

Potential Role of Group IIC-*attC* Introns in Integron Cassette Formation^{▽†}

Grégory Léon^{1,2} and Paul H. Roy^{1,2*}

Centre de Recherche en Infectiologie, Centre Hospitalier Universitaire de Québec,¹ and Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Université Laval,² Québec, Canada

Received 22 May 2009/Accepted 20 July 2009

Integrans are natural expression vectors in which gene cassettes are integrated downstream of a promoter region by a site-specific recombinase. Gene cassettes usually consist of a single gene followed by a recombination site designated *attC*. A major unanswered question is how a gene becomes associated with an *attC* site. Here, we investigate the potential role of a specific lineage of group IIC introns, named group IIC-*attC*, in cassette formation. Group IIC-*attC* introns preferentially target *attC* while retaining the ability to target transcriptional terminators. We show using a PCR-based mobility assay with *Escherichia coli* that the *S.ma*.I2 intron from the genome of a clinical isolate of *Serratia marcescens* can target both *attC* site and putative terminator motifs of resistance genes. Quantitative results showed that *S.ma*.I2 is more efficient in targeting various *attC* sequences than three group IIC-*attC* introns (54 to 64% sequence identity) from the genomes of environmental isolates. We also show that purified group IIC-*attC* intron-encoded reverse transcriptases have both RNA-dependent and DNA-dependent DNA polymerase activities *in vitro*. These data permit us to suggest a new model for gene cassette formation, in which a group IIC-*attC* intron targets separately a transcriptional terminator adjoining a gene and an isolated *attC*, joins the gene and the *attC* by homologous recombination, and then splices and reverse transcribes a gene-*attC* RNA template, leading to the formation of a cassette.

Integrans and gene cassettes are considered important genetic elements in the evolution of multiresistance plasmids and transposons in gram-negative bacteria (39). Integrans can be categorized as mobile or chromosomal depending on their genomic location (i.e., on plasmids or on chromosomes). Mobile integrans and their cassettes are known to have a role in the dissemination of antibiotic resistance genes, whereas the majority of chromosomal integron cassettes include open reading frames (ORFs), most of whose products have no known function (21). Integrans are natural expression vectors composed of an integrase gene (*intI*) followed by a cassette promoter region and an integrase-specific recombination site (called an *attI* site) into which new cassettes are integrated (12). Gene cassettes usually consist of one promoterless gene associated with a distinct integrase-specific recombination site (called an *attC* site) that is located downstream of the gene. *attC* sites found in mobile integron cassettes exhibit little sequence similarity but contain a characteristic central palindromic sequence (11). *attC* sites can vary considerably in length (57 to 141 bp), and their sequence similarities are restricted primarily to the boundaries, which correspond to two pairs of conserved inverted repeats, 1L-2L and 2R-1R (40). Usually, a unique *attC* site is associated with one gene and is named by reference to the gene (e.g., the aminoglycoside resistance gene *aadA1* and the *attC_{aadA1}* site). In contrast, *attC* sequences from chromosomal integrans with large cassette arrays are usually closely related and species specific, despite marked differences

in the cassette ORF codon usage (3, 6, 10, 15, 32). These observations suggest that ORFs and *attC* sites of cassettes have independent origins and are associated by an unknown mechanism that plays a crucial role in integron evolution. By taking into account the unique structural characteristics of most integron cassettes (i.e., the absence of promoters, the paucity of noncoding sequence, and the presence of only one gene), it has been suggested that integron cassette genesis involves reverse transcription of mRNA molecules in an organism that encodes a reverse transcriptase (RT) (11). In such a model, the *attC* site may either be part of the original transcript (such as a Rho-independent transcription terminator) or added by an unknown mechanism from an isolated source of inverse repeat sequences (12, 30). The idea of reverse transcription became pertinent when we first reported the presence of a group IIC intron identified as *S.ma*.I2 inserted into a mobile integron between a structural gene and its associated *attC* site in the multiresistant *Serratia marcescens* SCH909 strain (4). Since then, a specific lineage of several group IIC introns, named the group IIC-*attC* introns, has been reported to be present in cassettes of mobile or chromosomal integrans (23, 27–29, 41, 46) or adjacent to isolated *attC* sites (18, 28).

Group II introns are catalytic RNAs and mobile retroelements that self-splice by means of a lariat intermediate (35). They are found in lower eukaryote organelle DNAs and also in several bacterial genomes. Group II introns are separated into several lineages based on their intron-encoded protein (IEP) sequences (36, 48). These mobile elements encode proteins having RT, RNA splicing (maturase), and sometimes DNA endonuclease (En) activities. The mobility of group II introns (known as retrohoming) involves invasion of a DNA target by an intron RNA and is mediated by a ribonucleoprotein complex composed of the excised intron lariat and the IEP (17). In

* Corresponding author. Mailing address: Centre de Recherche en Infectiologie, CHUQ, Pavillon CHUL, Local RC-709, 2705 boul. Lau-rier, Québec, Québec, Canada G1V 4G2. Phone: (418) 654-2705. Fax: (418) 654-2715. E-mail: Paul.Roy@crchul.ulaval.ca.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

[▽] Published ahead of print on 24 July 2009.

the first step, cleavage of the DNA top strand occurs by reverse splicing of the excised intron RNA at the intron target site. In the next step, bottom strand cleavage is catalyzed by the IEP's En domain. Then the 3' end of the cleaved bottom strand is used as a primer by the IEP's RT domain for reverse transcription of the intron RNA. In the final step, cellular repair processes complete the insertion of the intron (degradation of the intron RNA template, second-strand DNA synthesis, and ligation) into the new target site (38). Retrohoming is highly site specific because the intron RNA base pairs to a complementary DNA target known as intron binding site 1 (IBS1), IBS2, and δ sequence (or IBS3) to help position the reverse-splice site, while the IEP recognizes a small number of specific bases (24, 37).

Bacterial group IIC introns are a subgroup of group II introns that differ from the other bacterial introns in many RNA secondary structure aspects (42). Importantly for this study, in group IIC introns fewer base pairs are required for target site specificity (5 bp compared to 13 bp for groups IIA and IIB) because of missing IBS2-exon binding site 2 (EBS2) and shorter IBS1-EBS1 as well as IBS3-EBS3 pairing motifs (17). Nevertheless, group IIC introns are site-specific retroelements that are inserted directly after transcriptional terminator motifs or other inverted repeats, such as *attC* sites (28, 31). It was shown previously that recognition of a secondary structure, using the folded inverted repeats, was implicated in the target site specificity (19, 29, 31). Another important characteristic of group IIC introns is the absence of a C-terminal En domain in the IEPs. It has been shown that group II introns that cannot carry out site-specific second-strand cleavage due to the lack of an En domain are still mobile, but the residual mobility shows a pronounced strand bias and dependence on replication (20, 47).

In the present study, we searched for and found evidence for the potential role of group IIC-*attC* introns in gene cassette formation. We used an *Escherichia coli* two-plasmid mobility assay in order to compare the frequencies of mobility of four group IIC-*attC* introns (having 54 to 64% sequence identity) into *attC* sites and into putative transcriptional terminator motifs of resistance genes. We also purified group IIC-*attC* intron-encoded RTs in order to screen for both RNA- and DNA-dependent DNA polymerase activities in vitro. Together, our findings support a new cassette formation model in which a group IIC-*attC* intron uses both its specificity for DNA targets with stem-loop motifs and its IEP activities for targeting and then joining distant genes and *attC* sites.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and nucleic acid extraction. Bacterial strains and plasmids are described in Table 1. *S. marcescens* SCH909, *Shewanella baltica* OS155, and *E. coli* strains were grown in Luria-Bertani (LB) broth at 37°C or at the temperature indicated below. When necessary, antibiotics were used at the following concentrations: ampicillin (Ap), 100 μ g/ml; and chloramphenicol (Cm), 68 μ g/ml. *Nitrosomonas europaea* was cultured as described previously (18). *Geobacter sulfurreducens* genomic DNA and the *S. baltica* OS155 strain were kindly provided by The Institute for Genomic Research and by the DOE Joint Genome Institute, respectively. Total DNA was isolated using a phenol-chloroform purification method (34).

PCR procedures. For PCRs the Phusion DNA polymerase (Finnzymes) was used according to the manufacturer's instructions. PCR primers were synthesized

using an ABI-3900 DNA synthesizer from Applied Biosystems Inc. (Foster City, CA) (Table 2).

Nucleic acid folding. For each target site (*attC* or putative transcriptional terminator), secondary structures of the homing strand were determined using the MFOLD 3.2 program (49).

