedersen-Lane, and M. Belfort.** 1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proc. Natl. Acad. Sci. USA **85:**1151–1155.

36. **Shub, D. A., M.-Q. Xu, J. M. Gott, A. Zeeh, and L. D. Wilson.** 1987. A family of autocatalytic group I introns in bacteriophage T4. Cold Spring Harbor Symp. Quant. Biol. **52:**193–200.

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

38. **Terzaghi, B. E., and W. E. Sandine.** 1975. Improved media for lactic streptococci and their bacteriophages. Appl. Microbiol. **29:**807–813.

39. **Tremblay, D. M., and S. Moineau.** 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. Virology **255:**63–76.

40. **Xu, M.-Q., S. D. Kathe, H. Goodrich-Blair, S. A. Nierzwicki-Bauer, and D. A. Shub.** 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. Science **250:**1566–1570.

41. **Xu, M.-Q., and D. A. Shub.** 1989. The catalytic core of the *sunY* intron of bacteriophage T4. Gene **82:**77–82.