NOTES

A Group IIC-Type Intron Interrupts the rRNA Methylase Gene of *Geobacillus stearothermophilus* Strain 10[∇]

Samuel E. Moretz[1] and Bert C. Lampson[2]*

Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12735 Twinbrook Parkway, Rockville, Maryland 20852,[1] and Department of Health Sciences, Box 70673, East Tennessee State University, Johnson City, Tennessee 37614[2]

Group IIC introns insert next to the stem-loop structure of rho-independent transcription terminators, thus avoiding intact genes. The insertion sites of 17 copies of the G.st.I1 intron from *Geobacillus stearothermophilus* were compared. One copy of the intron was found to interrupt an open reading frame (ORF) encoding an rRNA methylase.

Reverse transcriptase (RT)-encoding retroelements are an increasingly diverse collection of DNAs found in bacterial genomes (4, 14). The most widely dispersed and most frequently encountered type of retroelement in bacterial genomes is the group II intron. Other types of bacterial retroelements include the multicopy single-stranded DNA (msDNA)-producing retrons (5) and the diversity-generating retroelements (9).

Group II introns are self-splicing DNAs that code for a functional RT and also produce a functional RNA ribozyme. An mRNA containing a copy of the intron can fold into a catalytically active secondary structure and splice itself out of the mRNA molecule, producing fused 5′ and 3′ exons and a free intron lariat. The multifunctional intron-encoded protein (IEP) greatly aids splicing by binding to the intron RNA and stabilizing the catalytic region during folding of the ribozyme (3, 15).

Group II introns of bacteria are also mobile DNAs that use a reverse transcription step catalyzed by the RT activity provided by the IEP. In the best-characterized mechanism, the IEP binds to the intron RNA lariat after splicing and forms a ribonucleoprotein (RNP) complex. The RNP complex then recognizes a specific "homing" site in a DNA target via base pairing between the intron RNA (exon binding sites [EBS]; EBS1 and EBS2 motifs) and DNA target exons (intron binding sites [IBS]; IBS1 and IBS2 motifs). This is followed by reverse splicing of the intron RNA into the top strand of the DNA target. The mobility process is completed when the bottom DNA strand at the target is cleaved by an endonuclease, also provided by the IEP. This cleaved bottom strand serves as a primer for the reverse transcription of the inserted intron RNA, thus copying the intron back into DNA (3, 15, 16).
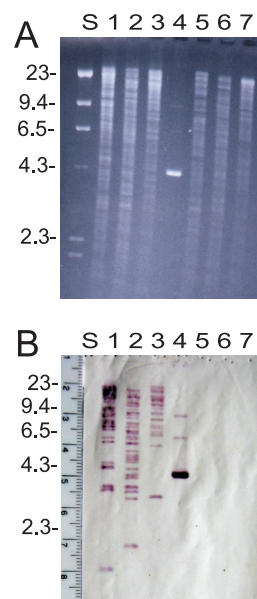

FIG. 1. Southern (DNA) hybridization of chromosome DNA from *G. stearothermophilus*, with the cloned IEP gene of the G.st.I1 intron as a probe. (A) Chromosome DNA from *G. stearothermophilus* strain 10 (lanes 1 to 3) and *G. stearothermophilus* strain ATCC 12980 (lanes 5 to 7) was digested to completion with PstI (lanes 1 and 5), HindIII (lanes 2 and 6), or BamHI (lanes 3 and 7). The digested chromosome was separated via agarose gel electrophoresis and stained with ethidium bromide. Lambda DNA digested with HindIII served as a size standard (lane S), with numbers on the left in kilobase pairs. The plasmid pTrt#16, containing a cloned copy of the IEP gene, was used as a positive control (lane 4) (21). (B) DNA from the agarose gel (A) was transferred to a Nytran filter via the Southern blot method. PCR DNA of the IEP gene from intron G.st.I1 was labeled with DIG and used as a hybridization probe. The DNA probe (IEP gene) hybridizes to many discrete restriction fragments in the chromosome of strain 10 (lanes 1 to 3). No hybridization signal appears against the chromosome DNA of ATCC strain 12980 (lanes 5 to 7).