Intron mobility (retrohoming) assay. Intron mobility was determined by PCR amplification of the targeted sites using an *E. coli* two-plasmid assay in which an intron cloned into pUC18 inserts into a target site cloned into pACYC184 (Table 1). The recipient clones were obtained by cloning full-length complementary oligonucleotides (Table 2) or purified PCR amplicons of the target sites tested. For each mobility assay, the intron donor plasmid (Ap^r) and the recipient plasmid (Cm^r) were transformed simultaneously in *E. coli* DH5- α competent cells and subjected to Ap and Cm selection. One colony of each double transformant was grown in 5 ml of LB medium in the presence of Ap and Cm at 30°C overnight. The overnight preculture (100 μ l) was pelleted, inoculated into 5 ml of LB medium with both antibiotics, and incubated at 37°C until the optical density at 600 nm was 0.5. Expression of the intron was then induced by addition of 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG), followed by incubation at 37°C for 3 h. Plasmid DNAs were extracted using a QIAprep kit (Qiagen). Mobility was screened using PACYC-5' and one of the four intron-specific primers (NeI1.for, GsI2.for, SbI1.for, or SmaI2.for) (Table 2).

The homing frequency of *S.ma*I2 was determined by colony patch hybridization. Purified plasmids (100 ng) from induced cultures were digested with PvuI. This enzyme cuts only in the intron donor plasmid, outside the intron sequence. *E. coli* DH5 α cells were transformed with the PvuI-digested products and plated on LB medium containing Cm to select for cells containing recipient plasmids that were or were not interrupted by the intron. To calculate the intron mobility efficiency, expressed as the percentage of recipient plasmids that received the intron, 100 isolated Cm^r colonies were patched onto plates containing LB medium with Cm. The patches were lifted onto nylon membranes and hybridized with a 5'-end-labeled probe (SmaI2.for) using [γ -³²P]ATP (3,000 Ci/mmol; GE Healthcare, Piscataway, NJ). For each target site tested, randomly chosen positive colonies were cultured, and the homing site was confirmed by PCR amplification of the target site (using the PACYC-5' and PACYC.rev primers) and sequencing.

MBP-IEP expression and purification. For expression and purification of soluble IEPs, we used the pMAL protein fusion and purification system of New England Biolabs (NEB). pMAL-c2X clones were expressed in *E. coli* Rosetta2(DE3)(pLysS) (Novagen) (Table 1). A single colony was inoculated into 5 ml of LB medium with 100 μ g/ml Ap and 34 μ g/ml Cm and grown at 30°C overnight. This preculture (500 μ l) was pelleted, inoculated into 50 ml of LB medium containing 1% glucose and the same antibiotics, and grown at 37°C to an optical density at 600 nm of 0.5, and this was followed by induction with 1 mM IPTG for 2 h at 37°C. Cells were harvested by centrifugation for 20 min at 4,000 \times g, resuspended in 10 ml of column buffer (20 mM Tris [pH 7.4], 200 mM NaCl, 1 mM EDTA), and kept at -20°C. Cells were lysed on ice by sonication using four 20-s 25-W bursts interspersed with 20-s rest periods. All subsequent steps were carried out at 4°C. Lysates were cleared by centrifugation for 30 min at 9,000 \times g, and the maltose binding protein (MBP)-IEP fusion proteins were purified on 500 μ l of amylose resin used according to the manufacturer's instructions (NEB). Fusion proteins were eluted using 2 ml of column buffer with 20 mM maltose, pooled, and concentrated using a Centricon YM-10 (Milipore). Finally, glycerol was added to a final concentration of 20%.

Biochemical assays. RNA-dependent (RT) and DNA-dependent DNA polymerase assays were carried out with 500 ng of MBP-IEPs, as determined by the Bradford assay (Bio-Rad) using bovine serum albumin (NEB) as the standard. RT assays were performed using 500 ng of the artificial template and primer substrates poly(rA) and oligo(dT)₁₂₋₁₈ (Amersham Biosciences) and 5 μ Ci [α -³²P]dTTP (3,000 Ci/mmol; GE Healthcare, Piscataway, NJ) in a 10- μ l reaction mixture consisting of 1 \times RT buffer (50 mM Tris-HCl [pH 8.3], 75 mM KCl, 3 mM MgCl₂, 5 mM dithiothreitol). Reverse transcription reactions were initiated by addition of purified MBP-IEPs to the reaction mixtures, which were incubated for 30 min at 42°C. To measure the incorporation of [α -³²P]dTTP, 5- μ l portions of the reaction mixtures were spotted onto DE81 paper (Whatman, Fairfield, NJ) and washed four times with 2 \times SSC (1 \times SSC is 150 mM NaCl plus 15 mM sodium citrate) for 5 min each time. The filters were then dried and quantified by Cerenkov counting using a Beckman LS1801 scintillation counter. RT activities were compared to the activity of the SuperScript III polymerase (100 U; Invitrogen).

DNA-dependent DNA polymerase activity was tested in the presence of activated (nicked with DNase I) calf thymus DNA (100 μ g/ml; Amersham Biosciences), dATP (10 μ M), dTTP (10 μ M), dGTP (10 μ M), and 1 μ M [γ -³²P]dCTP. Reactions were performed with 25- μ l (total volume) mixtures at 37°C using the

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Description or relevant characteristics	Source or reference
<i>E. coli</i> strains		
DH5- α	<i>supE44</i> Δ <i>lacU169</i> (ϕ 80 <i>lacZ</i> Δ <i>M15</i>) <i>hsdR17</i> <i>recA1</i> <i>endA1</i> <i>gyrA96</i> <i>thi-1</i> <i>relA1</i>	Invitrogen
Rosetta2 (DE3)(pLysS)	F ⁻ <i>ompT</i> <i>hsdSB</i> (r _B ⁻ m _B ⁻) <i>gal</i> <i>dcm</i> (DE3) pLysSRARE2 (Cm ^r)	Novagen
Other strains		
<i>S. marcescens</i> SCH909	Strain carrying three integrons and one group IIC intron	4
<i>N. europaea</i>	Strain carrying one integron and two identical group IIC intron copies	18
<i>S. baltica</i> OS155	Strain carrying two identical group IIC intron copies without an adjacent <i>attC</i> site or integrase gene	
Plasmids		
pNeI1	2,376-bp <i>SacI</i> / <i>HindIII</i> PCR fragment amplified from the genome of <i>N. europaea</i> (containing 2,007 bp from the full-size intron copy in the integron, 101 bp from exon 1, and 268 bp from exon 2) cloned in pUC18 digested by <i>SacI</i> / <i>HindIII</i>	
pNeI1-EBS1	pNeI1 clone in which the intron EBS1 was changed by site-directed mutagenesis (5'-ACAU \rightarrow 5'-ACAA)	
pGsI2	2,408-bp <i>SacI</i> / <i>HindIII</i> PCR fragment amplified from the genome of <i>G. sulfurreducens</i> (containing 1,908 bp from the full-size intron copy distant from the integron, 250 bp from exon 1, and 250 bp from exon 2) cloned in pUC18 digested by <i>SacI</i> / <i>HindIII</i>	
pSbI1	2,449-bp <i>SacI</i> / <i>HindIII</i> PCR fragment amplified from the genome of <i>S. baltica</i> OS155 (containing 1,950 bp from the full-size intron, 247 bp from exon 1, and 252 bp from exon 2) cloned in pUC18 digested by <i>SacI</i> / <i>HindIII</i>	
pSmaI2	2,469-bp <i>SacI</i> / <i>HindIII</i> PCR fragment amplified from the genome of <i>S. marcescens</i> SCH909 (containing 1,971 bp from the full-size intron, 245 bp from exon 1, and 253 bp from exon 2) cloned in pUC18 digested by <i>SacI</i> / <i>HindIII</i>	
pSmaI2-YAHH	Clone derived from pSmaI2 by site-directed mutagenesis to replace the intron ORF nucleotides GACGAT with CACCAT (positions 1311 to 1316) and the corresponding amino acid residues D256 and D257 with H256 and H257	
pMAL-c2X	Cloning vector with the MBP gene (<i>malE</i>) for purification of a fusion protein	NEB
pNetr	1,314-bp blunt <i>HindIII</i> PCR fragment of the entire intron ORF amplified from pNeI1 and cloned into pMAL-c2X digested by <i>XmnI</i> and <i>HindIII</i>	
pNetr-YAHH	Clone derived from pNetr by site-directed mutagenesis to replace the intron ORF nucleotides GACGAT with CACCAT (positions 1314 to 1319) and the corresponding amino acid residues D238 and D239 with H238 and H239	
pGstr	1,305-bp blunt <i>HindIII</i> PCR fragment of the entire intron ORF amplified from pGsI2 and cloned into pMAL-c2X digested by <i>XmnI</i> and <i>HindIII</i>	
pattCaadA1	Annealed oligonucleotides attCaadA1.for and attCaadA1.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i> ; the attCaadA1 sequence is from the <i>aadA1</i> cassette in the integron of Tn21	
pattCdfrA1	147-bp <i>HindIII</i> / <i>BamHI</i> PCR fragment of attCdfrA1 amplified from pLQ427 and cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i> ; the attCdfrA1 sequence is from the <i>dfrA1</i> cassette in integron 1 of <i>S. marcescens</i> SCH909	18
pattCsat	Annealed oligonucleotides attCsat.for and attCsat.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i> ; the attCsat sequence is from the <i>sat</i> cassette in the integron of Tn7	
pattCpse1	Annealed oligonucleotides attCpse1.for and attCpse1.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i> ; the attCpse1 sequence is from the <i>pse1</i> cassette in the integron of <i>P. aeruginosa</i> SCH481 (clinical isolate)	
pattCbla _{IMP}	188-bp <i>HindIII</i> / <i>BamHI</i> PCR fragment of attCbla _{IMP} amplified from pLQ439 and cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i> ; the attCbla _{IMP} sequence is from the <i>bla</i> _{IMP} cassette in the integron of <i>S. marcescens</i> AK9373	18
pattCaacA1a-orfG	162-bp <i>HindIII</i> / <i>BamHI</i> PCR fragment of attCaacA1a-orfG amplified from pLQ440 and cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i> ; the attCaacA1a-orfG sequence is from the <i>aacA1a-orfG</i> cassette in the integron of Tn2424	18
pattCVCr	199-bp <i>HindIII</i> / <i>BamHI</i> PCR fragment of a VCR amplified from pSU38::ORF1-cat and cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i>	22
pblaSHV-5	Annealed oligonucleotides blaSHV-5.for and blaSHV-5.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i>	
pblaSHV-5 mut.	Annealed oligonucleotides blaSHV-5_mut.for and blaSHV-5_mut.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i>	
pblaCMY-8	Annealed oligonucleotides blaCMY-8.for and blaCMY-8.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i>	
pblaCMY-8 mut	Annealed oligonucleotides blaCMY-8_mut.for and blaCMY-8_mut.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i>	
pblaCTX-M-3	Annealed oligonucleotides blaCTX-M-3.for and blaCTX-M-3.rev cloned in pACYC184 digested by <i>HindIII</i> / <i>BamHI</i>	

RT reaction buffer. DNA synthesis was stopped after 10 min by addition of 600 μ l of cold 5% trichloroacetic acid-1% sodium pyrophosphate, which was followed by nucleic acid precipitation on ice for 10 min. The samples were filtered (Millipore) and washed with 250 μ l of 5% trichloroacetic acid-1% sodium

pyrophosphate to measure the labeled DNA by scintillation counting (filter-based assay). The background was determined by a parallel incubation with purified MBP. DNA-dependent DNA polymerase activities were compared to the activity of Biotools DNA polymerase (5 U; Biotools).

TABLE 2. Oligonucleotides used in this study

Oligonucleotide	Sequence
NeI1_SacI.....	5'-ACGCCGAGCTCAGGGGCAGACAC
NeI1_HindIII.....	5'-TCAGATCAAAAGCTTTTCCTTGTGG
GsI2_SacI.....	5'-TTTTGAGCTCATAAACATCGCTGATGTGAATC
GsI2_HindIII.....	5'-TTTTAAGCTTAATCCGGAATTGGAGACCCC
SbI1_SacI.....	5'-TTTTGAGCTCTTTACTCAAGTAATCATCG
SbI1_HindIII.....	5'-TTTTAAGCTTCTAATGAGCAGACCAAAGATTG
SmaI2_SacI.....	5'-TTTTGAGCTCGGACAGTCTCATCGAGCTG
SmaI2_HindIII.....	5'-TTTTAAGCTTACGAAGCGTATGAAATCGCG
NeI1_EBS1.for.....	5'-GAGCTGGCACAAAGACAGCGAAGCCCAACGG
NeI1_EBS1.rev.....	5'-CCGTTGGGCTTCGCTGTCTTGTGCCAGCTC
SmaI2_YAAH.for.....	5'-GCACGTTACGCTCATCACATAATCATTTTAG
SmaI2_YAHH.rev.....	5'-CTAAATGATTATGTGATGAGCGTAACGTGC
Netr.for.....	5'-ATGCACCGAGCCTTGAACCAAGAC
Netr.rev.....	5'-TTTTAAGCTTATCAACCCTTGTCCGGCCC
Netr_YAHH.for.....	5'-CGCCCGCTACGCCACCATGTGATCATCC
Netr_YAHH.rev.....	5'-GGATGATCACATGGTGGGCGTAGCGGGCG
Gstr.for.....	5'-ATGCCCGTGGAGCGCTCATTGACCG
Gstr.rev.....	5'-TTTTAAGCTTATTACCGGGCCGTGGCCGG
attCaadA1.for.....	5'-AGCTTGTCTAACAATTCGTTCAAGCCGACGCCGCTTCGCGGCGCGGCTTAA CTCAAGCGTTAGATG
attCaadA1.rev.....	5'-GATCCATCTAACGCTTGAGTTAAGCCGCGCCGCGAAGCGGCGTTCGCGCTTG AACGAATTGTTAGACA
attCdfA1.for.....	5'-AAAAAAGCTTTACCAAATCTGGCAAAAGGG
attCdfA1.rev.....	5'-AAAAGGATCCTCGATACTTCGGCGATCACC
attCaacA1a-orfG.for.....	5'-TTTTAAGCTTTGATCAACCGAAATTTGCCC
attCaacA1a-orfG.rev.....	5'-TTTTGGATCCGCGTTTGGATGAAATTTGCCC
attCsat.for.....	5'-AGCTTGCTAACAATTCATTCAAGCCGACACCGCTTCGCGGCGCGGCTTAA TTCAGGAGTTAAACG
attCsat.rev.....	5'-GATCCGTTTAACTCCTGAATTAAGCCGCGCCGCGAAGCGGTGTTCGCGCTTG AATGAATTGTTAGGCA
attCpse1.for.....	5'-AGCTTGGCTAACAAGGCCATCAAGTTGACGGCTTTTCCGTCGCTTGTTTTG TGGTTTAAACGCTACGCTACCACAAAACAATCAACTCCAAAGCCGCAACTTA TGGCGGCGTTAGACG
attCpse1.rev.....	5'-GATCCGTCTAACGCGCCATAAGTTGCGGCTTTGGAGTTGATTGTTTGTG GTAGCGTAGCGTTAAACCACAAAACAAGCGACGGAAGCCGTCAACTTG ATGGCCTTGTTAGCCA
attCblaimp.for.....	5'-AAAAAAGCTTAAAAACCATCAAAACCAAGC
attCblaimp.rev.....	5'-AAAAGGATCCGGTGCTGTTGGTACAGATGC
VCR.for.....	5'-TTTTAAGCTTTTAATGGTAAGTTATGGG
VCR.rev.....	5'-TTTTGGATCCATAGCTCTCATGATATCTCC
blaSHV-5.for.....	5'-AGCTTAACCCGGCGGTGGCCGCGCGCTTATCCGGCCCCGAGCACCTCGC AGGCGTGCCGGGCG
blaSHV-5.rev.....	5'-GATCCGCCCCGGCACGCCTGCGAGGTGCTGCGGGCCGGATAACGCGCGCG GCCACCGCCGGGTTA
blaSHV-5_mut.for.....	5'-AGCTTAACAAGGCGGTGGCCGCGCGCTTATCCGGCCCCGAGCACCTCGC AGGCGTGCCGGGCG
blaSHV-5_mut.rev.....	5'-GATCCGCCCCGGCACGCCTGCGAGGTGCTGCGGGCCGGATAACGCGCGCG GCCACCGCCTTGTTA
blaCMY-8.for.....	5'-AGCTTGACAGAAAGAGGGCGGTACATTTCGGTGAATGTGCCGCCCTTTTTCT TGGTGCTGGGGGAAG
blaCMY-8.rev.....	5'-GATCCTTCCCCCAGCACCAGAAAAAGGGCGGCACATTACCGAATGTACC GCCCTCTTTCTGTCA
blaCMY-8_mut.for.....	5'-AGCTTGACAACAAGAGGGCGGTACATTTCGGTGAATGTGCCGCCCTTTTTCT GGTGCTGGGGGAAG
blaCMY-8_mut.rev.....	5'-GATCCTTCCCCCAGCACCAGAAAAAGGGCGGCACATTACCGAATGTACC GCCCTCTTGTGTCA
blaCTX-M-3.for.....	5'-AGCTTGTAATAGCGAAACGGAATGGGGAAACTCATTCCGTTTTTGTATTAT CGCCTTAGACGGCG
blaCTX-M-3.rev.....	5'-GATCCGCGCTCTAAGGCGATAAAACAAAAACGGAATGAGTTTCCCATTC GTTTCCGCTATTACA
PACYC-5'.....	5'-TGAGCACCTGAAGTCAGCC
PACYC.rev.....	5'-ATGAGCCCCGAAGTGGCGAGC
NeI1.for.....	5'-CATAGAAGGCCGAGGCAGGC
GsI2.for.....	5'-AGCCGACGAACAGAAACAGC
SbI1.for.....	5'-GCGAGCTGATGAACAAGAAC
SmaI2.for.....	5'-GCAAATCTGTGAGCTGATGA

RESULTS

Analysis of the genomic context of group IIC-*attC* introns. In recent years, sequencing of the genomes of clinical and environmental bacteria has led to the identification of 71 bacterial genomes with full-length or fragmented group IIC-*attC* introns (28). Fifteen of the full-length introns are in mobile integrons in antibiotic resistance cassettes conferring resistance to β -lactams, aminoglycosides, or trimethoprim (Fig. 1A) (4, 16, 23, 27, 29, 41). Five of them are in chromosomal integrons in gene cassettes that have no known homologues (except for the *N. europaea* integron, which contains an ORF possibly related to *qacE*, which mediates resistance to quaternary ammonium compounds) (Fig. 1B) (18, 28). Three do not have any nearby integrase gene homolog but are inserted into an *attC* site (Fig. 1C) (18, 28).