* Corresponding author. Mailing address: Dept. of Health Sciences, Box 70673, East Tennessee State University, Johnson City, TN 37614. Phone: (423) 439-4572. Fax: (423) 439-4562. E-mail: Lampson@etsu.edu.
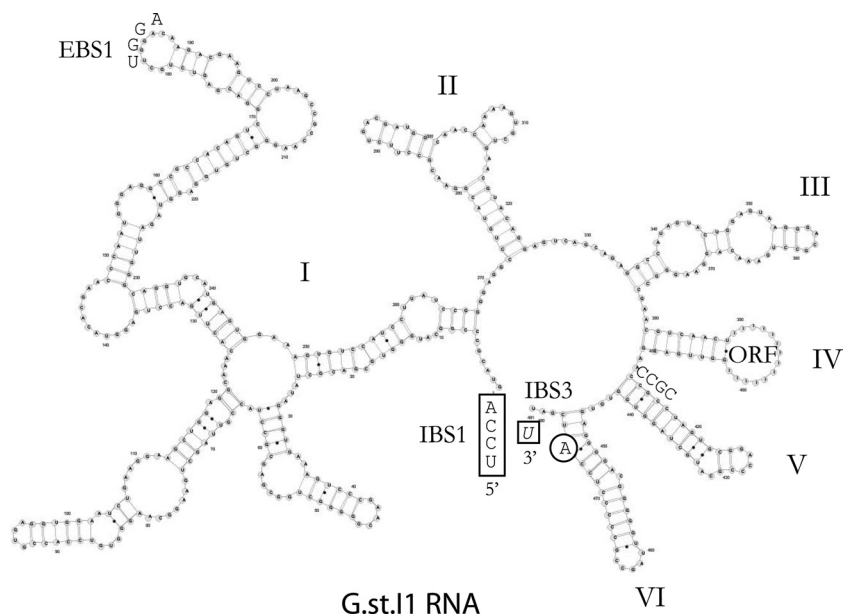
FIG. 2. The secondary structure of the G.st.I1 intron RNA is indicative of the group IIC-type introns. The DNA sequence of intron copy 12B43 was used to infer the structure of the intron RNA ribozyme. The six conserved stem-loops (domains) present in all group II introns are designated with large roman numerals. Atypical structures commonly seen in group IIC introns include the following. The IBS1-EBS1 base pairing interaction between exon and intron regions involves a short 4-bp interaction. Also, there appears to be an absence of an EBS2 pairing region in this intron. In addition, the conserved domain V is composed of a short 12-bp stem and an atypical CCGC sequence at the base of this stem-loop structure. The other structures shown include the "bulging A" (circled A), where the spliced lariat RNA will form, and the ORF, which designates the very large region of domain IV containing the IEP gene sequence not shown here.

Many IEPs of bacterial group II introns do not contain an endonuclease function and thus do not integrate into a specific homing site. One of the most proficient of these non-site-specific introns is the bacterial subclass C-type element (group IIC). Bacterial group II introns are divided into subclasses A to E based on the phylogenetic relationships of their IEPs and the structure of the folded intron ribozyme (13, 19). After mobilization, group IIC-type introns insert into intergenic regions of the chromosome adjacent to rho-independent transcription terminators (10). This is because the RNP complex recognizes the stem-loop structure that may form in the DNA target at transcription terminators or at *attC* sites of integrons (6, 8) rather than at a specific DNA sequence. Thus, group IIC introns can insert nonspecifically at the end of transcription units and avoid disruption of vital host genes (12).