In both mobile and chromosomal integrons, a single intron copy is usually present in the last cassette and in the direction opposite that of cassette transcription. Figure 1A shows that various clinical strains isolated in Korea (e.g., *Acinetobacter* genomospecies and *Enterobacter cloacae* strains) contain virtually identical intron copies in mobile integrons that contain related resistance gene cassettes (16, 46). Figures 1B and 1C show that in the *N. europaea* and the *G. sulfurreducens* genomes, second identical intron copies have *attC* sites at their 5' ends but no recognizable ORF at the other end. Another intron *attC*, without any nearby *intI* homolog, is present in the metagenome of an uncultured *Shewanella* sp. strain (accession no. AACY020561240) (45). Sequence analysis revealed two intron copies inserted into distinct *attC* sites and separated by a 480-bp noncoding sequence (Fig. 1C). Homologs of the 480-bp sequence between the two introns have been found to be cassettes with *Vibrio cholerae* repetitive DNA sequences (VCRs) in multiple copies in the chromosomal integron of *Vibrio cholerae* N16961 (accession no. AE003853) that contains 176 cassettes. Non-protein-encoding cassettes occurring in several copies are intriguing features of *Vibrio* integrons, notably *Vibrio vulnificus* integrons (5).

Analysis of the homing sites showed that most introns are inserted into *attC* sites of various sequences in the consensus TTGT/T (IBS1-IBS3) motif that is located downstream of imperfect inverted repeat sequences (28) (see Fig. S1 in the supplemental material). Interestingly, compilation of the target sites for introns in mobile integrons showed that there was a preponderance of the 60-bp *attC_{aadA1}* site at the 5' ends of several introns. However, in some cases a resistance gene other than the expected *aadA1* gene was found at the 3' end of the same introns. Instead, we found the *qacF* resistance gene in *S. marcescens* (accession no. AY030343), the putative fosfomycin resistance gene *fosX* in *Pseudomonas putida* (accession no. AY065966), and the *bla_{VIM-2}* resistance gene in *Klebsiella pneumoniae* (accession no. DQ153218) and *Providencia rettgeri* (accession no. AY887109) (Fig. 1A). Lack of correspondence between the 5' and 3' exon sequences was also observed in *Salmonella enterica* serovar Typhimurium (accession no. AY204504) for the *attC_{cmIA}* site and the *aadB* resistance gene. Together, these observations suggest that genes and *attC* sites with independent origins are brought together by group IIC-*attC* introns and that intron-containing cassettes could represent intermediates in cassette assembly.

In vivo mobility of cloned group IIC-*attC* introns into various *attC* sites. We used a two-plasmid mobility assay with *E. coli* to compare the levels of mobility of group IIC-*attC* introns *S.ma.I2* (accession no. AF453998), *N.e.I1* (accession no. AL954747), *G.s.I2* (accession no. AE017180), and *S.b.I1* (accession no. CP000563) into various *attC* sites (Fig. 2A; see Materials and Methods and Fig. S2 in the supplemental material). We used these introns because their IEPs share only 48 to 58% amino acid identity and because they are present either in gene cassette arrays of integrons (*S.ma.I2*, *N.e.I1*, and *G.s.I2*) or in an isolated *attC* site without a nearby integrase gene (*S.b.I1*) (28). Six *attC* sites that were successfully targeted by *S.ma.I2* (29) were subcloned into pACYC184 in order to remove unnecessary sequences (i.e., resistance gene and other recombination sites) from the previous recipient plasmids. Additionally, we cloned a VCR site (22), which was not previously tested as a homing site (Table 1).

Intron mobility was screened by PCR and/or colony patch hybridization methods (see Materials and Methods). Figure 2B shows the results of a sensitive PCR assay performed with PACYC-5' and one of the four intron-specific primers in order to amplify the 3' intron-exon integration junction. PCR bands corresponding to specific homing events were detected for all introns. At least two of the seven *attC* sites (*attC_{aadA1}* and *attC_{sat}*) were successfully targeted by all of the introns tested. These two sites are the smallest sites (60 bp) and share 88.1% sequence identity. Sequencing of the PCR products showed that the four introns inserted specifically between the IBS1-IBS3 motifs (5'-TTGT/T) located in the bottom strand sequence of all *attC* sites tested (data not shown). These results confirmed the results of the previous in vivo mobility assays performed with *S.ma.I2* (29). Interestingly, *S.ma.I2* is the only intron that targeted all *attC* sites, including the newly tested VCR site. Analysis of the folded structures of the various target sites showed that the VCR bottom strand folds into a stem-loop structure with a large central loop composed of unpaired bases (see Fig. S2 in the supplemental material). Compared to the three-base loops of most *attC* sites tested in this study, the large central loop of VCR might affect recognition of either stem or IBS1-IBS3 motifs by the incoming introns. We also included in this assay the pNeI1-EBS1 mutant clone, in which the 5'-ACAU→ACAA EBS1 mutation of *N.e.I1* could result in stronger EBS1-IBS1 pairing, and the pSmaI2-YAHH mutant clone, which potentially suppresses the RT activity of the *S.ma.I2* IEP. Figure 2B shows that the single-base EBS1 change allowed the *N.e.I1* mutant to target two additional *attC* sites (*attC_{dfpA1}* and *attC_{pse1}*) compared to the wild-type intron. As expected, mobility into *attC* sites was not detected for the *S.ma.I2* YAHH mutant.

In order to compare the levels of mobility of *S.ma.I2* for the distinct *attC* sites, we used the colony patch hybridization method. Under the experimental conditions used, the *S.ma.I2* mobility frequency (defined as the percentage of homing site plasmids containing introns) varied from <1.0% for the *attC_{dfpA1}* and VCR sites to 95.5% for the *attC_{pse1}* site (Table 3). Results shown in Table 3 were confirmed using the PACYC-5' and PACYC-rev primers in a semiquantitative PCR assay in which the intensity of the PCR products correlates with the mobility frequency values (see Fig. S3 in the supplemental material). Together, the results shown in Fig. 2B and Fig. S3 in the

supplemental material suggest that the *N.e.I1*, *G.s.I2*, and *S.b.I1* introns have low mobility frequencies into *attCs* compared to *S.ma.I2*. Colony patch hybridization assays confirmed that the mobility frequencies of the *N.e.I1*, *G.s.I2*, and *S.b.I1* introns into *attC_{good41}* were <1.0% (data not shown).

Assays of in vivo mobility of *S.ma*.I2 into putative transcriptional terminators of resistance genes. Assays of the mobility of group IIC-*attC* introns into mutated *attC* sites showed that a structural motif, rather than a defined homing site sequence, is involved in target site recognition along with the short IBS1-