Multiple copies of the G.st.I1 intron. Recently, a group II intron-like RT was isolated from the thermophile *Geobacillus stearothermophilus* and shown to display RT activity at hot temperatures (21). To clone the entire intron plus flanking DNA, the previously cloned IEP gene was used as a probe in a Southern hybridization experiment. Chromosome DNA from *G. stearothermophilus* strain 10 (17) and ATCC strain 12980 was digested to completion with a restriction endonuclease (HindIII, EcoRI, or PstI), separated by gel electrophoresis, transferred to a Nytran membrane, and hybridized with a digoxigenin (DIG)-labeled DNA probe consisting of the IEP gene. The DNA probe hybridized to numerous (around 20) discrete restriction fragments in the chromosome of strain 10 (Fig. 1B, lane 2). No hybridization signal was detected against the chromosome of ATCC strain 12980 (Fig. 1B, lanes 5 to 7). The hybridization experiment indicates that perhaps 20 or

more copies of the G.st.I1 intron are present at different locations on the chromosome of strain 10 but absent from the ATCC strain.

Structure of intron RNA. The DNA sequence of the G.st.I1 intron was used to infer the folded structure of the intron RNA ribozyme by comparison with a known, closely related intron, B.h.I1, from *Bacillus halodurans* (19). As shown in Fig. 2, the G.st.I1 intron RNA can fold into a complex series of stem-loop structures designated I through VI (11). However, this intron RNA has an atypical structure that is commonly seen in a group IIC-type intron. For example, the highly conserved domain V contains a shorter stem of 12 bp rather than the more typical 14-bp stem in this stem-loop domain. In addition, the conserved sequence at the base of the stem in domain V is CCGC rather than the more common RAGC (19). Also atypical for most introns, the G.st.I1 intron contains a short IBS1-EBS1 pairing (4 bp rather than the usual 6 bp). Also, there appears to be a complete absence of the EBS2 pairing site in the G.st.I1 intron RNA and the IBS3-EBS3 pair is only a single base (Fig. 2). All of these abbreviated structures have been observed in well-characterized subclass C-type introns (12, 20).