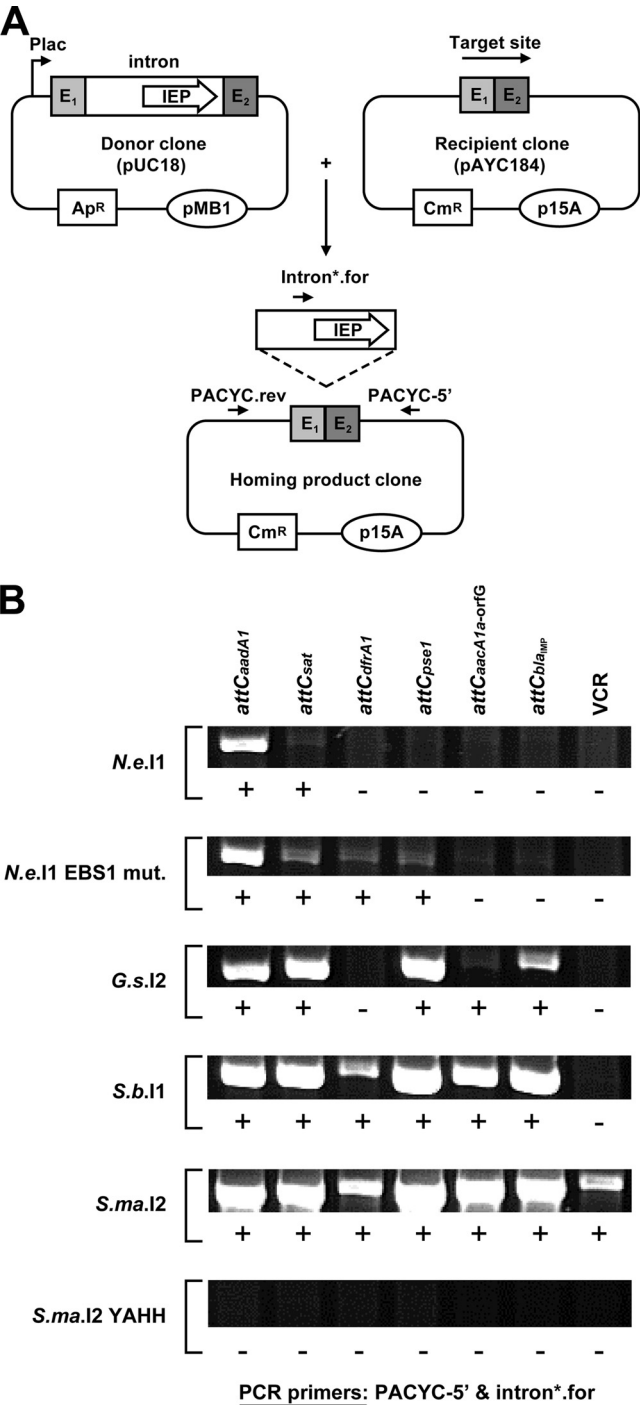


FIG. 2. (A) Two-plasmid mobility assay with *E. coli*. E₁ and E₂, exon sequences; Plac, *lac* promoter; Ap^R and Cm^R, ampicillin and chloramphenicol resistance genes, respectively; pMB1 and p15A, plasmid replication origins. “Intron*.for” indicates one of four intron-specific primers, NeI1.for, GsI2.for, SbI1.for, or SmaI2.for (Table 2). (B) In vivo mobility of group IIC-*attC* introns into various *attC* site sequences. Agarose gels contained the PCR products obtained using PACYC-5' and one of the four intron-specific primers. These primers amplify the 3' intron-exon integration junction and are specific for homing into the bottom strand of the cloned *attC* sites. +, positive PCR result for integration; –, negative PCR result for integration. Secondary structures of the bottom strand sequence (homing strand) of all *attC* sites tested are shown in Fig. S2 in the supplemental material.

TABLE 3. In vivo mobility frequency of *S.ma*.I2 for various target sites

Target ^a	Homing strand ^b	Orientation ^c	<i>S.ma</i> .I2 mobility frequency (%) ^d
<i>attC_{aadA1}</i>	bs	Lag	18.0 ± 2.5
<i>attC_{sat}</i>	bs	Lag	15.0 ± 1.0
<i>attC_{dfcA1}</i>	bs	Lag	<1.0
<i>attC_{pse1}</i>	bs	Lag	95.5 ± 2.9
<i>attC_{aacA1a-orfG}</i>	bs	Lag	4.5 ± 1.9
<i>attC_{blaIMP}</i>	bs	Lag	4.3 ± 1.7
VCR	bs	Lag	<1.0
Term_ <i>bla_{CMY-8}</i>	bs	Lag	<1.0
Term_ <i>bla_{CMY-8}</i> mutant	bs	Lag	<1.0
Term_ <i>bla_{SHV-5}</i>	bs	Lag	<1.0
Term_ <i>bla_{SHV-5}</i> mutant	bs	Lag	<1.0
Term_ <i>bla_{CTX-M-3}</i>	ts	Lead	<1.0

^a The target was a gene cassette *attC* site (*attC* or VCR) or putative transcriptional terminator (Term).

^b *S.ma*.I2 homing specificity for the bottom strand (bs) or top strand (ts) sequences based on the PCR products shown in Fig. 2B and 3.

^c Orientation of the homing strand sequence relative to the direction of DNA replication, lagging or leading strand.

^d Percentage of recipient plasmids that received the intron. The mobility frequencies were determined by colony patch hybridization with an intron-specific probe and are the means ± standard deviations of positive hybridization signals of at least three independent experiments.

IBS3 motifs (19, 29). In this study, we suggest that the group IIC-*attC* introns are involved in the recruitment of distant genes and *attC* sites. We tested whether *S.ma*.I2, which showed the highest mobility frequencies for various *attC* sites, could also target transcriptional terminator sequences, which are the preferred targets of other group IIC introns (31). We searched GenBank databases for potential IBS1-IBS3 and stem-loop motifs in putative transcriptional terminators of antibiotic resistance genes. Candidate sequences (designated Term sites) downstream of the plasmid-borne β-lactamase genes *bla_{CMY-8}* (NCBI Entrez gene ID 6383913; in *K. pneumoniae* plasmid pK29; accession no. EF382672), *bla_{SHV-5}* (gene ID 1446571; in *E. coli* plasmid p1658/97; accession no. AF550679), and *bla_{CTX-M-3}* (in *Citrobacter freundii* plasmid pCTX-M3; accession no. AF550415) were cloned and tested as homing sites for *S.ma*.I2 in *E. coli* (Table 1). Figure 3A shows agarose gels containing the PCR products corresponding to the 3' intron-exon integration junction in the Term sites indicated. PCR bands corresponding to *S.ma*.I2 mobility were detected with the wild-type Term_ *bla_{CMY-8}* site in the bottom strand sequence and with the Term_ *bla_{CTX-M-3}* site in the top strand sequence. Sequencing of the PCR amplicons showed intron mobility into the 5'-CTGT/C motif (IBS1-IBS3), which overlaps the stop codon of *bla_{CMY-8}*, and into the 5'-TTGT/T motif, which is located on the top strand downstream of a potential stem-loop motif in the Term_ *bla_{CTX-M-3}* sequence (Fig. 3B; see Fig. S4 in the supplemental material). Other group IIC introns have been found to be joined to transcriptional terminators in nature (31); however, the orientation and strand specificity of *S.ma*.I2 insertion appear to be determined by the position of the IBS1 and IBS3 motifs (19).

The importance of complementary EBS1-IBS1 motifs was confirmed by mutation of putative IBS1 motifs for the consensus 5'-TTGT sequence in the Term_ *bla_{CMY-8}* and Term_ *bla_{SHV-5}* bottom strand sequences. A two-base mutation of the

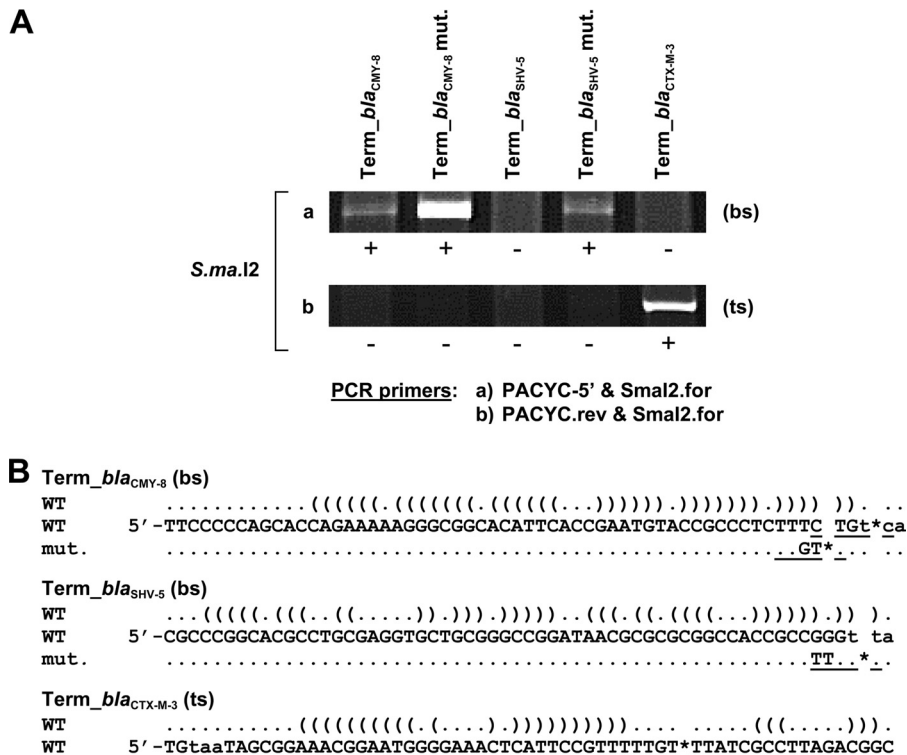


FIG. 3. *S.ma.I2* intron mobility into putative transcriptional terminators (Term sites) of various resistance genes. (A) Agarose gels containing the PCR products obtained using primer pair “a” or “b” in order to detect insertion of *S.ma.I2* into putative IBS1-IBS3 motifs in bottom strand (bs) or top strand (ts) sequences, respectively. (B) *S.ma.I2* homing site for each Term site based on the sequencing results. The asterisks indicate the intron insertion sites between the four-base IBS1 motif and the one-base IBS3 motif (underlined). Lowercase letters indicate the gene stop codons. *Vienna* outputs (above each sequence) indicate base pairing from a potential stem-loop structure. A period indicates sequence identity with the wild type. Secondary structures of the homing strand are shown in Fig. S4 in the supplemental material. WT, wild type; mut., mutant.