Insertion occurs adjacent to transcription terminators. The 5′ and 3′ exon DNAs of 14 different copies of the G.st.I1 intron (designated b through m, o, and p) (Fig. 3) were derived from BLAST searches of the unfinished genome sequence of strain 10 (7). These DNA sequences were compared with three cloned copies of this intron (12B43, 8A6, and 9A22) produced from this study. These three cloned copies were obtained from a plasmid library containing HindIII restriction fragments of the strain 10 chromosome and recovered via colony hybridization, using the IEP gene as a probe. The aligned DNA se-
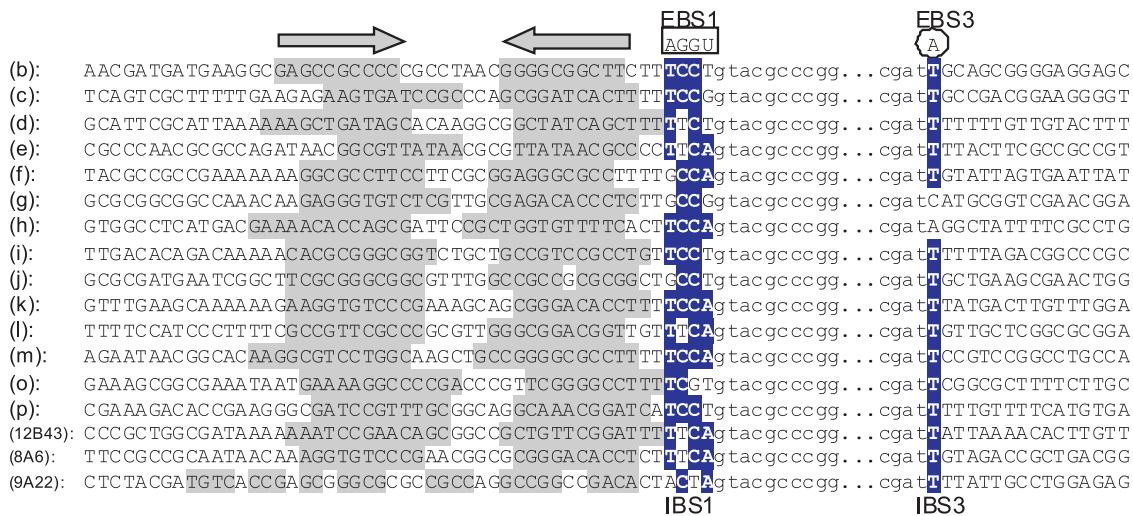
EBS1   EBS3
AGGU   A

```
(b):     AACGATGATGAAGGCGAGCCGCCCCCGCCTAACGGGGCGGCTTCTTTCCTgtacgcccgg...cgatTGCAGCGGGGAGGAGC
(c):     TCAGTCGCTTTTTGAAGAGAAGTGATCGCCCAGCGGATCACTTTTCCGgtacgcccgg...cgatTGCCGACGGAAGGGGT
(d):     GCATTCGCATTAAAAAAGCTGATAGCACAAGGCGGCTATCAGCTTTTCTgtacgcccgg...cgatTTTTTTGTTGTACTTT
(e):     CGCCCAACGCGCCAGATAACGGCGTTATAACGCGTTATAACGCCCCTTCAgtacgcccgg...cgatTTTACTTCGCCGCCGT
(f):     TACGCCGCCGAAAAAAAGGCGCCTTCCTTCGCGGAGGGCGCCTTTTGCCAgtacgcccgg...cgatTGTATTAGTGAATTAT
(g):     GCGCGGCGGCCAAACAAGAGGGTGTCTCGTTGCGAGACACCCTCTTGCCGgtacgcccgg...cgatCATGCGGTCGAACGGA
(h):     GTGGCCTCATGACGAAAACACCAGCGATTCCGCTGGTGTTTTCACTCCAgtacgcccgg...cgatAGGCTATTTTCGCCTG
(i):     TTGACACAGACAAAAACGCGGGCGGTCTGCTGCCGTCCGCCTGTTCCTgtacgcccgg...cgatTTTTTAGACGGCCCGC
(j):     GCGCGATGAATCGGCTTCGCGGGCGGCGTTTGGCCGCCGCGCGGCTGCCTgtacgcccgg...cgatTGCTGAAGCGAACTGG
(k):     GTTTGAAGCAAAAAAGAAGGTGTCCCGAAAGCAGCGGGACACCTTTTCCAgtacgcccgg...cgatTTATGACTTGTTTGGA
(l):     TTTTCCATCCCTTTTCGCCGTTCGCCGCGCGTTGGGCGGACGGTTGTTTCAgtacgcccgg...cgatTGTTGCTCGGCGCGGA
(m):     AGAATAACGGCACAAGGCGTCCTGGCAAGCTGCCGGGGCGCCTTTTCCAgtacgcccgg...cgatTCCGTCCGGCCTGCCA
(o):     GAAAGCGGCGAAATAATGAAAAGGCCCCGACCCGTTCGGGGCCTTTCGTgtacgcccgg...cgatTCGGCGCTTTTCTTGC
(p):     CGAAAGACACCGAAGGGCGATCCGTTTGCGGCAGGCAAACGGATCATCCTgtacgcccgg...cgatTTTTGTTTTCATGTGA
(12B43): CCCGCTGGCGATAAAAAAATCCGAACAGCGGCCGCTGTTCGGATTTTTCAgtacgcccgg...cgatTATTAAAACACTTGTT
(8A6):   TTCCGCCGCAATAACAAAGGTGTCCCGAACGGCGCGGGACACCTCTTTCAgtacgcccgg...cgatTGTAGACCGCTGACGG
(9A22):  CTCTACGATGTCACCGAGCGGGCGCGCCGCCAGGCCGGCCGACACTACTAgtacgcccgg...cgatTTTATTGCCTGGAGAG
```
IBS1   IBS3

FIG. 3. The upstream and downstream (exon) sequences adjacent to each of the 17 copies of the G.st.I1 intron are compared. With the exception of intron copy 9A22, all the intron copies appear to insert adjacent to a transcription terminator stem-loop structure (gray-shaded sequences below the large arrows) in the genome of strain 10. Intron DNA is in lowercase letters, and the blue-shaded sequences are the proposed IBS1 pairing region of the upstream exon and the single-base IBS3 pairing region downstream of the intron.

quences revealed that indeed most intron copies appear to have inserted just 3′ to a transcription terminator-like stem-loop followed by several Ts (gray-shaded DNA sequence under the arrows in Fig. 3). In addition, at the immediate 5′ junction with the intron, most copies contain at least 3 of the 4 bases of the IBS1 motif (blue-shaded TCCA in Fig. 3), and at the immediate 3′ junction with the intron, most copies contain the single-base IBS3 site (blue-shaded T in Fig. 3). These common properties of the different insertion sites for intron G.st.I1 are typical of other group IIC-type introns described by others (12).

Intron copy 9A22 interrupts an ORF. From the aligned DNAs in Fig. 3, it is clear that one of the intron copies, designated 9A22, has a very atypical insertion site. Namely, the 5′ exon DNA does not appear to contain a sequence capable of forming a strong stem-loop-type transcription terminator. Also, there is a poorly conserved IBS1 motif at the 5′ junction with this intron copy (Fig. 3). For these reasons, a larger region around the insertion site for copy 9A22 (and 12B43 and 8A6) was examined for intergenic DNA and open reading frames (ORFs). The complete, known genome sequence of *Geobacillus kaustophilus* was used as a guide to annotate the ORFs around the insertion site of these intron copies (18).

The intron copy 9A22 is unusual because it does not appear to have inserted adjacent to a transcription terminator (Fig. 3). Indeed, when DNA around this insertion site is analyzed, intron copy 9A22 is found to interrupt a small ORF encoding an rRNA methylase (Fig. 4C). This intron copy is inserted in the same orientation (i.e., into the coding strand) as the ORF, at codon 203 of a total of 247 codons. None of the other intron copies, 12B43 and 8A6 (Fig. 4) and b through p (not shown), appear to insert into an ORF.

This appears to be the first report of a group IIC intron that interrupts a chromosomally encoded ORF. The rRNA methylase encoded by this ORF is part of a large family of structurally similar methyltransferases related to the SpoU methyl-
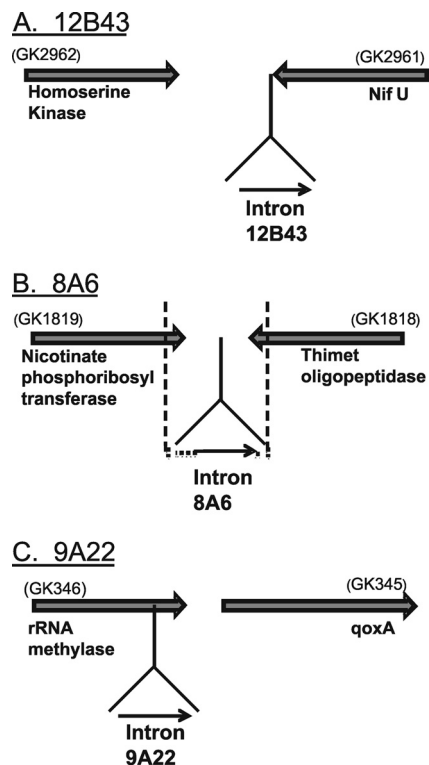