5'-TTTC sequence located upstream of the Term_ *bla*_{CMY-8} insertion site (see above) to 5'-TTGT resulted in a stronger PCR band corresponding to a specific insertion of *S.ma.I2* into the new 5'-TTGT/T (IBS1-IBS3) site. Similarly, mutation of the 5'-GGGT sequence that overlapped the stop codon of *bla*_{SHV-5} on the bottom strand to 5'-TTGT resulted in a weak PCR band corresponding to a specific insertion of *S.ma.I2* into the new 5'-TTGT/T (IBS1-IBS3) site. In contrast to the positive results for *S.ma.I2*, in similar mobility assays with the *N.e.I1*, *G.s.I2*, and *S.b.I1* introns negative PCR results were obtained for all of the Term sites (data not shown).

We compared the levels of mobility of *S.ma.I2* for the distinct Term sites using the colony patch hybridization method (Table 3). Under the experimental conditions used, the *S.ma.I2* mobility frequencies were <1.0% for both wild-type and mutated Term sites.

***N.e.I1*- and *G.s.I2*-encoded RTs have DNA-dependent DNA polymerase activity.** Based on the structural characteristics of most integron cassettes (i.e., the absence of promoters, the paucity of noncoding sequences, and the presence of only one gene), it has been suggested that cassette formation involves reverse transcription of assembled gene-*attC* sequences at the RNA level (4, 11, 12, 30). Several studies showed the RT activity of purified group II IEP or ribonucleoprotein preparations (25, 26, 31, 38, 44). Here, we searched for RNA-dependent (RT) and DNA-dependent DNA polymerase activi-

ties of purified IEPs from two distinct introns, *N.e.I1* and *G.s.I2*. The IEPs of these introns (Netr and Gstr, respectively; 59.4% amino acid identity) were expressed in *E. coli* and affinity purified using the pMAL protein fusion and purification system (see Materials and Methods). The purified ~90-kDa MBP-IEP fusions (~40 kDa for the MBP and ~50 kDa for the IEP) were tested for RT and DNA-dependent DNA polymerase activities using poly(rA)-oligo(dT)₁₂₋₁₈ and activated calf thymus DNA substrates, respectively (see Materials and Methods). For the RT assays, we used a purified MBP preparation as a negative (or background) control and a commercial RT, SuperScript III, as a positive control. For the DNA-dependent DNA polymerase assays, we used a purified MBP preparation as a negative control and a commercial DNA polymerase, the Biotools DNA polymerase, as a positive control. Figure 4 shows that the MBP-Netr and MBP-Gstr fusions displayed high RT activities (comparable to the SuperScript III activity) with the poly(rA)-oligo(dT)₁₂₋₁₈ substrate, whereas no significant activity was detected with MBP alone. As expected, the RT activity of the MBP-Netr protein was abolished by point mutations of the conserved aspartate residues in the YADD motif (changed to YAAH) of the IEP RT domain. Figure 4 shows that the MBP-Netr and MBP-Gstr fusions also displayed significant DNA-dependent DNA polymerase activities with the calf thymus DNA substrate. The MBP and MBP-Netr YAAH proteins displayed no activity, and the activity of the

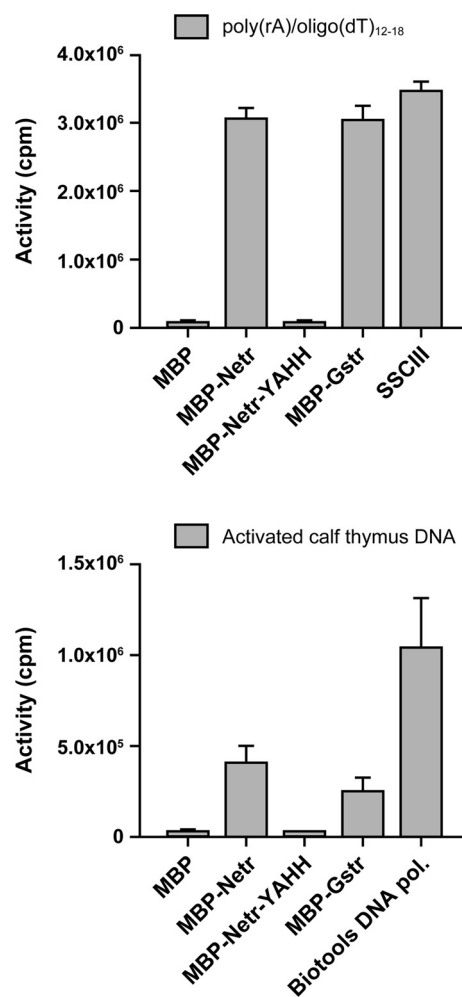


FIG. 4. RT activity of purified group IIC-*attC* IEPs. Quantitative RNA-dependent (RT) and DNA-dependent activities of purified MBP-IEPs were assayed using poly(rA)-oligo(dT)₁₂₋₁₈ and activated calf thymus DNA substrates, respectively. Assays were performed in triplicate, and the error bars indicate standard deviations. Purified MBP was used as a negative control. The commercial Superscript III (SSCIII) and Biotools DNA polymerase enzymes were used as positive controls for the RT and DNA-dependent DNA polymerase (pol.) activities, respectively.

Biotools DNA polymerase used as a positive control was at least twice that of the wild-type MBP-IEP fusions.

DISCUSSION

Integrans and gene cassettes are ancient structures that play a major role in bacterial genome evolution by increasing the exchange of exogenous genes. Lateral gene transfer of integrans allows bacteria to adapt to different environmental pressures, such as antibiotic use in human medicine or in agriculture. However, little information is available about the origins of the resistance genes and the way in which they are assembled into gene cassettes. The large cassette arrays of some chromosomal integrans are potential sources of integran cassettes for mobile integrans (7, 13, 33, 43). Analysis of the gene cassette metagenome of environmental bacteria showed that

most of the chromosomal integrans include ORFs (whose products have no known function), and occasionally a chromosomal integran includes a single antibiotic resistance gene (2, 14, 21). However, the few antibiotic resistance cassettes found in chromosomal integrans cannot explain the large variety of resistance cassettes found in mobile integrans (8, 9). Interestingly, potential progenitors of the *cat* (chloramphenicol acetyltransferase) and *oxa2* (β -lactamase) cassette-associated genes have been found in the chromosomes of *Pseudomonas aeruginosa*, *Agrobacterium tumefaciens*, and *Synechocystis* (30), but not as cassettes. Despite 63 to 70% nucleotide identity between the gene cassettes and the corresponding chromosomal genes, *attC* sites have not been found downstream of the ORFs. These examples strongly suggest an independent origin for the genes and the *attC* sites and a mechanism of cassette formation. Furthermore, analysis of known integran cassettes suggested that their genesis involves an intermediate step involving reverse transcription of mRNA molecules in order to explain the absence of promoters, the paucity of noncoding sequence, and the presence of only one gene (11, 12, 30). However, this previous model did not explain how genes and *attC* sites are assembled or the origin of the RT. Recent studies showed that group IIC introns can target various transcriptional terminators (and that some group IIC introns target *attC* sequences) and that their ORFs code for a multidomain protein with RT activity (29, 31).

In this study, we investigated whether group IIC-*attC* introns could be involved in cassette formation by targeting and then joining distant genes and *attC* sites. Using an experimental system, we showed that the *S.ma.I2* intron from a clinical isolate is more efficient in targeting various *attC* sites than the *N.e.I1*, *G.s.I2*, and *S.b.I1* introns from environmental isolates are. Analysis of flanking exon sequences of group IIC-*attC* introns from the mobility assays and their natural genomic contexts showed that these introns recognize a consensus homing site, 5'-TTGT/T (IBS1-IBS3), which is at the 3' end of most *attC* sites (bottom strand) (28, 29). We suggest that the homing specificity of these introns may explain the precise juxtaposition of the structural gene stop codon and the inverted core site of the *attC* site in many integran cassettes (11). Furthermore, we showed that *S.ma.I2* targets in vivo at low levels non-*attC* sequences located downstream of resistance genes. Based on the genomic context, we suggest that the targeted sequences correspond to putative bidirectional transcriptional terminators. The data show that the target sites of group IIC-*attC* introns are not limited to *attC* sites but also include putative transcriptional terminator motifs, potentially leading to gene-intron intermediates in cassette formation. Despite the substantial bacterial genomic sequencing that has been done, none of the members of the group IIC-*attC* lineage has been found to be associated with putative terminators in nature. The low mobility frequencies observed here for the Term sites might be one of the reasons for this. An alternative explanation is that the intermediates are unstable.