FIG. 4. Insertion sites for G.st.I1 intron copies 12B43, 8A6, and 9A22. (A) Intron copy 12B43 inserts (in the opposite orientation) adjacent to a transcription terminator stem-loop for the ORF coding for NifU. (B) Intron copy 8A6 inserts adjacent to a terminator stem-loop in an intergenic region between two converging ORFs. The dashed lines indicate a unique DNA sequence not present in the related genome of *G. kaustophilus*. (C) Intron copy 9A22 interrupts an ORF coding for a rRNA methylase by insertion at codon 203 of the 247-codon ORF.
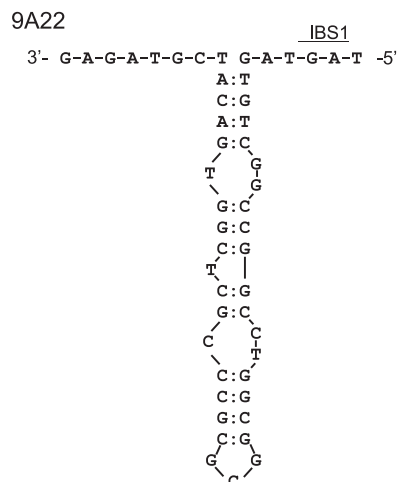
```
 9A22
                                    IBS1
   3'- G-A-G-A-T-G-C-T G-A-T-G-A-T -5'
                       A:T
                       C:G
                       A:T
                       G:C
                     T    G
                           G
                       G:C
                       G:C
                       C:G
                     T    |
                       C:G
                       G:C
                     C    C
                           T
                       C:G
                       C:G
                       G:C
                       C:G
                     G    G
                       C
```

FIG. 5. A postulated stem-loop structure that may form in the bottom DNA strand lies adjacent to the intron copy 9A22. A large but imperfect hairpin structure from the bottom strand of DNA is shown. This stem-loop lies immediately 5′ to the IBS1 motif at the ORF insertion site of the G.st.I1 intron (copy 9A22) and might form during lagging-strand synthesis of the replicating chromosome. Such a structure may have served as an integration signal for the intron.

ase of *Escherichia coli* (1). This is a distinct group of methyltransferases (termed SPOUT) broadly involved in RNA processing of tRNA and rRNA. All 8 completed *Geobacillus* genomes currently posted in GenBank contain this rRNA methylase gene (or a very similar gene). However, none appear to be interrupted with an intron (or other genetic element), as is the case for the strain 10 genome. It is not clear why in this one instance (copy 9A22) the G.st.I1 intron inserted into an ORF with presumably no adjacent terminator stem-loop. This ORF region does not appear to show any similarity to an integron *attC* site (no similarity to the 1L-1R and 2L-2R conserved regions) (6). However, an analysis of the bottom strand of the DNA at the insertion site revealed a long but imperfect stem-loop structure that may possibly form in single-stranded DNA, as might occur during lagging-strand synthesis of the replicating chromosome. This imperfect hairpin structure (Fig. 5), which lies immediately adjacent to the IBS1 motif, may have served as an aberrant integration signal for the G.st.I1 intron.

Several attempts were made to detect splicing of the G.st.I1 intron, but it was not observed. This included *in vivo* experiments in which total RNA from strain 10 cells was used to produce specific amplified cDNA of spliced intron mRNA (2). Also, plasmid clones of two intron copies (12B43 and 9A22) were used to produce runoff transcripts. These RNAs were then used to produce specific amplified cDNA copies of any spliced products (2). Finally, an *in vitro* experiment was done using radiolabeled intron RNA placed in a splicing reaction buffer to detect spliced products via electrophoresis.

Splicing and retrotransposition of group IIC introns have been observed and biochemically defined (2, 12, 20). We did not detect any genetic or structural differences in this intron (G.st.I1) to explain our failure to observe splicing. Thus, we expect that this intron is able to splice *in vivo* and therefore

would not block the expression of the rRNA methylase gene, which could be essential to the host cell.

Nucleotide sequence accession numbers. The cloned copies 12B43, 8A6, and 9A22 of the G.st.I1 intron with flanking DNA have been deposited in GenBank under the accession numbers HM566446, HM566447, and HM368526, respectively.