The data presented here suggest a new model for integran cassette formation (Fig. 5). In this model, the first step involves homing of identical group IIC-*attC* introns to a transcriptional terminator adjoining a gene and to an isolated *attC* site (step 1). The targeted gene can be an ORF or a resistance gene with a putative bidirectional terminator, as shown here in vivo with

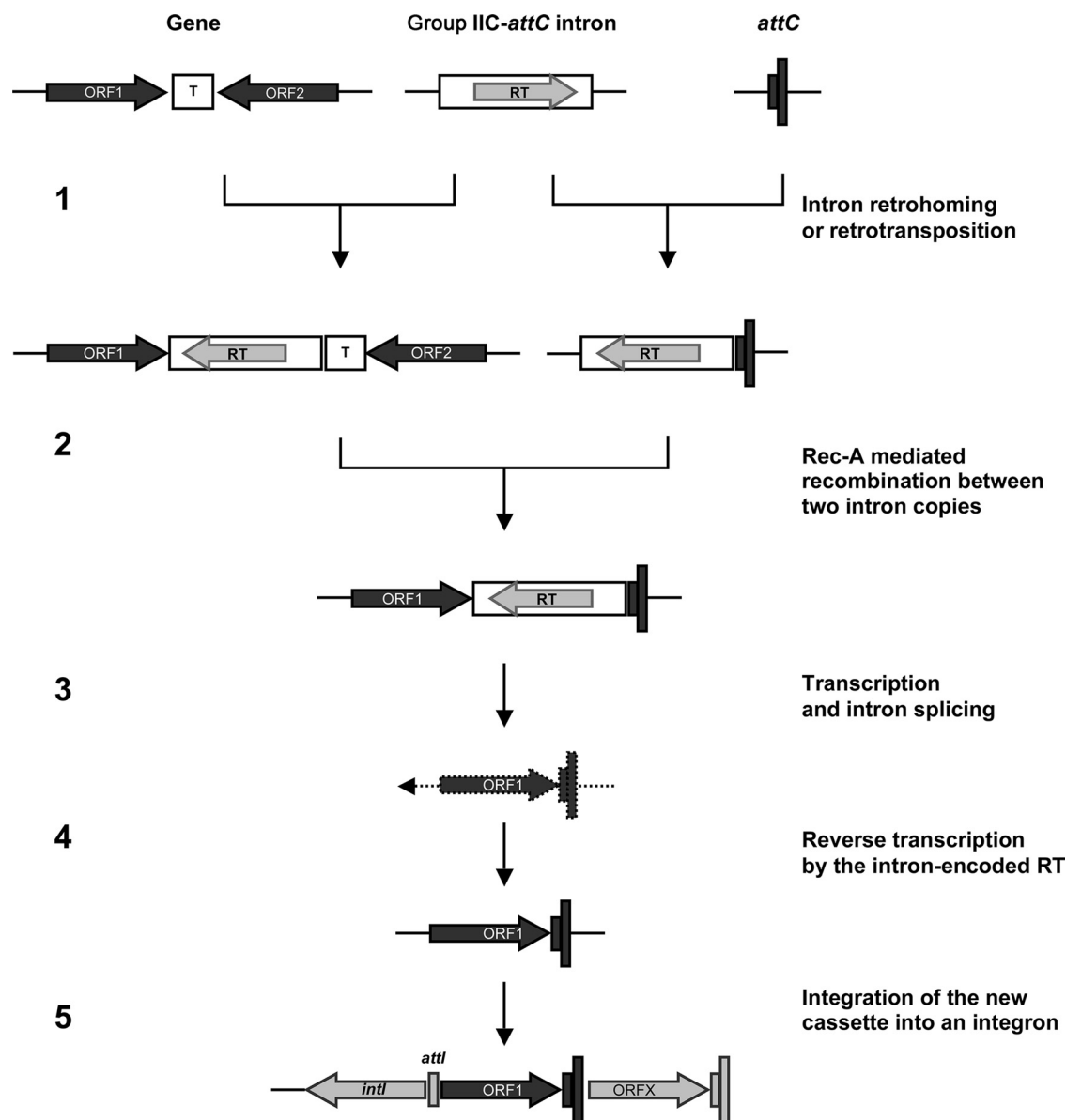


FIG. 5. Model for integrin cassette neoformation mediated by a group IIC-attC intron, adapted from a previously suggested mechanism (4). The proposed model involves two independent homing events placing one intron copy immediately downstream of a gene (ORF) and another intron copy in an isolated attC site (step 1). The next step involves recombination, possibly RecA mediated, between the two introns (step 2). The next event involves splicing of the intron (step 3), followed by reverse transcription using an unknown primer and the gene-attC RNA as the template (step 4). The final step involves integration of the cassette into the integron by an integrase (step 5).

S.ma.I2 (Fig. 3). Isolated intron-attCs have been found in *N. europaea*, *G. sulfurreducens*, and *Shewanella* sp. genomes without any nearby *intI* homolog (Fig. 1C). A recent study of the distribution of group IIC-attC intron target sites in complete bacterial genomes showed that single or multiple copies of a particular site (with $\geq 90\%$ nucleotide identity) are present in diverse organisms (28). Moreover, some targets are not confined to specific clades in bacterial orders. The next step involves recombination, possibly RecA mediated, between the two introns (step 2). Potential ORF-intron-attC cassette intermediates have been found in the *Shewanella* sp. metagenome (accession no. AACY020561240) (Fig. 1C) and in several mobile integrons (Fig. 1A). These intermediates may represent

“frozen” intermediates that are unable to complete the process or that are selected for by the expression of cassettes from the internal (intron) promoter oriented toward exon 1 (G. Léon, C. Quiroga, and P. H. Roy, unpublished data), particularly in *K. pneumoniae* (accession no. AJ971342), *S. marcescens* (accession no. AF453998), and *E. coli* (accession no. AY785243) clinical isolates (Fig. 1A). The next step involves full-length transcription of the cassette intermediate bottom strand sequence (i.e., the IEP coding strand) and splicing of the intron (step 3). This normally takes place in the gene’s chromosomal context and is not yet associated with an integron. Ligated gene-attC RNAs were detected by RT-PCR following the self-splicing of *N.e.I1* (data not shown) and *S.ma.I2* introns from

gene-intron-*attC* transcripts in vitro (29). The next step involves reverse transcription of the joined gene-*attC* RNA by the intron-encoded RT using an unknown primer or priming mechanism (step 4). We showed that purified RTs from two group IIC-*attC* introns (*N.e.I1* and *G.s.I2*) have both RNA- and DNA-dependent DNA polymerase activities in vitro. To our knowledge, this is the first report of a significant DNA-dependent DNA polymerase activity for bacterial group II IEPs (1). The final step involves integration of the cassette by site-specific recombination (step 5). While not all integron-containing strains contain group IIC-*attC* introns, cassettes formed in strains that do contain them would be exchanged by lateral transfer of their integrons, associated with plasmids or transposons. The unique structural characteristics of gene cassettes and the rapid appearance of new antibiotic resistance cassettes in mobile integrons (e.g., extended-spectrum β -lactamase and carbapenemase cassettes, such as *bla*_{IMP} and *bla*_{VIM}), which do not have homologs among chromosomal integron cassettes, suggest that there is an RT-based cassette formation mechanism that is boosted by the selective pressure of antibiotics. As more genome project sequences are assembled and analyzed, it is likely that more group IIC-*attC* introns will be shown to occur in genera in which chromosomal integrons are present and data will ultimately show the contribution of these introns to the cassette neoformation process.

ACKNOWLEDGMENTS

We thank Melanie Martin for technical assistance with the DNA polymerase assays.

This work was supported by Canadian Institutes of Health Research grant MT-13564 to P.H.R. G.L. was supported by a fellowship from Canadian Institutes of Health Research Strategic Training Initiatives in Health Research grant STP-59324.