REFERENCES

1. **Anantharaman, V., E. V. Koonin, and L. Aravind.** 2002. SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. J. Mol. Microbiol. Biotechnol. **4:**71–75.
2. **Chee, G.-J., and H. Takami.** 2005. Housekeeping recA gene interrupted by group II intron in the thermophilic *Geobacillus kaustophilus*. Gene **363:**211–220.
3. **Lambowitz, A. M., and S. Zimmerly.** 2004. Mobile group II introns. Annu. Rev. Genet. **38:**1–35.
4. **Lampson, B. C.** 2007. Prokaryotic reverse transcriptases, p. 403–420. *In* J. Polaina and A. P. MacCabe (ed.), Industrial enzymes: structure, function, and applications. Springer, Dordrecht, Netherlands.
5. **Lampson, B. C., M. Inouye, and S. Inouye.** 2005. Retrons, msDNA, and the bacterial genome. Cytogenet. Genome Res. **110:**491–499.
6. **Leon, G., and P. H. Roy.** 2009. Group IIC intron mobility into *attC* sites involves a bulged DNA stem-loop motif. RNA **15:**1543–1553.
7. **Lewis, S., B. Perry, N. Fares, R. Morales-Diaz, and B. Roe.** 2010. *Bacillus (Geobacillus) stearothermophilus* genome sequencing—strain 10. *Bacillus (Geobacillus) stearothermophilus* Genome Sequencing Project. University of Oklahoma, Norman, OK. http://www.genome.ou.edu/bstearo.html.
8. **MacDonald, D., G. Demarre, M. Bouvier, D. Mazel, and D. N. Gopaul.** 2006. Structural basis for broad DNA-specificity in integron recombination. Nature **440:**1157–1162.
9. **Medhekar, B., and J. F. Miller.** 2007. Diversity-generating retroelements. Curr. Opin. Microbiol. **10:**388–395.
10. **Mitra, A., A. Kandavelmani, H. V. Jayashree, and V. Nagaraja.** 2009. Occurrence, divergence and evolution of intrinsic terminators across eubacteria. Genomics **94:**110–116.
11. **Qin, P. Z., and A. M. Pyle.** 1998. The architectural organization and mechanistic function of group II intron structural elements. Curr. Opin. Struct. Biol. **8:**301–308.
12. **Robart, A. R., S. Wooseok, and S. Zimmerly.** 2007. Insertion of group II intron retroelements after intrinsic transcription terminators. Proc. Natl. Acad. Sci. U. S. A. **104:**6620–6625.
13. **Simon, D. M., N. A. Clarke, B. A. McNeil, I. Johnson, D. Pantuso, L. Dai, D. Chai, and S. Zimmerly.** 2008. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA **14:**1704–1713.
14. **Simon, D. M., and S. Zimmerly.** 2008. A diversity of uncharacterized reverse transcriptases in bacteria. Nucleic Acids Res. **36:**7219–7229.
15. **Singh, R. N., R. J. Saldanha, L. M. D'Souza, and A. M. Lambowitz.** 2002. Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and auto-regulates translation. J. Mol. Biol. **318:**287–303.
16. **Smith, D., J. Zhong, M. Matsuura, A. M. Lambowitz, and M. Belfort.** 2005. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. Genes Dev. **19:**2477–2487.
17. **Stenesh, J., and R. A. Roe.** 1972. DNA polymerase from mesophilic and thermophilic bacteria. I. Purification and properties of the enzyme. Biochim. Biophys. Acta **272:**156–166.
18. **Takami, H., Y. Takaki, G. J. Chee, S. Nishi, S. Shimamura, H. Suzuki, S. Matsui, and I. Uchiyama.** 2004. Thermoadaption trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. Nucleic Acids Res. **32:**6292–6303.
19. **Toor, N., G. Hausner, and S. Zimmerly.** 2001. Co-evolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA **7:**1142–1152.
20. **Toor, N., A. R. Robart, J. Christianson, and S. Zimmerly.** 2006. Self-splicing of a group IIC intron: 5′ exon recognition and alternative 5′ splicing events implicate the stem-loop motif of a transcriptional terminator. Nucleic Acids Res. **34:**6461–6471.
21. **Vellore, J., S. E. Moretz, and B. C. Lampson.** 2004. A group II intron-type open reading frame from the thermophile *Bacillus (Geobacillus) stearothermophilus* encodes a heat-stable reverse transcriptase. Appl. Environ. Microbiol. **70:**7140–7147.