REFERENCES

- Beauregard, A., M. J. Curcio, and M. Belfort. 2008. The take and give between retrotransposable elements and their hosts. *Annu. Rev. Genet.* **42**:587–617.
- Boucher, Y., M. Labbate, J. E. Koenig, and H. W. Stokes. 2007. Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* **15**:301–309.
- Boucher, Y., C. L. Nesbo, M. J. Joss, A. Robinson, B. C. Mabbutt, M. R. Gillings, W. F. Doolittle, and H. W. Stokes. 2006. Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. *BMC Evol. Biol.* **6**:3.
- Centron, D., and P. H. Roy. 2002. Presence of a group II intron in a multiresistant *Serratia marcescens* strain that harbors three integrons and a novel gene fusion. *Antimicrob. Agents Chemother.* **46**:1402–1409.
- Chen, C. Y., K. M. Wu, Y. C. Chang, C. H. Chang, H. C. Tsai, T. L. Liao, Y. M. Liu, H. J. Chen, A. B. Shen, J. C. Li, T. L. Su, C. P. Shao, C. T. Lee, L. I. Hor, and S. F. Tsai. 2003. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res.* **13**:2577–2587.
- Coleman, N., S. Tetu, N. Wilson, and A. Holmes. 2004. An unusual integron in *Treponema denticola*. *Microbiology* **150**:3524–3526.
- da Silva, A. C., J. A. Ferro, F. C. Reinach, C. S. Farah, L. R. Furlan, R. B. Quaggio, C. B. Monteiro-Vitorello, M. A. Van Sluys, N. F. Almeida, L. M. Alves, A. M. do Amaral, M. C. Bertolini, L. E. Camargo, G. Camarotte, F. Cannavan, J. Cardozo, F. Chambergo, L. P. Ciapina, R. M. Cicarelli, L. L. Coutinho, J. R. Cursino-Santos, H. El-Dorry, J. B. Faria, A. J. Ferreira, R. C. Ferreira, M. I. Ferro, E. F. Formighieri, M. C. Franco, C. C. Greggio, A. Gruber, A. M. Katsuyama, L. T. Kishi, R. P. Leite, E. G. Lemos, M. V. Lemos, E. C. Locali, M. A. Machado, A. M. Madeira, N. M. Martinez-Rossi, E. C. Martins, J. Meidanis, C. F. Menck, C. Y. Miyaki, D. H. Moon, L. M. Moreira, M. T. Novo, V. K. Okura, M. C. Oliveira, V. R. Oliveira, H. A. Pereira, A. Rossi, J. A. Sena, C. Silva, R. F. de Souza, L. A. Spinola, M. A. Takita, R. E. Tamura, E. C. Teixeira, R. I. Tezza, M. Trindade dos Santos, D. Truffi, S. M. Tsai, F. F. White, J. C. Setubal, and J. P. Kitajima. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**:459–463.
- Fluit, A. C., and F. J. Schmitz. 1999. Class 1 integrons, gene cassettes, mobility, and epidemiology. *Eur. J. Clin. Microbiol. Infect. Dis.* **18**:761–770.
- Fluit, A. C., and F. J. Schmitz. 2004. Resistance integrons and super-integrons. *Clin. Microbiol. Infect.* **10**:272–288.
- Gillings, M. R., M. P. Holley, H. W. Stokes, and A. J. Holmes. 2005. Integrons in *Xanthomonas*: a source of species genome diversity. *Proc. Natl. Acad. Sci. USA* **102**:4419–4424.
- Hall, R. M., D. E. Brookes, and H. W. Stokes. 1991. Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. *Mol. Microbiol.* **5**:1941–1959.
- Hall, R. M., and C. M. Collis. 1995. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.* **15**:593–600.
- Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleischmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477–483.
- Holmes, A. J., M. R. Gillings, B. S. Nield, B. C. Mabbutt, K. M. Nevalainen, and H. W. Stokes. 2003. The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ. Microbiol.* **5**:383–394.
- Holmes, A. J., M. P. Holley, A. Mahon, B. Nield, M. Gillings, and H. W. Stokes. 2003. Recombination activity of a distinctive integron-gene cassette system associated with *Pseudomonas stutzeri* populations in soil. *J. Bacteriol.* **185**:918–928.
- Jeong, S. H., K. Lee, Y. Chong, J. H. Yum, S. H. Lee, H. J. Choi, J. M. Kim, K. H. Park, B. H. Han, S. W. Lee, and T. S. Jeong. 2003. Characterization of a new integron containing VIM-2, a metallo-beta-lactamase gene cassette, in a clinical isolate of *Enterobacter cloacae*. *J. Antimicrob. Chemother.* **51**:397–400.
- Lambowitz, A. M., and S. Zimmerly. 2004. Mobile group II introns. *Annu. Rev. Genet.* **38**:1–35.
- Léon, G., and P. H. Roy. 2003. Excision and integration of cassettes by an integron integrase of *Nitrosomonas europaea*. *J. Bacteriol.* **185**:2036–2041.
- Léon, G., and P. H. Roy. 2009. Group IIC intron mobility into *attC* sites involves a bulged DNA stem-loop motif. *RNA* **15**:1543–1553.
- Martinez-Abarca, F., A. Barrientos-Duran, M. Fernandez-Lopez, and N. Toro. 2004. The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res.* **32**:2880–2888.
- Mazel, D. 2006. Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* **4**:608–620.
- Mazel, D., B. Dychinco, V. A. Webb, and J. Davies. 1998. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**:605–608.
- Michael, G. B., M. Cardoso, and S. Schwarz. 2008. Molecular analysis of multiresistant porcine *Salmonella enterica* subsp. *enterica* serovar Bredeney isolates from southern Brazil: identification of resistance genes, integrons and a group II intron. *Int. J. Antimicrob. Agents* **32**:120–129.
- Mohr, G., D. Smith, M. Belfort, and A. M. Lambowitz. 2000. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* **14**:559–573.
- Moran, J. V., S. Zimmerly, R. Eskes, J. C. Kennell, A. M. Lambowitz, R. A. Butow, and P. S. Perlman. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol. Cell. Biol.* **15**:2828–2838.
- Munoz-Adelantado, E., J. San Filippo, F. Martinez-Abarca, F. M. Garcia-Rodriguez, A. M. Lambowitz, and N. Toro. 2003. Mobility of the *Sinorhizobium meliloti* group II intron RmInt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. *J. Mol. Biol.* **327**:931–943.
- Pan, J. C., R. Ye, H. Q. Wang, H. Q. Xiang, W. Zhang, X. F. Yu, D. M. Meng, and Z. S. He. 2008. *Vibrio cholerae* O139 multiple-drug resistance mediated by *Yersinia pestis* pIP1202-like conjugative plasmids. *Antimicrob. Agents Chemother.* **52**:3829–3836.
- Quiroga, C., and D. Centron. 2009. Using genomic data to determine the diversity and distribution of target site motifs recognized by class C-*attC* group II introns. *J. Mol. Evol.* **68**:539–549.
- Quiroga, C., P. H. Roy, and D. Centron. 2008. The *S.ma.I2* class C group II intron inserts at integron *attC* sites. *Microbiology* **154**:1341–1353.
- Recchia, G. D., and R. M. Hall. 1997. Origins of the mobile gene cassettes found in integrons. *Trends Microbiol.* **5**:389–394.
- Robart, A. R., W. Seo, and S. Zimmerly. 2007. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc. Natl. Acad. Sci. USA* **104**:6620–6625.
- Rowe-Magnus, D. A., A. M. Guerout, L. Biskri, P. Bouige, and D. Mazel. 2003. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res.* **13**:428–442.
- Rowe-Magnus, D. A., A. M. Guerout, P. Ploncard, B. Dychinco, J. Davies, and D. Mazel. 2001. The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc. Natl. Acad. Sci. USA* **98**:652–657.

34. Sambrook, J., and D. W. Russell. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35. Schmelzer, C., and R. J. Schweyen. 1986. Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell* **46**:557–565.
36. 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.
37. Singh, N. N., and A. M. Lambowitz. 2001. Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol.* **309**:361–386.
38. 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.
39. Stokes, H. W., and R. M. Hall. 1989. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol.* **3**:1669–1683.
40. Stokes, H. W., D. B. O'Gorman, G. D. Recchia, M. Parsekhian, and R. M. Hall. 1997. Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol. Microbiol.* **26**:731–745.
41. Sunde, M. 2005. Class I integron with a group II intron detected in an *Escherichia coli* strain from a free-range reindeer. *Antimicrob. Agents Chemother.* **49**:2512–2514.
42. Toor, N., G. Hausner, and S. Zimmerly. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**:1142–1152.
43. Vaisvila, R., R. D. Morgan, J. Posfai, and E. A. Raleigh. 2001. Discovery and distribution of super-integrons among pseudomonads. *Mol. Microbiol.* **42**:587–601.
44. 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.
45. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66–74.
46. Yum, J. H., K. Yi, H. Lee, D. Yong, K. Lee, J. M. Kim, G. M. Rossolini, and Y. Chong. 2002. Molecular characterization of metallo-beta-lactamase-producing *Acinetobacter baumannii* and *Acinetobacter* genomospecies 3 from Korea: identification of two new integrons carrying the *bla*_{VIM-2} gene cassettes. *J. Antimicrob. Chemother.* **49**:837–840.
47. Zhong, J., and A. M. Lambowitz. 2003. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.* **22**:4555–4565.
48. Zimmerly, S., G. Hausner, and X. Wu. 2001. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res.* **29**:1238–1250.
49. Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**:3406–3415.